ELSEVIER

Cognitive Science 25 (2001) 287–313

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

A computational model of frontal lobe dysfunction: working memory and the Tower of Hanoi task

Vinod Goel[a,*], S. David Pullara[b], Jordan Grafman[c]

[a]*Department of Psychology, York University, Toronto, Canada and University of Aberdeen, Aberdeen, Scotland*

[b]*School of Computer Science, Simon Fraser University, Vancouver, Canada*

[c]*Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA*

Abstract

A symbolic computer model, employing the perceptual strategy, is presented for solving Tower of Hanoi problems. The model is calibrated—in terms of the number of problems solved, time taken, and number of moves made—to the performance of 20 normal subjects. It is then "lesioned" by increasing the decay rate of elements in working memory to model the performance of 20 patients with lesions to the prefrontal cortex. The model captures both the main effects of subject groups (patients and normal controls) performance, and the subject groups (patients and normal controls) by problem difficulty interactions. This leads us to support the working memory hypothesis of frontal lobe functions, but for a narrow range of problems. © 2001 Cognitive Science Society, Inc. All rights reserved.

Keywords: Frontal lobes; Computational model; Working memory; Problem solving; Tower of Hanoi; Short-term memory; 3CAPS; Production systems; Planning; Executive functions

1. Introduction

Working memory—the capacity to temporarily hold information in mind in the absence of external aids (and usually while some operation is performed on it)—is one strong candidate for a unitary account of the function of the dorsolateral prefrontal cortex (Goldman-Rakic, 1987). The idea can be traced back to Jacobsen's (1936) studies with lesioned

* Corresponding author. Tel. +011-44-1224-273-931; fax: +011-44-1224-273-426.
E-mail address: vgoel@yorku.ca (V. Goel).

PII: S0364-0213(01)00037-4

primates performing the delayed response task. In this task, one of two food wells are baited in full view of an animal, and then after a short delay (of a few seconds to minutes) the animal is allowed to reach into one of the two wells to retrieve the food. Successful performance on this task requires the animal to form and maintain an association between the food and a specific food well for the duration of the delay period. Jacobsen reported that primates with frontal lobe lesions were unable to perform the task. He interpreted the deficit as a loss of "immediate" memory, what we today call working memory.

Goldman-Rakic (1987; 1992; 1994) has followed up on this work with a series of lesion studies and single-cell recording studies in primates using tasks related to the delayed response paradigm. She has put forward the hypothesis that "the behavioral ineptitude of patients . . . on various neuropsychological tests may be reducible to an impairment of the mechanism(s) by which symbolic representations are accessed and held on line to guide behavior in the absence of (or in spite of) discriminative stimuli in the outside world" (Goldman-Rakic, 1987, p. 380).

A number of PET [O^{15}] studies have reported frontal lobe activation in working memory tasks. For example, Jonides et al. (1993) briefly presented subjects with groups of dots on a screen (200 ms) and then tested their ability to remember the location of dots after a 3 second delay. They reported activation in right-hemisphere prefrontal, occipital, parietal and premotor corticies which they associated with spatial working memory. Paulesu et al. (1993) asked subjects to remember six visually presented letters of the alphabet. They were asked to rehearse the letters (i.e. phonologically encode them) silently while waiting for a probe. A probe letter was then presented and subjects had to determine if the probe matched any of the six letters. They localized the phonological store to the left supramarginal gyrus.[1]

Roberts et al. (1994) used a concurrent task paradigm to see whether overloading working memory in a normal human population leads subjects to perform certain tasks in a manner similar to frontal lobe patients. They loaded subjects' working memory with an arithmetic task and then tested them on the Antisaccade task. In this task subjects are presented with a cue on screen. The natural tendency is to saccade to the cue. The task requires the subject to saccade away from the cue. Frontal lobe patients have difficulty doing this. Roberts et al. report that in the memory load condition normal subjects' performance is indistinguishable from that of frontal lobe patients.

Kimberg & Farah (1993) built a computer model that performs four well known frontal lobe tasks: motor sequencing, the Stroop task, the Wisconsin Card Sorting Test, and the context memory task. They report that when they "lesioned" the model by weakening the associations between elements in working memory, the performance of the model approximates frontal lobe patient performance as it has been reported in the literature, though they did not fit their model to any specific patient data.

We further this exploration of frontal lobe dysfunction by following up on the latter study. A symbolic computer model of another well-known frontal lobe task—the Tower of Hanoi—is presented. We fit the model—in terms of problems solved, time taken, and number of moves made—to the performance of 20 normal controls employing the perceptual strategy for solving the task. It is then "lesioned" by increasing the decay rate of elements in working memory to model the performance of 20 frontal lobe patients. The model captures both the main effects of subject groups (patients and normal controls) performance, and the

subject groups (patients and normal controls) by problem difficulty interactions. As such, it provides support for the working memory hypothesis for the dorsolateral prefrontal cortex. We conclude by emphasizing the limitations/scope of working memory explanations.

1.1. Tower of Hanoi task

The Tower of Hanoi is a puzzle consisting of three pegs and several disks of varying size. Given a start state, in which the disks are stacked on one or more pegs, the task is to reach a goal state in which the disks are stacked in descending order on a specified peg. There are three constraints on the transformation of the start state into the goal state. (1) Only one disk may be moved at a time. (2) Any disk not being currently moved must remain on a peg. (3) A larger disk may not be placed on a smaller disk.

The Tower of Hanoi is an instance of a $2^n - 1$ moves problem (where n is the number of disks that must be stacked). So for example, a 3 disk problem requires a minimum of 7 moves to complete, while a four disk problem requires at least 15 moves for completion. The number of disks in a problem provides a good measure of difficulty.

The start and goal states of our problems, along with the minimum number of moves required for solution, and the order of presentation are specified in Fig. 1. Our problems on the surface are all 5 disk problems. However, our start states are intermediate states where one or more of the disks have already been stacked or are in some intermediate position. So we will classify our problems into three levels of difficulty (easy, medium, and hard) by the minimum moves required for solution. Problems 5, 6, and 7 are the easy problems (mean of 7 moves); 1, 3, and 4 are the medium problems (mean of 10.6 moves); and 2, 8, and 9 are the hard problems (mean of 14.3 moves).

The Tower of Hanoi task is widely used as an experimental and diagnostic tool in the neuropsychology literature to gauge problem solving abilities. It is considered to be especially sensitive to frontal system dysfunction (Grafman, Litvan, Massaquoi, & Stewart, 1992; Mazzocco, Hagerman, & Pennington, 1992; Roberts, et al., 1994) and has become a staple in the evaluation of patients with frontal lobe lesions. It is routinely interpreted as a planning task in the neuropsychology literature and patients who display performance difficulties are said to have a “planning” deficit (Grafman, et al., 1992; Mazzocco, et al., 1992; Shallice, 1982; Shallice, 1988; Spitz, Minsky, & Bessellieu, 1985). The rational underlying this interpretation seems to be that, to successfully complete the task, subjects need to “look ahead” several levels deep and solve the problem in their heads, before physically moving any disks. If they are unable to solve the problem, it follows they were incapable of searching through the moves in their heads, and therefore they must have a “planning” or “look ahead” deficit.

However, there are reasons to believe that it is not an ideal problem to study planning (Goel & Grafman, 1995; Goel, Grafman, Tajik, Gana, & Danto, 1997) but under certain circumstances—when subjects use a strategy called the “perceptual strategy”—it is a interesting task to study working memory capacity in a complex problem solving situation.

There are a number of distinct strategies for solving the Tower of Hanoi problem. Simon (1975) discusses four of them. Different strategies make different demands on computational

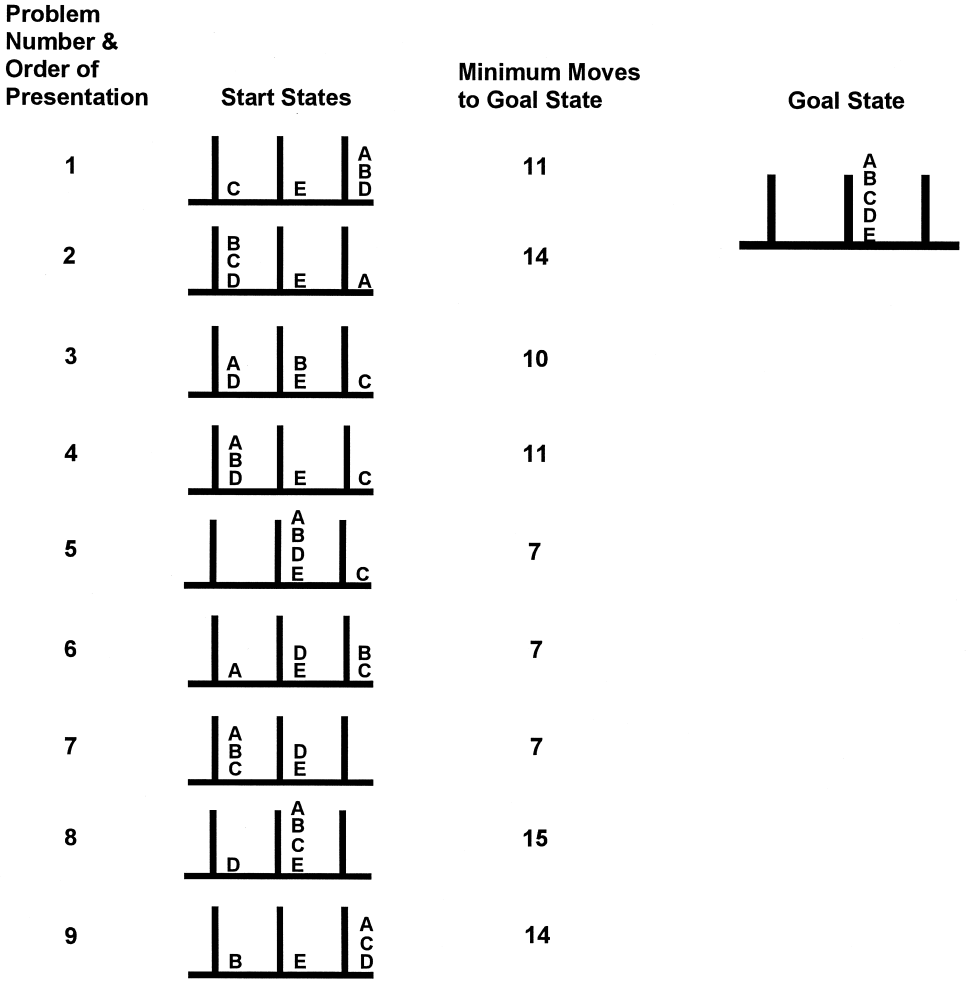


Fig. 1. Tower of Hanoi problems, ordered in the sequence administered.

resources (including working memory) so it is crucial to start with the strategy subjects are actually using. Goel and Grafman (1995) identified the perceptual strategy as the one being used by the normal control and patient subjects whose data we are modeling. In fact, of the four strategies identified by Simon (1975) it is the only one that can be applied at any intermediate state in the problem tree. Since all of our problems have an intermediate state as a start state, subjects are pretty well confined to this strategy.

