Available online at www.sciencedirect.com
ScienceDirect
Brain and Cognition 66 (2008) 11–20
Brain and Cognition
www.elsevier.com/locate/b&c

Counterintuitive and alternative moves choice in the Water Jug task ☆

Hassina P. Carder *, Simon J. Handley, Timothy J. Perfect

School of Psychology, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

Accepted 30 April 2007
Available online 6 June 2007

Abstract

MOVE problems, like the Tower of London (TOL) or the Water Jug (WJ) task, are planning tasks that appear structurally similar and are assumed to involve similar cognitive processes. Carder et al. [Carder, H.P., Handley, S.J., & Perfect, T.J. (2004). Deconstructing the Tower of London: Alternative moves and conflict resolution as predictors of task performance. *The Quarterly Journal of Experimental Psychology 57a, 8*, 1459–1483] showed that one predictor of TOL performance was the number of alternative move choices there were at a given point in the solution. In two experiments an individual move experienced on the WJ task was manipulated (perceptually consistent/counterintuitive) along with the number of alternative moves there were to choose between. A verification paradigm was employed in which participants made speeded judgements about the correctness of a move. Results showed performance was consistent with the application of a perceptual strategy accompanied by a process involving the evaluation of non-redundant alternative moves. These are discussed in the context of recent research that has examined the impact of executive dysfunction on Water Jug performance [Colvin, M.K., Dunbar, K., & Grafman, J. (2001). The effects of frontal lobe lesions on goal achievement in the Water Jug task. *Journal of Cognitive Neuroscience, 13*, 1129–1147].

Keywords: Executive function; Memory; Planning; Problem solving; Normal adults; Performance; Counterintuitive; Problem solving strategy; Water Jug; Tower of London

1. Introduction

One aspect of cognition that is uniquely human is the ability to hypothesise about the future, and this is one way that we can direct and control our lives to achieve our goals. Planning has been proposed as a higher-level cognitive function that is closely associated with executive control processes (e.g. Baddeley & Hitch, 1974; Norman & Shallice, 1980/1986; Burgess, 1997; Morris & Ward, 2005). Planning has been defined as a form of problem solving in situations where there are two distinct states of affairs, and the agent wants to get from one state to the other but it is not clear how the gap is best bridged. Hence, planners model a path from A to B, assessing the consequences of proposed actions and revising their plans until a set of actions are determined that can effectively attain the goal state (Goel, 2002). At the heart of this definition is the concept that different actions are considered when planning and problem solving, and this is a key theme of the current research.

In the information processing literature human cognition has been understood by building process models that can determine the route from A to B. Typically the tasks employed are MOVE problems such as the Tower of Hanoi (TOH), Tower of London (TOL) (Shallice, 1982, 1988) and the Water Jug task. These tasks are also traditionally employed as planning tasks in psychology and neuropsychology. There has been a good deal of research that focussed on the disk moving tasks (TOH and TOL), but much less is known about the Water Jug task (originally described by Luchins, 1942), which is the focal task of the current paper.

☆ This research is funded by an ESRC grant to the first author, No. RES-000-22-0607.
* Corresponding author. Fax: +44 (0) 1752233392.
E-mail address: hcarder@plymouth.ac.uk (H.P. Carder).

0278-2626/$ - see front matter
doi:10.1016/j.bandc.2007.04.006

In the Water Jug task there are three vessels of different sizes and a quantity of water. The aim is to transfer the starting distribution of water to a goal state by pouring water between the vessels. Fig. 1 illustrates the Water Jug task in its traditional form, and the reader is invited to try to solve this problem. The rules of the task are that you must carry on pouring until the jug is empty or the target jug is full. There are two valid solutions for this problem, one with eight moves and one with seven moves, which are explained in Appendix A.

The Water Jug task is similar to the disk moving tasks as both require the transformation from a start state to a goal state, and both require the manipulation of material (i.e. water/disks) between three different regions (i.e. jars/pegs), also all MOVE tasks have counterintuitive moves which appear to move away from the goal state, but are necessary for an optimal solution (Morris, Miotto, Feigenbaum, Bullock, & Polkey, 1997; Welsh, Satterlee-Cartmell, & Stine, 1999).