The basic perceptual strategy is the most obvious and "natural" strategy. It can be applied at any point in the state space but does not lead to the shortest solution path. In its simplest form, it is primarily stimulus-driven, makes few demands on working memory (if disks $=<$ 3), and does not require the discovery or execution of arcane rules. The primary concept required is that of the "largest disk blocking the movement of a given disk." The strategy is the following (Simon, 1975):[2]

(1) if all n disks are placed on the target peg, stop; else
(2) find the next disk (i) to be placed on the target peg
(3) if there are smaller disks on top of disk *i*, clear them
(4) clear disks smaller than *i* off the target peg
(5) move disk *i* to the target peg
(6) go to 1

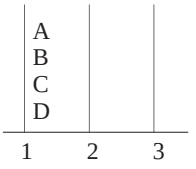
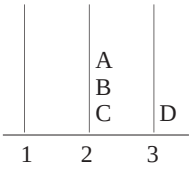
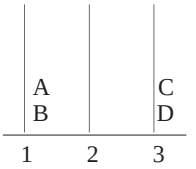
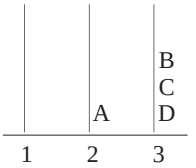
In this form the perceptual strategy is transparent and easy to execute. Because it is queuing off of the current disk configuration at each cycle, it does not require the simultaneous retention of multiple subgoal stacks (so long as the number of disks is not greater than the number of pegs). However, if the number of disks is greater than the number of pegs (i.e. greater than 3), then clearing the source peg to move disk *i* will block the target peg, and clearing the target peg will block the source peg, resulting in an infinite loop. The way to overcome this difficulty is to put the current subgoal stack in working memory on hold and instantiate additional subgoal stack(s). These additional stacks are used to make (and undo) certain temporary moves before proceeding with the strategy (see example below).

Let us consider an example. Table 1 shows a trace of our algorithm taking an *optimal* path through a 4-disk problem (peg 1 = source, peg 3 = target). The rows 1 through 4 of the Table correspond to the placement of disks of depth 4 through 1 respectively. Therefore, steps 1 through 4 correspond to the solution of a 4 disk problem. Steps 2 through 4 represent an *optimal* solution to a three disk problem (peg 2 = source, peg 3 = target). Column 2 specifies the goal that is being satisfied. Columns 3 and 4 indicate activity in subgoal stacks. The start state of the problem at each step is given in column 5 of the Table. The reference given in brackets after each step is to the line numbers of the algorithm specified below.

The first thing to note is that an increase in the number of disks to be placed results in an quantitative increase in the number of subgoals required for completion. Moving the deepest disk in the 3-disk problem requires 2 subgoals while moving the deepest disk in the 4 disk problem requires 4 subgoals. There is however, a more interesting, qualitative difference between the 3- and 4-disk problems. In the 3-disk problems, the two subgoals are completed sequentially so there is limited load on working memory. In the 4-disk problem, the first-level subgoal (clear D 1) results in an impasse (C $\rightarrow$?) as discussed above. Because of the ordering constraint on disks, none of the other pegs can accept the C disk. The satisfaction of the first subgoal (clear D 1) must be interrupted and the subgoal stack must be retained in memory while a second-level subgoal (clear base 2) is instantiated and completed. After the completion of this second-level subgoal, the subject must reactivate and complete the first-level subgoal. The next subgoal is to clear base 3. However, this third subgoal can only be satisfied by undoing the original subgoal (clear D 1). This is an example of the goal-subgoal conflict requiring a backward move mentioned in the above discussion. Finally, the first subgoal (clear D 1) stack must be reinstantiated after the completion of subgoal 3 (clear base 3). Only then can the original subgoal (clear D 1) finally be completed. Thereafter it becomes a 3-disk problem and the subgoals can be completed sequentially. A *nonoptimal* strategy increases both the number of subgoals considered sequentially and the number of subgoal stacks that must be retained simultaneously.

Goel & Grafman (1995) analysed data from 20 patients with focal frontal lesions solving

Table 1
Trace of perceptual strategy algorithm solving a 4-disk (and 3-disk) problem (optimal solution path). The start states are given in column 5. The goal state is to stack all the disks on peg 3. Goals and subgoals are shown in bold text. Disk movements are shown in italics. The numbers in brackets refer to line numbers of the algorithm specified below. "?" indicates an impasse. '**' indicates a backward move that undoes a previously completed subgoal.

	Goal Stack	Subgoals Level 1	Subgoals Level 2	Start States
1.	**D → base 3** (given)	**Clear D 1** (2b) *A → base 2* (3c/2a) *B → base 3* (3b/2a) *C → ?*	**Clear base 2** (2c) *A → B 3* (3a/2a/1)	
		C → base 2 (3a/2a)		
		Clear base 3 (2c) *A → D 1* (3c/2a)** *B → C 2* (3a/2a/1)		
		Clear D 1 (2b) *A → B 2* (3a/2a/1)		
	D → base 3 (2a/1)			
2.	**C → D 3** (given)	**Clear C 2** (2a) *A → D 3* (3c/2a) *B → base 1* (3a/2a)		
		Clear D 3 (2c)		
		A → B 1 (3a/2a/1)		
	C → D 3 (2a/1)			
3.	**B → C 3** (given	**Clear B 1** (2b) *A → base 2* (3a/2a)		
	B → C 3 (2a/1)			
4.	**A → B 3** (given)			
	A → B 3 (2a/1)			

the Tower of Hanoi and reported that when patients solved the problem they performed as well as the normal controls, however, patients failed to solve the problems more often than controls. But unlike much of the literature, they argued that patients' poor performance did not warrant any clear conclusions about "planning" or "look ahead" functions. They proposed alternative explanations involving working memory limitations and a "goal-subgoal conflict resolution" difficulty, and suggested that the latter may be a special problem for patients. We will here reuse the Goel and Grafman (1995) data to fit our computer model and offer a more refined explanation of frontal lobe patient performance based on our simulation. We will argue that a normal failure of "goal-subgoal conflict resolution", in conjunction with

working memory deficits, is *sufficient* to account for frontal lobe patient performance on the Tower of Hanoi. Claims of necessity cannot be made on the basis of a computer model.

2. Computer model

The computer model was built in a hybrid symbolic-connectionist architecture called 3CAPS (the third version of the Concurrent Activation-based Production System) developed by Just and Carpenter (Just & Carpenter, 1992). It focuses on the central role of working memory in constraining storage and processing (both spatially and temporally). It has been used to model a number of cognitive tasks including mental rotation, text processing, normal and aphasic sentence comprehension, human computer interaction and automated telephone interaction.

The 3CAPS architecture consists of a set of productions, a knowledge base, and a control mechanism, as follows:

1) A set of productions, (also called condition-action rules, or if-then rules), containing a left-hand side (LHS) and a right-hand side (RHS). The LHS contains patterns or conditions which when satisfied (i.e. matched against elements of working memory) cause the actions on the RHS to be executed. Productions fire once for every successful match of each condition on the LHS. The productions reside in procedural memory. The evaluation/execution of all productions in procedural memory constitutes a cycle.
2) A knowledge base contains information necessary for solving problems. This is the declarative memory of the system. All elements (also called facts) in declarative memory have an activation energy level associated with them. In order for LHS conditions to match a fact, its activation energy must equal or exceed the LHS threshold, which is defined at system startup. Any fact with less energy will not match in a LHS condition. Various mechanisms exist in 3CAPS for facts to lose or gain energy during the firing of productions.

   Facts can reside in different memory pools. The default memory pool is called the common pool. Each memory pool has an activation capacity, which refers to the total amount of energy available for memory elements in the pool. That is, the sum of the activation values of all the facts in a given memory pool cannot exceed the pool's capacity. If the addition of a fact into a pool causes its capacity to be exceeded, then the energy of all other memory elements in the pool are reduced in accordance to a specific heuristic.

   The common memory pool is exceptional in that it has unlimited capacity. Facts in the common pool are said to be in long-term memory. Facts in all other pools are said to be in working memory. LHS conditions can match facts in any memory pool, unless a specific pool is specified in the condition.
3) A control mechanism determines which productions are to be matched to the knowledge base, what order they are to be executed, and how conflicts are to be resolved. 3CAPS executes all productions in parallel, in any given cycle.[3] What this really means is that any change a production makes to declarative memory will not take place

until the end of the cycle. Thus actions on the RHS of a production will not have any effect on other productions during the same cycle. With this control strategy, the order that productions fire in is unimportant, and hence there is no need for conflict resolution.

2.1. Basic perceptual strategy algorithm

The Basic Perceptual Strategy was instantiated in the algorithm specified below. (The term "source peg" refers to the peg from which a disk is being moved. The "target peg" is the peg to which the disk will be moved to. The "auxiliary peg" is then the other peg. The term "goal peg" refers to the peg on which the disks are to be stacked.)

0. Halt if one of the following conditions have been met:
   - 0a. No goals remain on the goal stack.
   - 0b. The goal state has been reached.
   - 0c. Time to solve the problem has expired.
1. If the current goal has been achieved, remove it from the top of the goal stack.
2. If the current goal is to move a disk, then check the following (in order):
   - 2a. If the source and target pegs are clear, then move the disk to the target peg.
   - 2b. If the source peg is covered, then set as subgoal to clear the source peg. (Clearing a peg means to remove all disks on top of the disk we want to move.)
   - 2c. If the target peg is covered, then set as subgoal to clear the target peg. (For example, if peg 1 has B on top, and we want to move it onto disk C on peg 3, but disk A is sitting on C, then we need to clear peg 3 until C is on top.)
3. If the current goal is to clear a peg, then select the target peg as follows (in order):
   - 3a. If there is only one disk to clear, then let the auxiliary peg be the target peg.
   - 3b. If there is more than one disk to clear and only one of the other two pegs can accept a disk (i.e. has no disk, or a larger disk on it), then let that peg be the target peg.
   - 3c. If there is more than one disk to clear and both of the other two pegs can accept disks, then randomly select one of them as the target peg.
   - 3d. If there is more than one disk to clear and neither of the other two pegs can accept disks, then randomly set one of them to the target peg.