Research that has employed the disk moving tasks has shown that performance is closely related to spatial memory (e.g. Owen, Downes, Sahakian, Polkey, & Robbins, 1990; Owen, Doyon, Petrides, & Evans, 1996; Gilhooly, Wynn, Phillips, Logie, & Della Sala, 2002; Handley, Capon, Copp, & Harper, 2002) and inhibitory processes

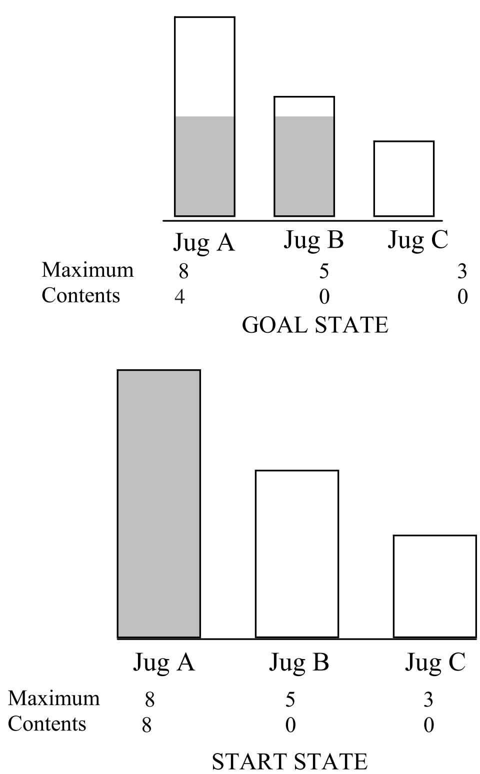


Fig. 1. The Water Jug task depicting vessels that can hold a maximum of 8 units of water, 5 units of water and 3 units of water. To solve the problem you must transfer the start state (8, 0, 0) into the goal state (4, 4, 0) by pouring water between vessels. Water must be poured until the pouring jug is empty or the target jug is full. The minimum number of pours required to solve this problem is 7.

(e.g. Goel & Grafman, 1995; Welsh et al., 1999; Miyake et al., 2000). The spatial memory component is thought to be involved in the representation and manipulation of spatial information in memory and has been clearly demonstrated (e.g. Handley et al., 2002; Phillips, Wynn, Gilhooly, Della Sala, & Logie, 1999; Gilhooly et al., 2002; Phillips, Gilhooly, Logie, Della-Sala, & Wynn, 2003). Inhibition is claimed to be involved in those situations when one alternative move is chosen over another. Specifically when a heavily preferred move is wrong and should be inhibited in favour of as less dominant but more effective alternative. For example, Miyake et al. (2000) employed CFA modelling to show that TOH performance loaded most heavily on an inhibitory control factor. An additional indirect line of evidence comes from the observation that frontal lobe patients perform poorly on these tasks (e.g. Shallice, 1982; Owen et al., 1990; Goel & Grafman, 1995; Morris et al., 1997). Several researchers have claimed that the presence of counterintuitive moves in the disk moving tasks elicits this poor performance (e.g. Morris et al., 1997; Goel & Grafman, 1995). In these cases it is necessary to inhibit the incorrect move that is consistent with the goal state in favour of a correct counterintuitive alternative move which appears to lead the solver away from the goal state.

There are some clues from the problem solving literature to suggest that the Water Jug and Tower tasks may invoke different problem solving strategies. The means-end analysis is one heuristic that provides a route into the depth of the problem space (Newell & Simon, 1972). This is where sub-goals are established that satisfy progress between the current state and the goal state. For example, Ernst and Newell (1969) developed the general problem solver (GPS) that could apply a means end heuristic to solve MOVE problems. GPS could successfully solve the TOH using this strategy but failed to solve the Water Jug task. Ernst and Newell (1969) argued that that the source of GPS′ failure was that it was ineffective at forward planning.

However, claiming that forward planning was not required, Atwood and Polson (1976) argued that participants need only use information about the current state and all its immediate successors to solve the Water Jug task. They argued that GPS′ failure was due to an overdependence on the subgoal strategy, when a more flexible approach would have been more successful. Atwood and Polson (1976) built an alternative model that could solve the task. It employed two sets of processes, a perceptual strategy and memory processes. The perceptual strategy compared the current state with all possible successive states to determine that which looked most similar to the goal. This is achieved by the calculation of an evaluation function (EVF) in which the goal contents were subtracted from the current contents (ignoring the sign) of the largest and middle sized jugs, to provide a numerical indicator of the perceptually similarity. The EVF formula and some example calculations are given in Appendix B. The second process was a memory function that allowed these products to be stored in memory, along with information about previous states that had been visited or evaluated, thus moves are either acknowledged as new or recognised as old. According to Atwood and Polson (1976), all alternative moves available are evaluated in the same way and the move that looks closest to the goal state is selected as the best route forward.

However, the Water Jug is highly demanding and the human brain does not have the processing capacity of a computer. There is evidence from TOL that alternative move search limits human planning performance in that task (Ward & Allport, 1997). In addition Carder (2004) and Carder, Handley, and Perfect (2004) presented experimental evidence that human planning performance was strongly predicted by the number of alternative move choices there were in TOL, such that performance worsened in problems with a high number of alternative moves. Therefore, the primary aim of the current paper is to determine the extent the alternative moves can predict Water Jug task performance, and explore the limitations of this search in humans.

Atwood and Polson (1976) went on to say that a move is deemed unacceptable if the best route forward is to a position that is less similar perceptually to the goal (i.e. the EVF value increases), because this would be counterintuitive to the perceptual strategy. However it is important to note that these moves are sometimes necessary to solve the task. They argued that a counterintuitive move will only be selected as a possible route forward after other alternatives have been tried, suggesting that these moves would be less acceptable to human participants.

This was explored further in research using frontal patients. There is plenty of evidence that the disk moving tasks are sensitive to prefrontal cortex functioning (e.g. Baker et al., 1996; Dagher, Owen, Boecker, & Brooks, 1999; Morris, Ahmed, Syed, & Toone, 1994; van den Heuvel et al., 2003). Following Goel and Grafman's (1995) argument that frontal patients fail on counterintuitive moves on TOH because of an inhibition deficit, Colvin, Dunbar, and Grafman (2001) predicted that frontal patients would also perform more poorly than normal participants in these moves in the Water Jug task. They argued that frontal patients would fail to inhibit the perceptual strategy and take a move which was actually more efficient, but that led to a position that looked further away from the goal state. Their results appeared to support the argument in that only 52% of frontal patients were able to solve the task in 30 min, compared to 89% of normal controls.

However, on closer inspection the premises that form this conclusion have significant weaknesses. Colvin et al. (2001) measured performance at the point of six different counterintuitive moves, but on only two of these was frontal patients' performance different from normal controls'. When patients were divided into groups by lesion laterality and compared to controls, those with left frontal and bilateral frontal lesions tended to make fewer counterintuitive moves, but again the difference was not significant. With no significant difference between groups on most of the counterintuitive moves, the actual evidence for Colvin et al.'s (2001) position is far from consistent. Furthermore Colvin et al.'s results showed significant differences between groups on moves that were *perceptually consistent* with the goal state. Specifically all patients made more moves that returned them to an earlier position in their solution (looping moves) than normal participants and the left and bilateral frontal groups also made significantly more returns to the start state (in the Water Jug task there is only ever one move that takes the solver forward, and all other moves result in a return to an earlier position or the start state). Hence, a possibly more reliable position is that patients failed to move on to new problem states. Consistent with this position, we have previously shown that older adults (who are said to show frontal like deficits including inhibition failure, West, 1996) showed more uncertainty as a result of increased routes forward than younger adults (Carder, Perfect, & Handley, 2003).

In fact, there are a number of methodological issues with Colvin et al.'s (2001) approach that make it difficult to interpret their findings. One key difficulty was that they had very little experimental control over the task; solution paths varied between participants and they could not control how many times or when a counterintuitive move appeared. In the present series of experiments we adopt a verification paradigm in which participants have to make speeded judgements about the appropriateness of individual Water Jug moves that are demonstrated by the computer. This allows a fine-grain analysis of performance on specific experimenter-chosen points in the solution. Crucially, it also enables the experimental manipulation of different factors such as EVF and the number of alternative moves. In the current research this experimental manipulation will determine the extent that (a) humans use a perceptual strategy to solve Water Jug problems, and (b) the extent that they consider the full range of alternative moves available.

In the two experiments reported here we employ Water Jug problems of just three moves. This was so we could avoid the errorful solutions associated with longer problems (participants can take 25 moves or more to solve a 7-move problem, Atwood & Polson, 1976). Participants witnessed the start state and watched the computer demonstrate a possible first move in the problem, which they had to evaluate as right or wrong. They knew the optimal solution length so they could be accurate. We manipulated the evaluation function value to empirically test whether counterintuitive moves were more difficult than perceptually consistent moves. To illustrate this manipulation, various 3-move problems can be excerpted from the problem given in Appendix A which the reader can easily visual in pictorial form. If the start state is 5/0/3 and the goal is 2/5/1, the best first move is to 5/3/0, which results in a state more similar perceptually to the goal (than 5/0/3) and a corresponding drop in EVF . However, if the start state is 8/0/0 and the goal is 6/2/0, the best first move is to 3/5/0 which is less similar to the goal state (than 8/0/0) and this causes EVF to increase so the move is counterintuitive. We also manipulated the number of alternative moves there are to choose between (which can be understood in Appendix A by the number of possible routes from a given position).