Note that all rules result in a new subgoal, except for 1 and 2a. Rule 1 removes the completed goal from the goal stack, and Rule 3 adjusts the internal state of the system to prepare for step 1.

The program solves the Tower of Hanoi problems by executing a 3CAPS implementation of the above algorithm. Since all problems have the same final state, (i.e. the stacking of all disks on peg 2), the program begins with the following "original" goals:

Depth 5: move disk E onto the base of peg 2
Depth 4: move disk D onto disk E on peg 2
Depth 3: move disk C onto disk D on peg 2
Depth 2: move disk B onto disk C on peg 2
Depth 1: move disk A onto disk B on peg 2

Like subjects, the program attempts to solve the problem by moving the largest disk to it's destination first, then the next largest, and so on. The "depth" value of a goal is a unique number representing it's position on the goal stack. The current goal is the one with the highest depth value. A subgoal generated in attempting to solve the current goal has a depth value one greater than the current goal. In any 3CAPS cycle, the program only tries to achieve the current goal. Once achieved, it is removed from the stack.

In step 3 of the algorithm, the subject is required to clear either a source or target peg before the goal (or subgoal) can be satisfied. In clearing the obstructing disk(s), a decision must be made as to which of the other two pegs it will be placed on. When there is only one disk to clear (as in step 3a of the algorithm), the correct move (irrespective of the configuration) is to move the blocking disk to the auxiliary peg. We found that our control subjects choose the correct peg about 75% of the time so we allowed the model to do the same. If there is more than one disk to clear from a peg (as in 3b–3d) and both other pegs can accept (3c) or not accept (3d) a disk, then in the absence of special knowledge subjects will choose either peg with equal probability.[4] We allowed the model to do the same.

Given the above algorithm and peg selection strategy, the program will find a correct solution given adequate time and memory resources. Our subjects did not have unlimited time and memory resources so we constrained these as below.

2.2. Parameters

We manipulated two parameters—move times and working memory activation—to fit the model to the normal control data. We scaled the time parameter as per the data and manipulated only the working memory parameter to fit the patient data. So our only free parameter in modeling patient data is working memory.

2.2.1. Move times

The subjects whose data we were modeling had a time limit of 2 minutes to solve each problem and occasionally ran out of time. A similar time constraint was needed for the model. But because symbolic computational models provide only a conceptual-level simulation there is no straight forward way of mapping them onto real-time performance. Therefore our absolute time scaling is arbitrary, but our relative times are motivated by common sense and data. The following two rules are used to determine relative move times: (i) executing a rule that requires physical movement of a disk will take longer than a rule that does not require a physical movement; (ii) it takes time to recall forgotten subgoals; the more subgoals that are forgotten, the longer it will take to recover.

Specifically we assigned a temporal duration to each of the program's rules and extracted a temporal penalty for recovery from forgetting every time a goal or subgoals were forgotten. The penalty for forgetting begins at a minimum cost, but shifts upwards with each goal forgotten. The range of the penalty remains fixed. Through experimentation, we found that allowing 1.5 units of time to elapse for rules that moved disks, and 1 unit of time to elapse for rules that did not move a disk, 1 to 4 units of time to elapse every time a goal/subgoal were forgotten along with a span shift of 1.5 units at every forgetting, gave a good

approximation of control subjects' time data and allowed us to calibrate the model and set a time limit of 2 "minutes" for each problem.

The Goel & Grafman (1995) data showed that frontal lobe patient times for solving the problems ($M = 85.0$, $SD = 4.5$) were 50% greater than normal control times ($M = 58.9$, $SD = 4.2$). We therefore scaled these parameters by 50% for patient simulations.

2.2.2. Working memory

Working memory is typically constrained by digit span and element duration. The working memory elements of interest to us are problem goals and subgoals. The goals are the five original goals specifying the solution to the problem statement. They are constant for all subjects and problems. Subgoals are the goals that subjects generate on route to achieving the solution.

Both goals and subgoals underwent decay based on the time they spent in working memory. Subgoals decayed according to the following formula:

$$\rho = e^{-at} \tag{1}$$

where t is the time the goal spent in memory, a is a constant, and the result p is the probability that the contents are forgotten. The curve generated by the equation when $a = -.15$ is graphed in Fig. 2 and approximates the short term memory curves for normal human subjects found in sources such as Peterson & Peterson (1959).

The decay of the original five goals was treated differently. Because they were part of the goal statement and could be easily remembered by encoding in a simple rule (place the disks in descending order) they were assumed to undergo a deeper level of encoding (Craik & Lockhart, 1972). Therefore, the time t that a goal spent in working memory was passed through equation (2) and t' was then passed on to the decay probability function and calculated as in (1) above:

$$t' = \max(0, \log(\max(1, t) - k)) \tag{2}$$

where t is the time the original goal has been in memory and k is a constant. A value of -2.7 for the constant k gave a good approximation of the normal control data.

2.3. Working memory requirements of our problems

Given our algorithm and the above parameters we can now specify the *differential* working memory requirements of our Tower of Hanoi problems. The easy problems (P5, P6, P7) will require some minimal simultaneous retention of multiple subgoal stacks 50% of the time during the placement of disk C (when peg selection is nonoptimal). When peg selection is optimal (50% of time) no simultaneous retention of multiple subgoal stacks is required in the easy problems. The medium problems (P1, P3, P4) place greater demands on working memory. Problems P1 and P4 can require simultaneous retention of multiple subgoal stacks twice, once during the placement of disk D, and again during the placement of disk C. The probability in each case is 50%. Thus there is a 75% chance that it will occur at least once. Problem P3 always requires simultaneous retention of multiple subgoal stacks at least once (during the placement of disk D) and may require it again during the placement of disk C (50% of the time). The three hard problems (P2, P8, and P9) all require simultaneous

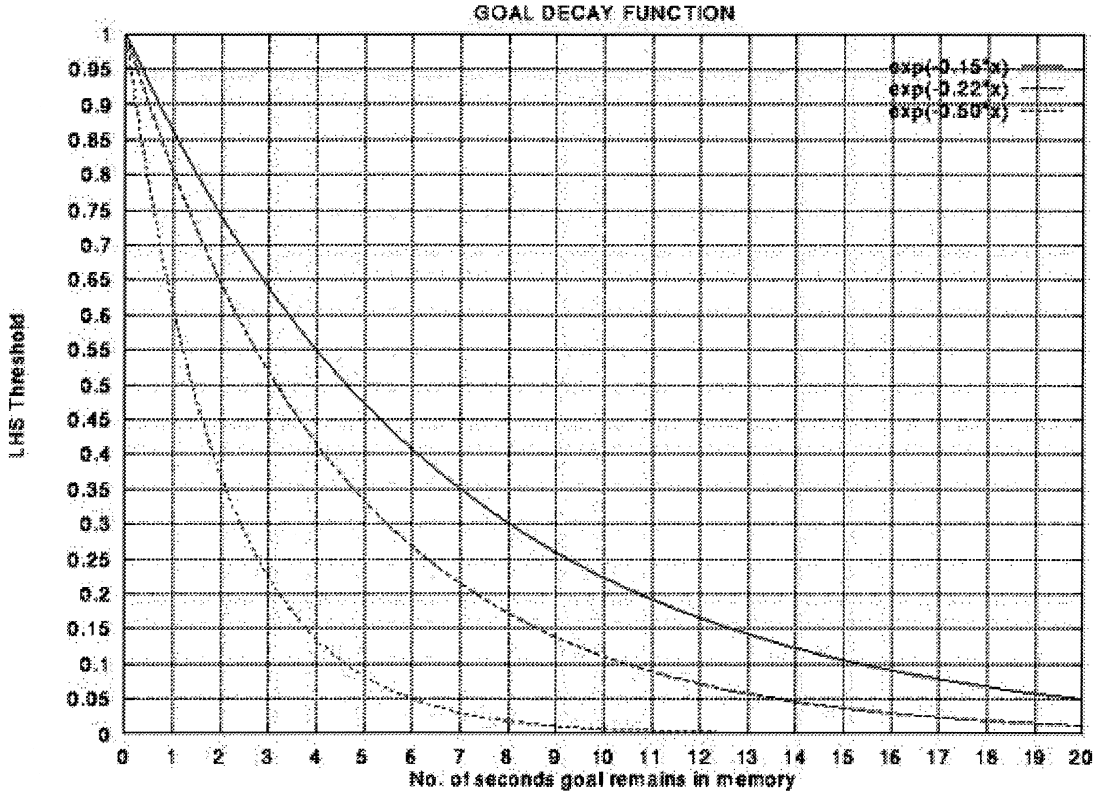


Fig. 2. Memory decay curve for different values of constant a.

retention of multiple subgoal stacks 100% of the time (and usually more than once). These differences in working memory requirements between the easy, medium, and hard problems will figure prominently in our explanation.[5]

3. Simulation results

The results of normal control and frontal lobe patient performance on the Tower of Hanoi are analysed in detail in Goel & Grafman (1995). Here we are interested in three sets of measures: (i) percent of problems solved, (ii) the number of moves required to solve problems, and (iii) the time required to solve problems. The data underwent a subject group (patients and controls) by problem difficulty (problems 1 to 9) ANOVA. Problem difficulty was a within-subject factor.