In summary in Experiment 1 we presented the Water Jug task via a verification paradigm in which participants watched the computer demonstrate a move, and had to make speeded judgements about whether it was correct to solve the problem in three moves. Half of the problems presented a counterintuitive move in which the EVF increased and half presented a move that was perceptually consistent with the goal state, as measured by a drop in EVF. In half of the problems there were two or three alternative moves to choose between, and in half of the problems there were four.

2. Experiment 1

2.1. Method

2.1.1. Participants

Twenty-three undergraduate participants from the University of Plymouth participated for a cash payment.

2.1.2. Procedure

Participants completed the task in a testing session that lasted about 30 min. They were issued 10 practice problems and then 32 experimental problems that were administered in a unique random order to each participant. What follows is a full description of the task.

2.2. Water Jug task

A verification version of the Water Jug task required that participants watched the computer demonstrate the first move in a series of problems, and they had to judge if this was right or wrong to solve it in the most efficient optimal solution.

Participants were introduced to the screen layout, which in pictorial form illustrated liquid distributed between jugs, as in Fig. 1. A goal state was given in the upper portion of the screen, and the starting position in the lower portion. Written instructions informed participants that each problem could be solved in three moves (including the move demonstrated). The pouring rule was also explained to them; i.e. when water was poured, it must continue until the poured-from jug is empty or the target-jug is full.

The computer displayed the starting position for 5000 ms and then performed a single move which took 2500 ms to complete. Immediately after the computer asked the participant to judge whether it was correct or not by clicking an appropriate on-screen button. Participants were told that their responses were being timed and that they should work as quickly as possible, while avoiding errors.

For every problem there is only one optimal move and two or three incorrect moves. (n.b. Atwood & Polson, 1976; define an optimal move as a move with the biggest drop in the evaluation function, but here the optimal move is part of the most efficient solution). Of the 32 problems administered, half demonstrated a correct move and half an incorrect move. In optimal moves, if the EVF reduced it was "perceptually consistent" and if it increased it was "counterintuitive". If there were three possible moves to choose between the alternative move factor was "low" and if there were four it was "high". Suboptimal moves also presented. These were matched to optimal moves in that the same start and goal states were presented but the move demonstrated was not optimal for the solution. However, the analysis focused on performance on optimal trials because a sub-optimal trial's EVF is different to its matched optimal trial. Consequently the factors manipulated in suboptimal trials do not correspond to the features of correctly demonstrated trials.

The problems comprised 3-move problems in which the maximum capacity of the jugs were 8,5,3 or 12,7,4 or 14,9,5 or 14,9,3 or 16,10,3 or 16,7,4. The problem spaces of these problems were mapped geometrically using barycentric coordinates onto their triangular-shaped problem spaces (e.g. Bogomolny, 2000) as illustrated in Appendix A. This provided a graphical representation of all legal capacity states, and all the possible transition routes between individual states. Of the problems selected, none of the solution paths used the same pour, so no problem state was presented more than once. To illustrate problem examples, the problems taken from the 8,5,3 apparatus are mapped onto the barycentric coordinates and are illustrated in Appendix A. For example a low/low problem we employed is 5/0/3 to 2/5/1: The move demonstrated was the first (to 5/3/0) (subsequent correct moves are to 2/3/3 and then 2/5/1).

The computer recorded the number of errors and latency (in milliseconds) of participants' judgements in each of the 32 trials.

2.2.1. Apparatus

All the tasks were presented on RM PCs with a 17″ screen. The resolution was set at 800×600 high colour (16 bit), and the task was presented via Windows XP. The participants interacted with the computer using a computer mouse.

2.2.2. Design

There were two kinds of Move Demonstrated (optimal/suboptimal) and two levels of the EVF factor (consistent with a perceptual strategy/counterintuitive). There were also two levels of the alternative moves factor (low, which had two or three legal moves/high where there were four legal moves). There were two dependent variables: errors and the median latency for correct trials in milliseconds.

2.3. Results

In the latency data, only trials the participant answered correctly were analysed. Consequently 6.9% (or 5 values) of the latency data were missing. These data were replaced by a regression calculation in which performance on other trials was used to predict the missing values. Data treatment did not alter the pattern of results that were observed. The raw data showed some positive skewing from a normal distribution and large standard deviations; hence we calculated the median values and the data were transformed using a logarithmic transformation for the analysis that follows. Descriptive statistics are given in Table 1 for the latency data and errors in Experiment 1.

A 2 (perceptually consistent/counterintuitive EVF) × 2 (low/high alternative moves) within-subjects ANOVA on the latency data for correctly judged decisions on problems that demonstrated optimal moves showed the EVF value effected participant's decision latencies, $F(1,22) = 4.8$, $MSE = 0.1$, $p < .05$, $\eta^2 = .178$. Moves that were perceptually consistent with the goal state were judged more quickly than those that were counterintuitive. There was an effect of alternative moves, $F(1,22) = 9.4$, $MSE = 0.1$, $p < .01$, $\eta^2 = .2099$, which showed performance was effected by the number of alternative moves there were to choose between, with more possible alternatives resulting in longer response latencies than those solutions with fewer moves to choose between.

The interaction between the two factors was marginally significant, $F(1,22) = 3.8$, $MSE = .179$, $p = .063$, $\eta^2 = .148$. This interaction is shown in Fig. 2 and suggests that the effect of alternatives was present only when counterintuitive moves were demonstrated. Newman-Keuls' post hoc tests revealed that the effect of alternatives in counterintuitive moves was significant ($p < .05$). However, in perceptually consistent trials it was not present ($p = .838$). The post hoc analysis revealed that participants spent more time judging counterintuitive moves with high alternatives than any other type move problem ($p < .02$). None of the other differences were significant ($p > .05$). This suggested that participants spent longer judging moves that were not consistent with their problem solving strategy, particularly when there were lots of other alternative moves to consider.

A 2 (perceptually consistent/counterintuitive EVF) × 2 (low/high alternative moves) within-subjects ANOVA on the error data showed no effect of EVF, $F(1,22) = 2.4$, $MSE = 0.9$, $p = .134$, $\eta^2 = .099$. The effect of alternative moves was marginally significant, $F(1,22) = 3.2$, $MSE = 1.1$, $p < .1$, $\eta^2 = .128$, but there was a significant interaction between the two factors $F(1,22) = 11.1$, $MSE = 1.1$, $p < .01$, $\eta^2 = .335$, which is illustrated in Fig. 3.

Table 1
Descriptive statistics showing the mean decision latency in milliseconds for correctly judged moves and errors (out of 4) in optimal trials in Experiment 1

	Low alternatives	High alternatives
Perceptually consistent EVF	12778 (9348) 1.2 (1.0)	12623 (7244) 0.9 (1.3)
Counterintuitive EVF	12968 (9126) 0.8 (1.2)	17999 (11617) 1.9 (1.2)

*SD*s are given in brackets. $N = 23$.

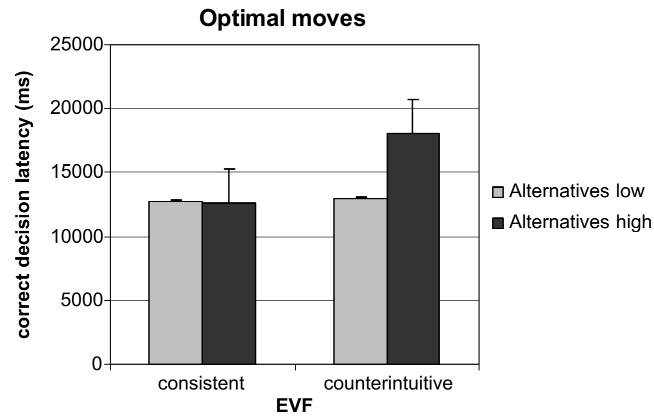


Fig. 2. Experiment 1: The effects of alternative moves and EVF on the response latency for correct decisions. Error bars show the standard error ($n = 23$).

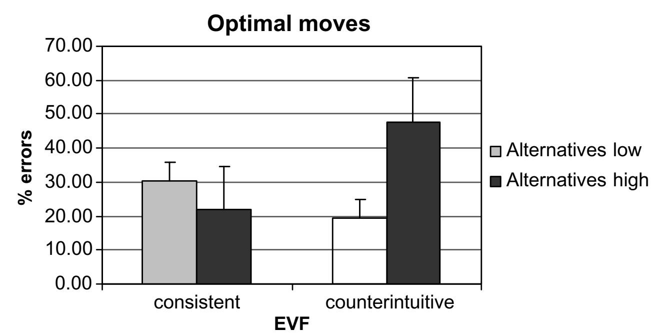


Fig. 3. Experiment 1: The effects of Move demonstrated, alternative moves and EVF on percentage error rates. Error bars show the standard error. ($n = 23$).