The main results for our purposes are the following: (1) Normal controls solved significantly more problems than patients ($M = 83.9\%$ ($SD = 36.9$) vs. 51.1% ($SD = 50.1$) respectively) ($F(1,304) = 21.8$, $p < .0001$) and there was a significant interaction between subject group (patients and controls) and problem difficulty ($F(8,304) = 2.3$, $p = .02$). These results are graphed in Fig. 3. (2) Normal controls took a mean of 13.8 ($SD =$

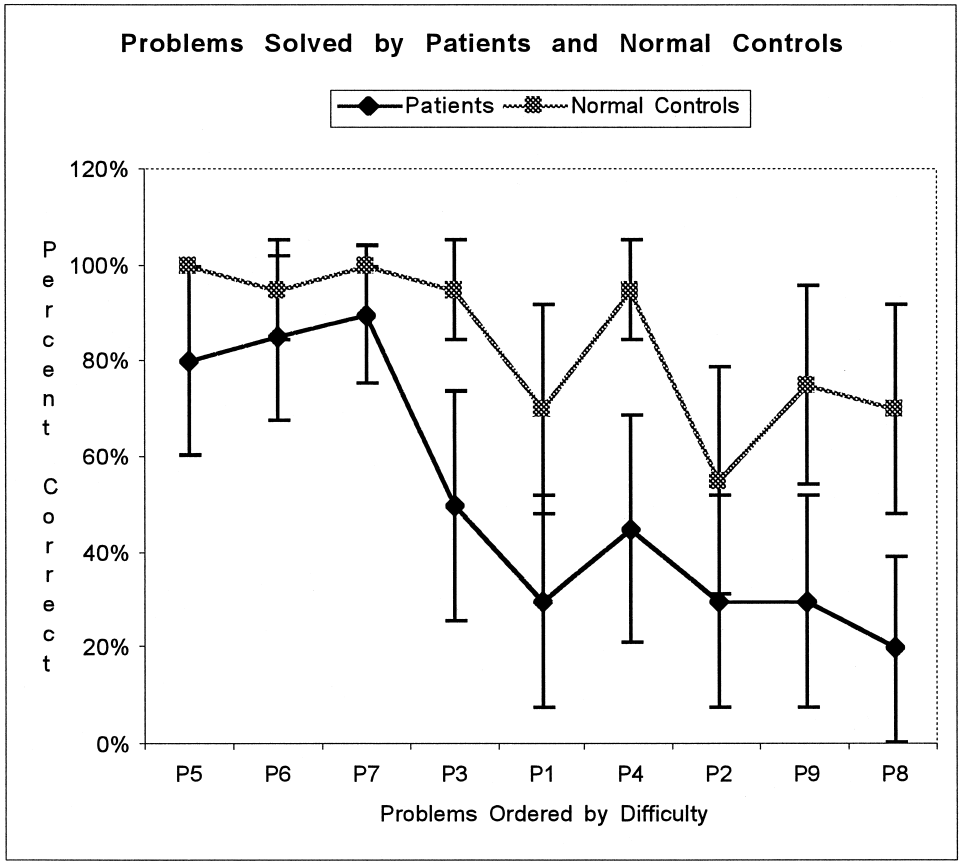


Fig. 3. Percent of problems correctly solved by normal controls and frontal lobe patients. Error bars depict the confidence interval at 95%.

6.1) moves to solve the problems compared to a mean of 11.6 ($SD = 5.3$) moves for the patients. This difference is significant ($F(1,304) = 7.7$, $p = .009$). However, there is no significant difference in the number of moves patients and controls ($M = 13.69$ ($SD = 4.6$) vs. 13.8 ($SD = 5.3$) respectively) made on problems solved correctly ($F(1,56) = 0.06$, $p = .81$). Both controls and patients required more than the minimum number of moves ($M = 10.67$) to solve the problems. (3) Patients with a mean time of 85.0 sec ($SD = 39.2$) took 50% longer than normal controls ($M = 58.9$ sec, $SD = 37.0$) to solve the problems. There is a near interaction in the time taken by subject groups (patients and controls) and problem difficulty ($F(8,304) = 1.8$, $p = .07$), and the main effects of subject groups and problem difficulty are significant ($F(1,304) = 16.2$, $p = .0003$ and $F(8,304) = 61.9$, $p < .0001$ respectively). Our computer model successfully captures these patterns of results.

The implication of these results are as follows: (1) The difficulties of the normal controls start with the hard problems where memory load is greatest. Patient difficulties start earlier with the medium difficulty problems which have a more modest memory load associated with them (hence the group by problem difficulty interaction). (2) When patients do solve the

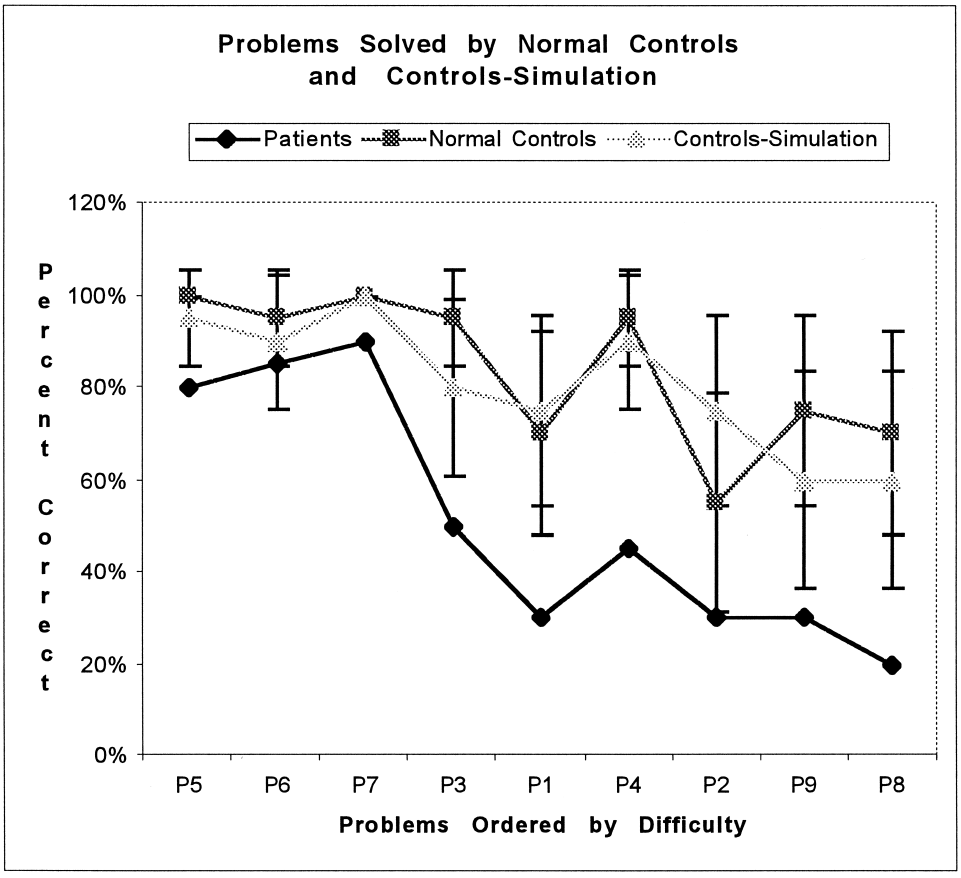


Fig. 4. Percent of problems correctly solved by normal controls and the controls-simulation. The frontal lobe patients' performance curve is shown as a point of reference. Error bars depict the confidence interval at 95%.

problems, they do so as efficiently as normal controls (based on number of moves). However, they fail to solve the problems more often. (3) Patients take 50% longer to make their moves.

It is also worth noting that, even though problems P1 and P4 are identical (except for the switching of the pegs), both the controls' and patients' performance is much better on problem P4 than P1. This is a function of learning. When people do the task, they learn from one problem to the next. So when they see the first problem repeated in the 4th position (except for the switching of the pegs) they do better.[6]

3.1. Fitting the NC data

We used the values of the parameters discussed above to simulate the Goel & Grafman (1995) normal control data with respect to problems solved, number of moves required to solve problems, and the time required to solve problems. Twenty trials, one corresponding to each subject were run. Fig. 4 graphs the problems solved data of the controls and the controls-simulation. The frontal lobe patient data is included as a reference point. As above,

the data underwent a subject group by problem difficulty (problems 1 to 9) ANOVA. The subject groups consisted of normal controls and the computer simulation of the normal controls (controls-simulation) and was a between-subject factor. Problem difficulty was a within-subject factor. An intraclass correlation coefficient (R_1), which combines a measure of correlation with a test in the difference of means (i.e. considers similarity in both the slopes and intercepts of the curves), is used as a measure of concordance between curves (Ebel, 1951; Kramer & Feinstein, 1981).[7] All reported intraclass correlations are significant. However, the magnitude of the correlation coefficients (R_1) are considered the more meaningful measures (Kramer & Feinstein, 1981).

There is no significant interaction between groups (controls and controls-simulation) and problem difficulty ($F(8,304) = 1.0$, $p = .42$), nor is there a significant difference between the overall performance of controls ($M = 83.9\%$, $SD = 16.5$) and the controls-simulation ($M = 80.6\%$, $SD = 14.5$) ($F(1,304) = 0.4$, $p = .53$). There is a significant difference between the controls-simulation and patient performance scores ($M = 80.6$ and $M = 50.6$ respectively) ($F(1,304) = 38.1$, $p < .0001$). The intraclass correlation coefficient for the controls and controls-simulation ($R_1 = 0.63$) is much higher than the corresponding coefficient for controls-simulation and patient data ($R_1 = 0.08$).

Fig. 5 graphs the number of moves normal controls and the controls-simulation used to solve the problems. Normal controls make a mean of 13.8 ($SD = 6.1$) moves per problem compared to a mean of 14.3 ($SD = 6.1$) for the controls-simulation. Again there is there no significant interaction between the control and controls-simulation groups and problem difficulty ($F(8,304) = 1.0$, $p = .40$) nor a significant main effect due to subjects ($F(1,304) = 0.97$, $p = .33$). The minimum number of moves for each problem are also graphed as a point of reference. The difference between the minimum moves ($M = 10.7$, $SD = 3.2$) and normal control moves ($M = 13.8$, $SD = 6.1$) approaches significance ($t(16) = -1.5$, $p = .14$) as does the difference between minimum moves and the controls-simulation moves ($M = 14.3$, $SD = 6.1$) ($t(16) = 1.8$, $p = .08$). The intraclass correlation coefficient for the number of moves made by controls and the controls-simulation ($R_1 = 0.84$) is much higher than the corresponding coefficient for controls-simulation and minimum moves ($R_1 = 0.17$).

Fig. 6 shows the time controls and the controls-simulation took to solve the problems. The normal controls took a mean of 58.9 sec ($SD = 37.0$) compared to 60.4 sec ($SD = 37.1$) for the simulation. There is no difference between these two means ($F(1,304) = 0.1$, $p = .75$). The solution times of the controls-simulation are significantly different from the solution times of the patients ($M = 85.0$ sec) ($F(1,304) = 18.8$, $p = .0001$). The intraclass correlation coefficient for the time taken by controls and the controls-simulation ($R_1 = 0.88$) is much higher than the corresponding coefficient for the time taken by the controls-simulation and patients ($R_1 = 0.19$).