Newman-Keuls' post hoc tests on the interaction revealed that while the effect of alternatives in counterintuitive moves was significant ($p < .01$), it was not observed in perceptually consistent moves ($p = .279$). The difference between counterintuitive, high alternative moves and perceptually consistent, high alternative moves was also significant, ($p < .01$), as was the difference with perceptually consistent, low alternative moves ($p < .05$) but no other differences were observed ($p > .05$). This showed that when there were lots of alternative moves to choose from participants were more likely to incorrectly reject a move if the move demonstrated was counterintuitive, but would accept moves that were consistent with a perceptual strategy even if there were lots of other moves available to choose from.

2.4. Discussion

In summary, Experiment 1 shows that the number of alternative moves there are to choose between predicts performance when the best move is counterintuitive, such that when there are fewer moves to choose between, responses are quicker. There was also an effect of EVF whereby Participants took longer and were more likely to reject a counterintuitive move that was correct than reject a perceptually consistent move that was correct. This EVF effect is consistent with Atwood and Polson's (1976) claim that participants who are unfamiliar with the Water Jug task solve the problem by taking moves that will move them perceptually closer to the goal state in a hill-climbing strategy, and with Colvin et al.'s (2001) observations that counterintuitive moves are more difficult. The finding that the number of alternatives impacts performance on counterintuitive but not perceptually consistent moves partly supports the claim that human participants engage in a breadth-based analysis of the problem. Although a breadth search of the problem space does occur, Atwood and Polson (1976) argued that all alternative moves are considered and that the move with the biggest drop in evaluation function is preferred. This data shows quite clearly that two moves that are of equal effectiveness in reaching the goal state vary in the extent that other alternative moves are considered. When the move is counterintuitive, alternative moves are considered however when the move is perceptually consistent they are not. In this case no additional alternative move search occurs.

Furthermore the error rates show clearly that participants do not always reject counterintuitive moves if there is a perceptually consistent move available as Atwood and Polson (1976) suggested. Rather they will accept a counterintuitive move on its first presentation, but they are more uncertain of it and take longer to evaluate it. This suggests that there are fundamental differences between computer and human problem solving strategies.

According to Atwood and Polson's (1976) problem solving theory, human participants should consider all alternative moves because of their mere presence. Therefore, the presence of extra additional routes forward should slow problem solving, even if these possible routes are completely fruitless. An alternative possibility is that participants can quickly screen out irrelevant alternative moves to avoid over burdening their already stretched cognitive resources unnecessarily. In Experiment 2 we therefore added an additional jug to the problems, giving more perceptual alternatives. However, two of the jugs were the same size and we told participations they could only use either one of these at any time. Thus the additional jug was logically redundant and did not offer any genuine alternative routes forward although the mere presence and number of alternatives was increased. We compared performance on the 4-jug version with that on matched 3-jug problems. We might expect that if the presence of alternatives engages any automatic processing then the fourth jug will result in longer latencies and more errors. Alternatively, if participants consciously reflect on possible alternatives they will recognise that fourth jug is redundant and be unaffected by it. In order to determine the number of participants required to demonstrate an effect of Jugs a power analysis was conducted. In Experiment 1 we showed that increasing the number of alternative moves by one (or two in some problems) attributed almost 30% of the variance in performance to this manipulation. We hypothesised that if Atwood and Polson were correct, then the four peg apparatus (which increases in the number of alternative moves to a maximum of six possible moves) would show a similar sized effect. We determined that a sample of 30 participants would secure a 90% probability of eliciting the effect if it existed, using Campbell and Thompson's (2002) power calculator.

3. Experiment 2

3.1. Method

3.1.1. Participants

Thirty participants from the University of Plymouth participated for a cash payment. None had participated in Experiment 1.

3.1.2. Procedure

Participants completed the task is a testing session that lasted about 45 min. They were issued 10 practice problems and then 32 experimental problems from either the 3- or 4-jug problems. These were administered in a unique random order to each participant. The order of 3- and 4-peg apparatus was completely counterbalanced across participants. Participants were then issued practice and experimental problems from the apparatus that they had not completed in the first round. The 3-jug task is described in Experiment 1, with the following modifications for the 4-jug problems.

3.2. 4-Jug Water Jug task

Participants were informed that there were two jugs of equal size and four jugs in total. They were informed that they could use all jugs in the solution as long as one of the equal sized jugs was empty in every move. The vessel size that was duplicated was counterbalanced between the three jugs in the problem. The jug that that the computer used in the solution was counterbalanced between the left and right hand of the two.

3.2.1. Design

There were two sizes of apparatus (3-jug or 4-jug) × two move demonstrated (optimal/suboptimal). Only optimal trials were used in the analysis that follows, as in Experiment 1. As in Experiment 1 there were two levels of the EVF factor (consistent with a perceptual strategy, and counterintuitive) and two levels of the alternative moves factor (low, and high). There were two dependent variables: errors and the median latency for correct trials in milliseconds.

3.3. Results

For the latency data only correct responses were analysed. Consequently 4.7% (or 9 values) in the latency data were missing and were replaced as described in Experiment

1. Data treatment did not alter the pattern of results that were observed. The raw data showed some positive skewing from a normal distribution and large standard deviations. Therefore, they were transformed using a logarithmic transformation for the analysis that follows. Descriptive statistics are given in Table 2 for latency and errors in Experiment 2.

A 2 (3-jug/4-jug) × 2 (perceptually consistent/counterintuitive EVF) × 2 (low/high alternative moves) within-subjects ANOVA on the latency data for correctly judged decisions showed that there was no effect of the number of Jugs, $F(1,29) = 1.5$, MSE = 0.6, $p = .237$, $\eta^2 = .048$, and the number of jugs did not interact with any other factor, Jug × EVF, $F(1,29) = 0.2$, MSE = 0.3, $p = .905$, $\eta^2 = .001$, Jug × alternative moves, $F(1,29) = 0.1$, MSE = 0.3, $p = 745$, $\eta^2 = .004$. Power analysis suggested that this was not due to a small sample size.

The effect of EVF observed in Experiment 1 was replicated, $F(1,29) = 16.9$, MSE = 0.3, $p < .001$, $\eta^2 = .369$. This can be seen in Fig. 4 and showed that counterintuitive moves took longer than moves that moved perceptually closer to the goal state.

Fig. 4 also illustrates an effect of alternative moves, as replicated from Experiment 1, $F(1,29) = 11.3$, MSE = 0.2, $p < .01$, $\eta^2 = .281$. This showed that increasing the number of alternative moves there were to choose between increased response latency. None of the remaining interactions were significant. EVF × alternative moves, $F(1,29) = 0.6$, MSE = 0.3, $p = .484$, $\eta^2 = .019$, Jugs X EVF × alternative moves, $F(1,29) = 0.1$, MSE = 0.2, $p = .956$, $\eta^2 = .000$.

A 2 (Perceptually consistent/counterintuitive EVF) × 2 (low/high alternative moves) within-subjects ANOVA on the error data where participants had rejected a move that was in fact correct, revealed no main effect of Jug, $F(1,29) = 1.3$, MSE = 1.3, $p = .260$, $\eta^2 = .044$. Neither were was there an effect of EVF, $F(1,29) = 0.1$, MSE = 1.0, $p = .704$, $\eta^2 = .005$, or alternative moves, $F(1,29) = 1.5$, MSE = 1.1, $p = .233$, $\eta^2 = .049$. However,

Table 2
Descriptive statistics showing mean decision latency in milliseconds for correctly judge moves and errors (maximum 4) in optimal trials in Experiment 2

	Low alternatives	High alternatives
3-jug Problems		
Perceptually consistent EVF	11718 (9538) 1.2 (0.8)	10954 (6659) 1.0 (1.1)
Counterintuitive EVF	13017 (9369) 0.8 (1.1)	16233 (8627) 1.2 (1.1)
MEAN	**12363 (9454) 1.0 (0.9)**	**13594 (7643) 1.1 (1.1)**
4-jug Problems		
Perceptually consistent EVF	12068 (8174) 1.3 (1.2)	11384 (5888) 0.9 (1.2)
Counterintuitive EVF	15484 (13941) 1.0 (1.2)	18490 (13660) 1.7 (1.3)
MEAN	**13776 (11058) 1.2 (1.2)**	**14937 (9774) 1.3 (1.3)**

Bold indicates the summary descriptive statistics collapsed across move type (perceptually consistent/counterintuitive) for both 3 and 4 jug problems. *SD*s are given in brackets. $N = 24$.