3.2. Fitting the patient data

Having approximated normal control performance across the three dimensions of interest (problems solved, number of moves per problems, and time per problem) we “lesioned” the model by (i) increasing the time parameters by 50% based on data reported above, and (ii)

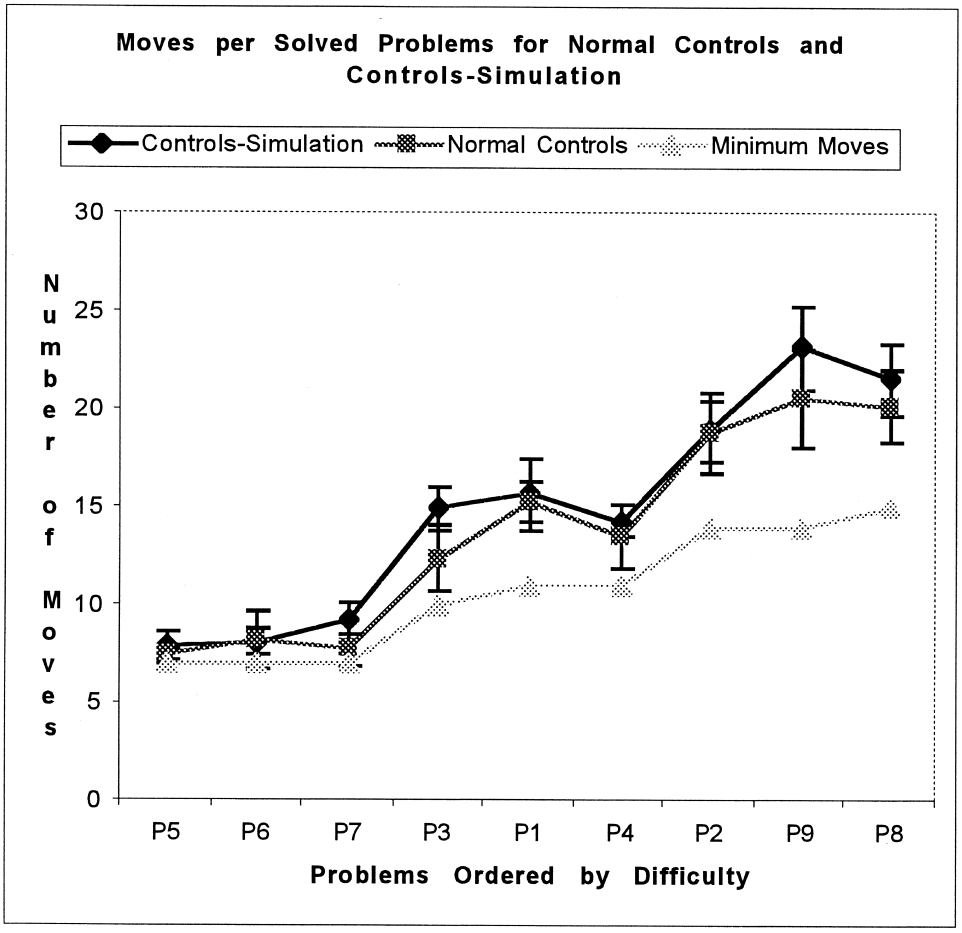


Fig. 5. Number of moves taken by normal controls and the controls-simulation to solve problems. The minimum moves for each problem are shown as a point of reference. Error bars depict the confidence interval at 95%.

decreasing the temporal duration of working memory elements using a value of $a = 0.22$ (see equation 1 and Fig. 2 above). Fig. 7 shows the solution curves for the controls-simulation and the impact on the curve of these manipulations.

The result of just increasing the time parameters is a simple nonselective downward shift in the solution curve from a mean of 80.6% ($SD = 0.40$) for the controls-simulation to 68.3% ($SD = 0.46$) for the time-simulation. The main groups (controls-simulation and time-simulation) effect is significant ($F(1,304) = 7.6$, $p < .05$) but there is no interaction between the subject groups (controls-simulation and time-simulation) and problem difficulty ($F(8,304) = 1.1$, $p = .35$) (as there is in the control and patient data).

If we compare the scores of the time-simulation with the patient scores ($M = 68.3\%$ ($SD = 46.0$) and $M = 51.1\%$ ($SD = 50.0$) respectively) we find a significant difference between these two means ($F(1,304) = 7.2$, $p = .01$) and a significant interaction between

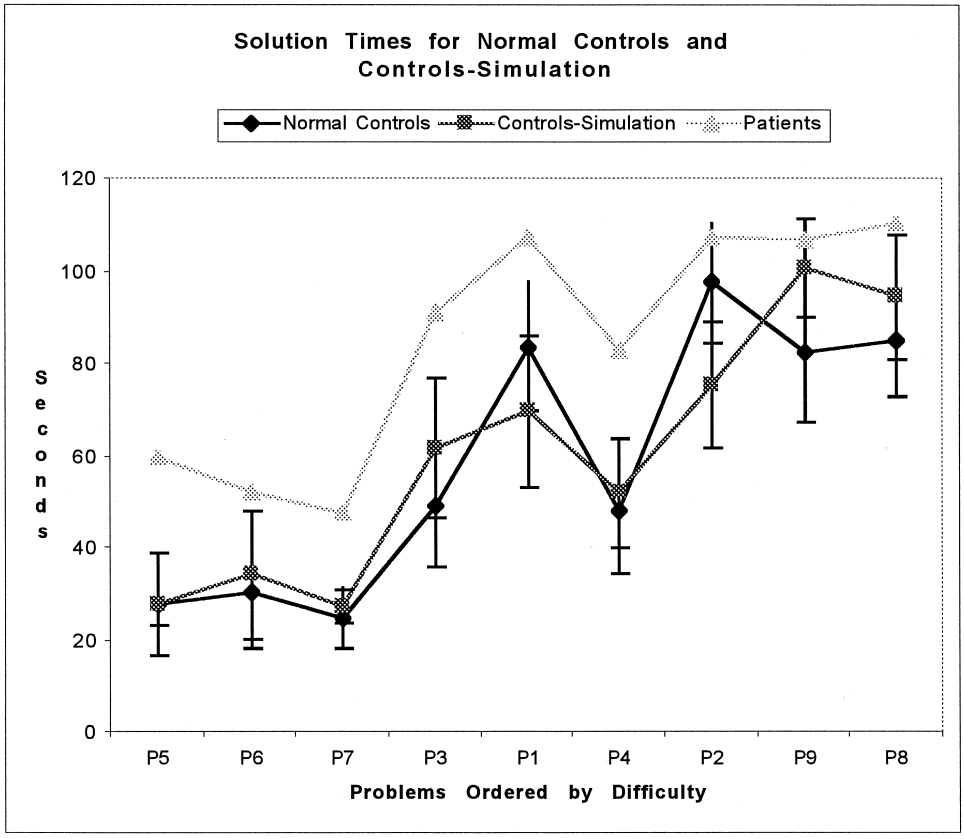


Fig. 6. Time taken by normal controls and the controls-simulation to solve problems. The frontal lobe patients' time curve is shown as a point of reference. Error bars depict the confidence interval at 95%.

groups (time-simulation and patients) and problem difficulty ($F(8,304) = 2.3$, $p = .01$), suggesting that the time-simulation is not capturing the patient data.

The consequences of introducing the working memory "deficit" are more selective than the consequences of the time parameters (Fig. 7). This is because, with the basic perceptual strategy, working memory is differentially involved as the number of disks to be placed increases. Therefore the effect is felt most on the hard and medium problems and is minimal on the easy problems. The resulting curve gives a very good approximation of the frontal lobe patient data. The patients-simulation solved a mean of 49.4% ($SD = 51.0$) of the problems compared to a mean of 51.1% ($SD = 50.0$) for the patients. There is no statistical difference between these means ($F(1,304) = 0.061$, $p = .81$). Neither is there any interaction between the groups (patients and patients-simulation) and problem difficulty ($F(8,304) = 0.72$, $p = .68$). Perhaps, most critically, the working memory manipulation does capture the interaction between subject groups (patients-simulation and normal controls) and problem difficulty that characterizes the patient and control data ($F(1,304) = 2.3$, $p = .02$).[8] The intraclass correlation coefficient for the number of problems solved by patients and the patients-simulation ($R_1 = 0.89$) is much higher than the corresponding

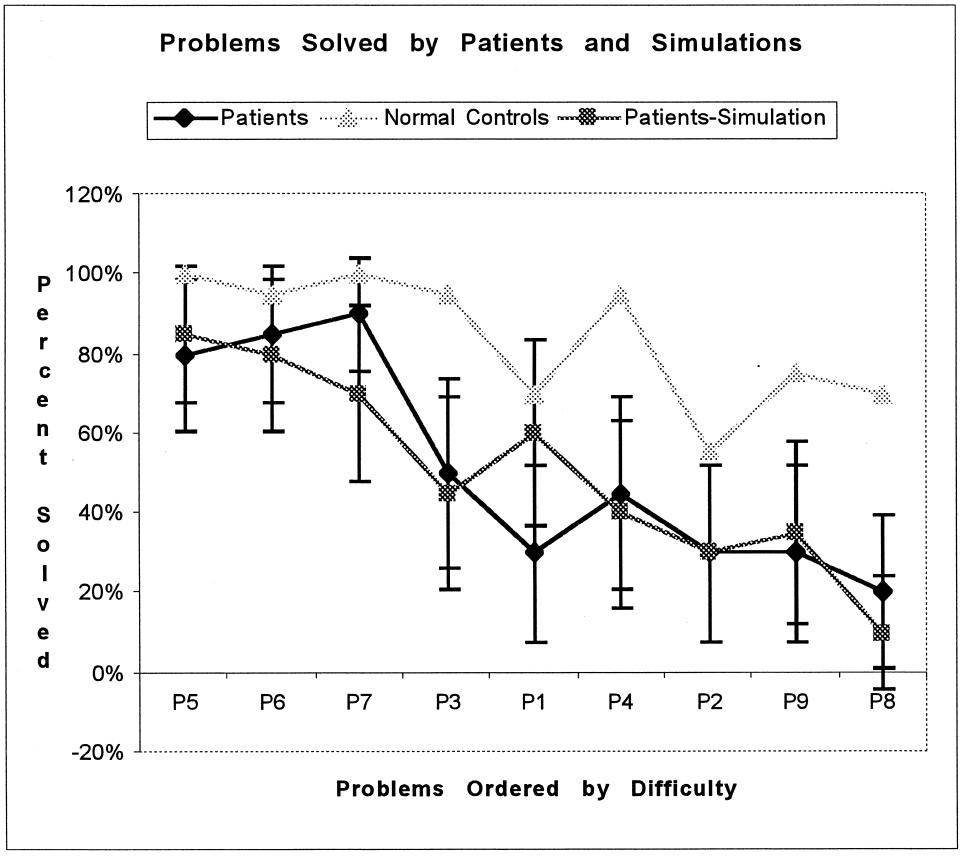


Fig. 7. Percent of problems correctly solved by frontal lobe patients and the controls-simulation (as a function of working memory manipulation). The normal controls' performance curve is shown as a point of reference. Error bars depict the confidence interval at 95%.

coefficient for the number of problems solved by the patients-simulation and controls (R_1 = 0.05).

There is also a good fit between the number of moves the patients-simulation and patients required to solve the problems (M = 11.9 and 11.7 for patients-simulation and patients respectively). (These figures represent the mean number of moves made during solved problems only.) There is no interaction between the patient and patients-simulation groups and problem difficulty ($F(2,34) = 0.21$, $p = .81$). Neither is there a significant main group effect of patients and patients-simulation ($F(1,17) = 0.65$, $p = .43$).[9] The intraclass correlation coefficient for the number of moves made by patients and the patients-simulation (R_1 = 0.97) is much higher than the corresponding coefficient for patients-simulation and minimum moves (R_1 = 0.24). These data are graphed in Fig. 8.

Finally, we want to examine the time it took patients and the patients-simulation to solve the problems (Fig. 9). The overall solution times for the patients-simulations was 82.6 sec (SD = 41.5) compared to 85.0 sec (SD = 39.2) for the patients. This main group effect is not significant ($F(1,304) = 0.82$, $p = .67$). Neither is there any interaction between

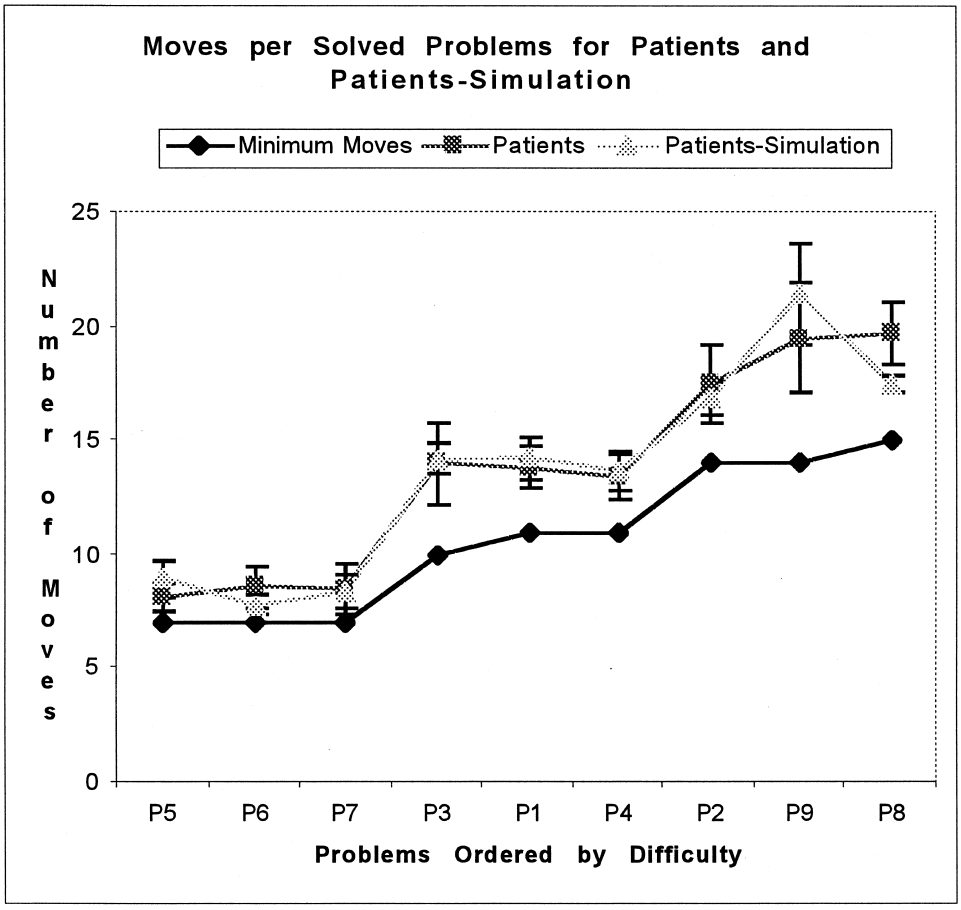


Fig. 8. Number of moves taken by frontal lobe patients and the patients-simulation to solve problems. The minimum moves for each problem are shown as a point of reference. Error bars depict the confidence interval at 95%.

groups and problem difficulty ($F(8,304) = 1.6$, $p = .11$). If we eliminate the first problem (because of leaning curve) the overall solution time mean for the patients-simulations becomes 82.9 sec ($SD = 42.2$) compared to 82.2 sec ($SD = 39.8$) for the patients and provides an improved fit to the data. (This main group effect is not significant ($F(1,266) = 0.009$, $p = .92$). Neither is there any interaction between groups (patients and patients-simulation) and problem difficulty ($F(7,266) = 0.76$, $p = .62$).) The intraclass correlation coefficient for the time taken by patients and the patients-simulation ($R_1 = 0.87$) is much higher than the corresponding coefficient for the time taken by the controls and patients-simulation ($R_1 = 0.20$).

4. Discussion and conclusion

The theoretically relevant aspects of the model are (i) the basic perceptual strategy algorithm peg selection strategy; (ii) time parameters; and (iii) the working memory param-

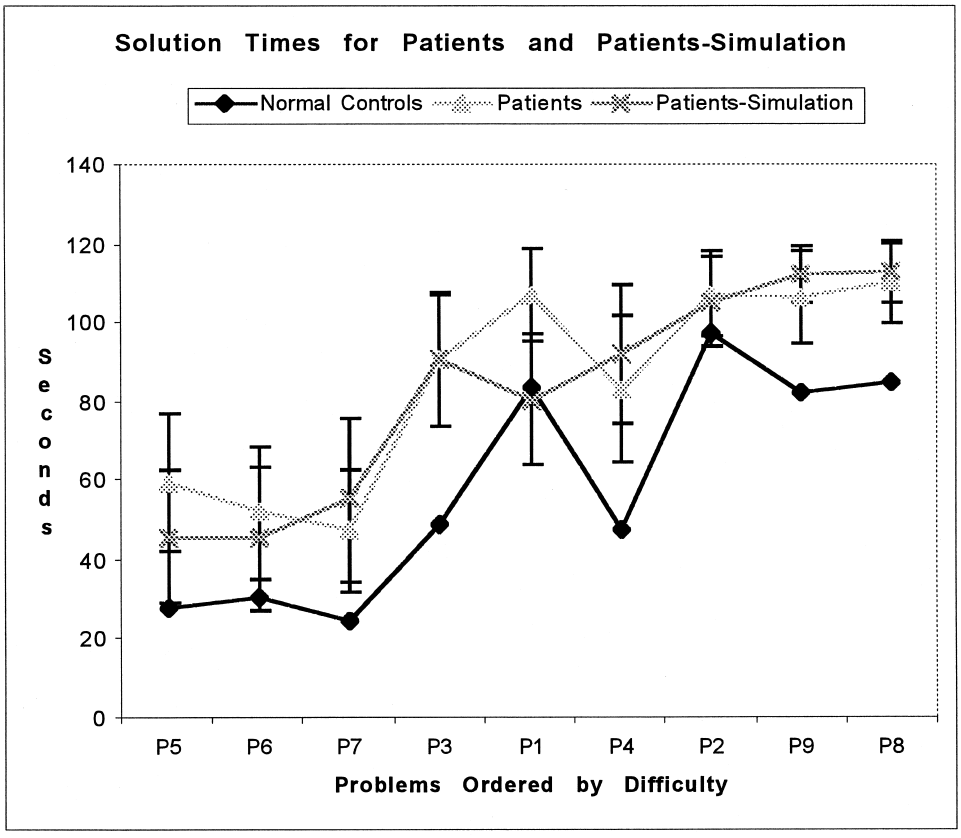


Fig. 9. Time taken by frontal lobe patients and the patients-simulation to solve problems. The normal controls' time curve is shown as a point of reference. Error bars depict the confidence interval at 95%.

eter. The first were left unchanged during the patient simulation so they serve the same roles in the control and patient simulations. We manipulated move times and working memory.

The perceptual strategy algorithm determines the path subjects take through the state space. As already noted above, there is considerable evidence that both our controls and patients were indeed using the perceptual strategy (Goel & Grafman, 1995). Other strategies such as the recursive strategy, or the move pattern strategy (Simon, 1975) have different memory requirements that will result in different performance characteristics if working memory is manipulated. The peg selection strategy also plays an important role in the behaviour of the model. The setting of these parameters determines goodness of peg selection. Nonoptimal peg selection increases both the frequency and duration of simultaneously maintaining multiple subgoal stacks. Nonoptimal peg selection occurs with 50% probability in both control and patient simulations during algorithm rules 3b–3d and with 25% probability during rule 3a (see discussion above).

The time parameter was adjusted based on empirical data. The effect of this adjustment was a simple, nonselective shift in the performance curve equally affecting problems of varying difficulty. This makes sense because it is just reducing the termination point.

Our only freely manipulated parameter is working memory. Its manipulation has a more selective effect on performance. The medium and hard problems are affected more than the easy problems. As noted above, this follows from the basic perceptual strategy algorithm. The requirement to simultaneously maintain multiple subgoal stacks in working memory is minimal in the easy problems, but becomes substantive in the medium and hard problems.

Our simulation results indicate that if we "lesion" our computer model of normal controls by increasing the time parameters (as per patient data), and increasing the rate of decay of activation of working memory elements, that is sufficient for the model to give a very good approximation—in terms of solution rate, times, and moves—of frontal lobe patient performance on the Tower of Hanoi task.

This explanation is a refinement of the explanation offered by Goel & Grafman (1995) in their original analysis of this patient data. They appealed to a combination of working memory and a goal-subgoal conflict resolution issue to explain frontal lobe patient performance on the Tower of Hanoi task. The idea is that when there is an odd number of disks, the problem requires a counter-intuitive, backward move, which superficially/locally takes you away from the global goal, to achieve the local subgoal. That is, one needs to differentiate between local subgoals and the global goal and acknowledge a conflict between the two, and at certain times inhibit the global goal and be guided by a local subgoal.

For example, in a 3-disk problem, where all the disks are stacked on peg P1 at the start state, and the goal state is to stack all the disks on peg P3, the globally correct first move (and the most natural move) is to place disk 1 on peg P2, because peg P3 needs to kept clear for disk 3. If disk 1 is placed on Peg 3 it will need to be moved off before disk 3 can be placed, so why not place it on P2 to begin with? However, if we begin by placing disk 1 on P2, then the only place to put disk 2 is P3. Unless we wish to reverse moves at this point, the next move must place disk 1 from P2 to P3. This clears P2 to accept disk 3. But it was P3 on which we wanted to place disk 3! The correct first move in this situation is to place disk 1 on peg P3, despite the fact that it violates the global goal and will need to be moved again. Locally, it is the correct thing to do.

This maps nicely onto the accepted frontal lobe deficit of failing to inhibit the prepotent (incorrect) response in favour of the alternative (correct) response. Goel & Grafman (1995) felt it necessary to appeal to this explanation, in addition to working memory, because they considered the medium difficulty problems to be 3-disk problems (because the number of moves required for their completion was closer to 7 than 14). As 3-disk problems, these problems should not have any substantive working memory requirements, and patients' performance on these problems should have been comparable to their performance on the easy problems. But there is a sharp decrease in patient performance between the easy and medium problems, suggesting some other mechanism is coming into play. Therefore Goel & Grafman appealed to failure of inhibition of the prepotent response to explain the drop in patient performance from the easy to medium problems. Our computer model allows us to refine this explanation.

The peg selection strategy in our model allows us to manipulate the "failure of inhibition of the prepotent response". The peg selection manipulation allows us to affect the frequency with which subjects inhibit the global goal in favour of a local subgoal. But unless the computer model has difficulty in maintaining the subgoal stacks in working memory, the

only effect of this manipulation is an increase in the time and number of moves required for solution. So peg selection on its own is not adequate to model patient performance. But normal levels of "failure of inhibition of the prepotent response", in conjunction with decreased temporal duration of elements in working memory, does give us results that do justice to patient data. The model captures both the main effects of patients and normal controls performance, and the subject groups (patients and normal controls) by problem difficulty interactions. In doing so, it suggests that working memory deficits (in conjunction with normal levels of inhibition) are sufficient (though maybe not necessary) to account for patient performance on the Tower of Hanoi task. It also demonstrates how a deficit in a basic computational mechanism—that one might think should result in a uniform cognitive deterioration—can result in selective cognitive deficits.

There are several existing computer models of frontal lobe dysfunction (Cardoso & Parks, 1998; Cohen & Servan-Schreiber, 1992; Dehaene & Changeux, 1991; Kimberg & Farah, 1993). Of these the first three are connectionist models, while the latter is a production-based model like ours. A comparison of our model with the Kimberg & Farah (1993) model—which also uses a working memory manipulation—is quite straight forward. Instead of explicitly manipulating the duration of elements in working memory, they weakened the connections among elements in working memory. This would result in elements receiving less activation from neighbors and thus their activation would fall below some threshold more quickly and would no longer trigger productions. In our model we explicitly increased the decay rate of working memory elements for the patients-simulation. But the net effect should be basically the same across the two models.

Dehaene & Changeux (1991) present a connectionist model of the WCST and then lesion the model to simulate frontal lobe patient performance. They look at 3 manipulations (i) sensitivity to negative feedback; (ii) retaining previously tested rules and avoiding retesting them; (iii) rejecting some rules on a priori grounds. They report that the first manipulation results in performance that most looks like patient data. While the details of their model are radically different from ours and they have no explicit working memory component, varying the sensitivity to negative feedback could certainly be modeled as a working memory manipulation in our system. Indeed Kimberg & Farah (1993) have done something very similar in their model.

Cohen & Servan-Schreiber (1992) present a connectionist model of schizophrenic performance in the Stroop task, the continuous performance test, and a lexical disambiguation task. They argue that the function of frontal cortex is to maintain the internal representation of "context information" and that schizophrenia is directly associated with abnormalities of the frontal cortex, in particular the disruption of this "context information" by which they mean "information held in mind in such a form that it can be used to mediate an appropriate behavioral response" (p. 46). This is certainly consistent with most notions of working memory. However, they go on to differentiate their notion of context information from working memory as follows (p. 46): "We usually think of short-term memory as storing recently presented information, the identity of which must later be retrieved—"declarative" representations in the sense used by Anderson (1983). In contrast, we think of internal representations of context as information stored in a form that allows it to mediate a response to the stimulus other than the simple reporting of its identity." So rather then being

information that is being directly manipulated by the processor, context information serves some supporting role in maintaining the activity of the working memory elements that will be directly manipulated. However, it is certainly possible that the same representations and mechanisms underlie their notion of context information and the more general notion of working memory. Indeed recent extensions of this model (Cohen, Braver, & O'Reilly, 1996; O'Reilly, Braver, & Cohen, 1999) have explicitly appealed to, and developed the notion of working memory. They provide interesting insight into how high-level, symbolic models, totally lacking in biological plausibility, can be mapped onto the biology of the prefrontal cortex.

Cardoso and Parks (1998) present a connectionist model of the Tower of Hanoi task and then "degrade" the model to simulate frontal lobe patient performance by scaling back the activation of the nodes in some subset of the network. This could potentially correspond to a working memory manipulation. However, what they are capturing in their model are the transformation rules (as associations) between input and output patterns of individual moves. In the degraded condition, what is being weakened is the knowledge the system has acquired about the transformation rules. They do not model the sequences of transformations and thus do not maintain and update a representation of subgoals and current states. This makes it difficult to compare it to our model, or indeed to evaluate it as a model of "executive functions".

We would like to conclude by comparing the working memory hypothesis with two other important computationally based hypothesis of frontal lobe dysfunction, the ideas of a central executive (Shallice, 1988; Shallice & Burgess, 1991) and structured event complexes (Grafman, 1994; Grafman, 1995), and by sounding a cautionary note about the scope and limitations of working memory explanations.

The central executive is, intuitively, the control mechanism of the cognitive system. These mechanisms are usually built into an architecture. So, for example, given the standard architecture of a Turing Machine one might equate the back and forth motion of the head with the central executive whereas the portion of the tape that is being read/written would be associated with working memory. In our 3CAPS architecture the built in control structure is that all productions that match fire in parallel.[10]

On the Norman and Shallice model (Shallice & Burgess, 1991) if exactly one production/operator matches a working memory element it simply fires. This is considered the routine case. However, if (i) no productions/operators match elements in working memory or (ii) several productions match working elements then control is passed onto a central executive or the supervisory attentional system (SAS). The SAS responds by altering the activation level of productions thus changing (i.e. increasing or decreasing) their probability of firing. In our particular model what this requires is that we account for frontal lobe patient performance by manipulating the firing thresholds of the LHS of productions that reside in long-term memory.

On Grafman's (1994; 1995) structured event complexes (SEC) model, SECs are large-scale knowledge structures that reside in long-term memory and guide much of our routine behaviour. In frontal lobe patients these structures can be damaged or at least their retrieval impaired. Certainly patients can retrieve individual events—even small sets of events.

However, some events may be retrieved out of order and others may demonstrate inappropriate duration of activation.

In terms of our computer model, these knowledge structures would correspond to the production rules of the perceptual strategy algorithm that reside in long-term memory. The SEC theory predicts difficulties in retrieving these rules. The activation of some rules may be impaired, others may be activated out of order, etc. The impairment occurs because of actual damage to the productions such as a change in their firing threshold.

So there are potentially interesting computational differences between the working memory hypothesis and the SAS and SEC hypothesis, though these distinctions only make sense given some strong assumptions about the functional architecture of the cognitive system (Pylyshyn, 1984). In our model the working memory hypothesis requires a manipulation of the activation of elements in working memory that trigger the LHS of productions, whereas the SAS and SEC theories require a manipulation of the actual firing thresholds of the productions. Given this, the question arises, might these alternate manipulations yield the same results as the working memory manipulation, and if so, are there any reasons for preferring one over the other.

If the manipulations of the activation levels of the LHS of productions were uniform across productions, then the net effect should be similar to our working memory manipulation (though conceptually it is a different manipulation). But both the SAS and SEC hypotheses require a selective and content sensitive manipulation of the firing levels of individual productions. The SAS account requires the specification of a mechanism that selects which productions to modify and to what extent. Needless to say such a mechanism is notoriously difficult to specify, and would require at least the power of a Turing Machine, thus raising the specter of an infinite regress. On the SEC account the manipulation of the firing thresholds is not carried out by some intelligent mechanism, but is just a chance function of a lesion.

So the question becomes, could we get the same performance results from our model if, instead of manipulating working memory, we *selectively* manipulated the firing threshold of *individual* rules in our algorithm. The answer may be ‘yes’, but it remains to be demonstrated. But an important consideration is that a working memory manipulation is a content-free structural manipulation, and such a manipulation—for reasons of elegance, coverage, parsimony—needs to be given preference over content sensitive manipulations, particularly in a content-free task like the Tower of Hanoi.

In conclusion, we want to sound a cautionary note about the scope of the working memory hypothesis. Frontal lobe patients exhibit performance deficits in a diffuse range of cognitive tasks and social situations (Stuss & Benson, 1986). Our results confirm that performance deficits in some tasks such as the Tower of Hanoi can be elegantly explained in terms of working memory deficits. Kimberg and Farah have shown similar results for the motor sequencing task, the Stroop task, the Wisconsin Card Sorting Test, and the context memory task. However, it is equally important to remember that many tasks in which the frontal lobes are implicated—cognitive estimation (Shallice & Evans, 1978), judgment tasks (Damasio, 1994), planning tasks (Goel & Grafman, 2000; Goel, et al., 1997) and inductive reasoning tasks (Goel & Dolan, 2000)—are unlike these modeled tasks.

Important dimensions along which these two sets of tasks differ are in terms of structure

and content. Tasks like the Tower of Hanoi, the WCST, or deductive reasoning are formal, well-structured, content-impoverished rule-governed tasks. If you can discover and/or understand the few rules then you can complete the task independent of any real-world knowledge. Working memory manipulations (which are content-free structural manipulations) are very effective in such situations as the current study, and Kimberg and Farah (1993) illustrate. Real-world, ill-structured tasks like planning and inductive reasoning require enormous amounts of judgement and knowledge about the world. It is harder to see the relevance of an appeal to working memory in these situations. (Damasio, 1994; Goel & Grafman, 2000; Goel, et al., 1997; Shallice & Burgess, 1991).

There is of course no contradiction here. In our search for unifying accounts of frontal lobe function, we sometimes forget that we are talking about one third of the human cortex. It would be truly amazing if it incorporated just one simple mechanism. It is perfectly consistent with the known facts that the frontal lobes should incorporate multiple mechanisms, one of which is working memory.

Notes

1. PET [O^{15}] studies of the Tower of London (a task related to the Tower of Hanoi) have shown cortical activation in similar (overlapping) areas (Baker, Rogers, Owen, Frith, Dolan, Frackowiak, et al., 1996; Rezai, Andreasen, Alliger, Cohen, Swayze, & O'Leary, 1993).
2. The "basic" perceptual strategy must be differentiated from the "sophisticated" perceptual strategy in which disks must be cleared in order of size, irrespective of what peg they are on (Simon, 1975).
3. We did not utilize the parallelism. We sequentialized our productions by inserting conditions that ensure that only one of them will fire during any given cycle. This seemed intuitively more plausible for our task and allowed us to better control instantiation of new goals, and avoid conflicts.
4. More accurately, if there is an odd number of disks to move, then the first disk should go to the goal peg; if there is an even number of disks, it should go to the other peg.
5. All of the problems require the retention of the goal and the current state.
6. We tested this by switching the ordering of problems 4 and 1 in one study with normal controls. Subjects performed better in problem 1.
7. The crucial claims with respect to the patient and control data have to do with the main effects of subject groups (patients and normal controls) performance, and the subject groups (patients and normal controls) by problem difficulty interactions. These aspects of the data are captured by the model and highlighted by the ANOVA. It is also desireable to have a "goodness of fit" measure between subject data and computer-simulation data. A chi square analysis is ruled out because the within-subject factor (problem difficulty) of the experiment design violates the independence assumption. A simple correlation analysis is also problematic. It captures the slopes of the curves but not the distance between them. Because the slopes of all the data points fall within a narrow range, the correlation between all of them is significant. On

this analysis, the correlation between the patient and control data is r = 0.81. This clearly misses important distinctions. To deal with this problem, we have reported the intraclass correlation, which takes into consideration not only the slopes of two curves, but also the absolute magnitude between them.

8. A post hoc analysis of score pairs on individual problems reveals that all pairs are statistically identical. The most divergent scores are for problem P1 ($t(38) = 1.3$, $p = .20$). This difference is due to the learning curve effect experienced by human subjects. Our computer model does not have a learning component and cannot model this.
9. The fact that these measures are made only for solved problems results in a reduced number of cases and an unbalanced statistical model. Therefore the nine problems were collapsed into 3 groups for this statistical analysis.
10. Though as noted above we bypassed this for a sequential control structure.

Acknowledgments

We are indebted to Marcel Just and Pat Carpenter for inviting the first author to the Cognitive Modeling in 3CAPS Workshop, Carnegie-Mellon University, July 14–20, 1996, and for making available the 3CAPS software. We thank Peter Roosn-Runge, Dan Kimberg and S. Dehaene for useful discussion and comments on earlier drafts of the manuscript. We also thank Chris Green and John Crawford on advice on statistical issues. This work was supported in part by grants from York University's Faculty of Arts funds, and NSERC Canada to the first author.

References

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge: Harvard University Press.

Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S., & Robbins, T. W. (1996). Neural Systems Engaged by Planning: A PET Study of the Tower of London Task. *Neuropsychologia*, *34*, 515–526.

Cardoso, J., and Parks, R. W. (1998). Neural Network Modeling of Executive Functioning with the Tower of Hanoi in Frontal Lobe—Lesioned Patients. In R. W. Parks, D. S. Levine, & D. L. Long (Eds.), *Fundamentals of Neural Network Modeling: Neuropsychology and Cognitive Neuropsychology and Cognitive Neuroscience*. Cambridge: MIT Press.

Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1996). A Computational Approach to Prefrontal Cortex, Cognitive Control, and Schizophrenia: Recent Developments and Current Challenges. *Philosophical Transactions of the Royal Society of London Series B, 351*, 1515–1527.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, Cortex, and Dopamine: A Connectionist Approach to Behavior and Biology in Schizophrenia. *Psychological Review*, *99*(1), 45–77.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Damasio, A. R. (1994). *Descartes' Error*. NY: Avon Books.

Dehaene, S., & Changeux, J.-P. (1991). The Wisconsin Card Sorting Test: Theoretical Analysis and Modeling in a Neuronal Network. *Cerebral Cortex*, *1*(Jan/Feb), 62–79.

Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika*, *16*, 407–24.

Goel, V., & Dolan, R. J. (2000). Anatomical Segregation of Component Processes in an Inductive Inference Task. *Journal of Cognitive Neuroscience*, *12*(1), 1–10.

Goel, V., & Grafman, J. (1995). Are Frontal Lobes Implicated in "Planning" Functions: Interpreting Data from the Tower of Hanoi. *Neuropsychologia*, *33*(5), 623–642.

Goel, V., & Grafman, J. (2000). The Role of the Right Prefrontal Cortex in Ill-structured Problem Solving. *Cognitive Neuropsychology*, *17*(5), 415–436.

Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A Study of the Performance of Patients with Frontal Lobe Lesions in a Financial Planning Task. *Brain*, *120*, 1805–1822.

Goldman-Rakic, P. S. (1987). Circuitry of Primate Prefrontal Cortex and Regulation of Behavior by Representational Memory. In V. B. Mountcastle, F. Plum, & S. R. Geiger (Eds.), *Handbook of Physiology: The Nervous System Vol. 5 Higher Functions of the Brain, Part 1* (pp. 373–417). Washington, D.C.: American Physiological Society.

Goldman-Rakic, P. S. (1992). Working Memory and the Mind. *Scientific American*(September), 111–117.

Goldman-Rakic, P. S. (1994). The Issue of Memory in the Study of Prefrontal Cortex. In A. M. Thierry (Eds.), *Motor and Cognitive Functions in the Prefrontal Cortex* (pp. 113–121). Berlin: Springer-Verlag.

Grafman, J. (1994). Alternative Frameworks for the Conceptualization of Prefrontal Lobe Functions. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology*, *Vol. 9* (pp. 187–202). Amsterdam: Elsevier.

Grafman, J. (1995). Similarities and Distinctions Among Models of Prefrontal Cortical Functions. In J. Grafman, K. J. Holyoak, & F. Boller (Eds.), *Structure and Function of the Human Prefrontal Cortex* (pp. 337–368). N.Y.: Annals of the New York Academy of Sciences.

Grafman, J., Litvan, I., Massaquoi, S., & Stewart, M. (1992). Cognitive Planning Deficit in Patients with Cerebellar Atrophy. *Neurology*, *42*(8), 1493–1496.

Jacobsen, C. F. (1936). Studies of Cerebral Functions in Primates. I. The Function of the Frontal Association Areas in Monkeys. *Comparative Psychology Monographs*, *13*, 1–60.

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., & Minoshima, S. (1993). Spatial Working Memory in Humans as Revealed by PET. *Nature*, *363*, 623–625.

Just, M. A., & Carpenter, P. A. (1992). A Capacity Theory of Comprehension: Individual Differences in Working Memory. *Psychological Review*, *99*(1), 122–149.

Kimberg, D. Y., & Farah, M. J. (1993). A Unified Account of Cognitive Impairments Following Frontal Lobe Damage: The Role of Working Memory in Complex, Organized Behavior. *Journal of Experimental Psychology: General*, *122*(4), 411–428.

Kramer, M. S., & Feinstein, A. R. (1981). Clinical Biostatistics LIV. The Biostatistics of Concordance. *Clin. Pharmacol. Ther*., *29*(1), 111–123.

Mazzocco, M. M. M., Hagerman, R. J., & Pennington, B. F. (1992). Problem Solving Limitations Among Cytogenetically Expressing Fragile X Women. *American Journal of Medical Genetics*, *43*, 78–86.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A Biologically-Based Computational Model of Working Memory. In A. Miyanke & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* NY: Cambridge University Press.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The Neural Correlates of the Verbal Component of Working Memory. *Nature*, *362*, 342–345.

Peterson, L. R., & Peterson, M. J. (1959). Short-term Retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198.

Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, Massachusetts: MIT Press.

Rezai, K., Andreasen, N. C., Alliger, R., Cohen, G., Swayze, V., & O'Leary, D. S. (1993). The Neuropsychology of the Prefrontal Cortex. *Arch Neurol*, *50*, 636–642.

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal Cognitive Processes: Working Memory and Inhibition in the Antisaccade Task. *Journal of Experimental Psychology: General*, *123*(4), 374–393.

Shallice, T. (1982). Specific Impairments of Planning. *Philosophical Transactions of the Royal Society of London*, *Series B*, *298*, 199–209.

Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Shallice, T., & Burgess, P. (1991). Higher-Order Cognitive Impairments and Frontal Lobe Lesions in Man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal Lobe Function & Dysfunction* Oxford: Oxford University Press.

Shallice, T., & Evans, M. E. (1978). The Involvement of the Frontal Lobes in Cognitive Estimation. *Cortex*, *14*, 294–303.

Simon, H. A. (1975). The Functional Equivalence of Problem Solving Skills. *Cognitive Psychology*, *7*, 268–288.

Spitz, H. H., Minsky, S. K., & Bessellieu, C. L. (1985). Influence of Planning Time and First-Move Strategy on Tower of Hanoi Problem-Solving Performance of Mentally Retarded Young Adults and Nonretarded Children. *American Journal of Mental Deficiency*, *90*(1), 46–56.

Stuss, D. T., & Benson, D. F. (1986). *The Frontal Lobes*. N.Y: Raven Press.