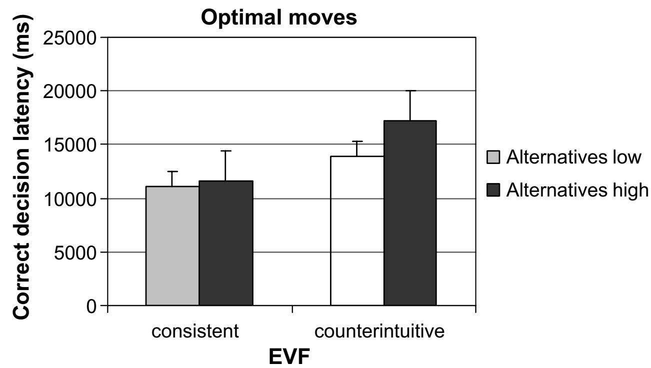


Fig. 4. Experiment 2: The effects of Move Demonstrated, alternative moves and EVF on the response latency for correct decisions. Error bars show the standard error ($n = 18$).

there was an interaction between EVF and alternatives, $F(1,29) = 6.4$, MSE = 1.9, $p < .05$, $\eta^2 = .181$, which is illustrated in Fig. 5.

Newman-Keuls' Post hoc tests on the interaction showed that the effect of alternatives on counterintuitive judgements was marginally significant, $p < .1$, but none of the other differences were significant ($p > .05$). None of the remaining interactions were significant; Jug X EVF, $F(1,29) = 2.1$, MSE = 1.0, $p = .161$, $\eta^2 = .067$, Jugs × Alternatives, $F(1,29) = 0.1$, MSE = 0.6, $p = 719$, $\eta^2 = .005$, Jugs × EVF × Alternatives, $F(1,29) = 0.8$ MSE = 1.0, $p = .373$, $\eta^2 = .027$.

3.4. Discussion

The aim of Experiment 2 was to examine the extent to which the mere presence of a greater number of alternatives would impact on performance, irrespective of the relevance of those alternatives. The findings of Experiment 2 showed that the perceptual presence of redundant alternative moves had no discernable effect on performance. This supports the idea that participants actively consider alternative move choices as required and can rapidly screen out redundant possibilities in the perceptual array. The main effects of EVF and alternative moves remained in the latency data,

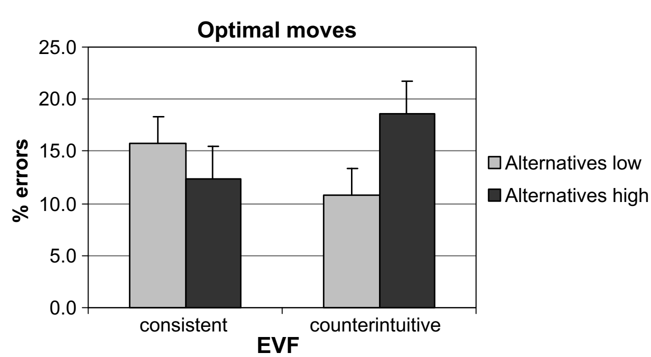


Fig. 5. Experiment 2: the effects of alternative moves and EVF on percentage of errors. Error Bars show the standard error ($n = 24$).

and replicated those effects demonstrated in Experiment 1. Whilst the interaction between EVF and alternative moves was not present in the latency data, it was present in the error data, suggesting as in Experiment 1 that the effect of alternatives is greater when a presented move is counterintuitive. It is possible that the interaction was no longer present because participants had an even greater burden placed on their cognitive resources with the extra jug and the associated rule, so they further pruned the amount of processing they engaged in.

4. General discussion

The primary goal of the present research was to determine the extent that performance on the Water Jug task was predicted by the perceptually consistency of a move as suggested by Colvin et al. (2001), and by the number of alternative moves that there were to choose between. This followed evidence that the number of alternative moves was a key predictor on a different executive task, TOL (Carder et al., 2004). We wished to ascertain if this finding could be generalised to other problem solving tasks. The evidence for this position was evaluated via a verification paradigm in which participants made speeded judgements about the effectiveness of a demonstrated move. We manipulated the alternative move demands and evaluation function (EVF) of demonstrated moves to achieve these goals, and obtained main effect of these factors in both experiments. The findings from the latency data in Experiment 1 and the error data in Experiments 1 and 2 demonstrated that these factors were interactive, such that additional alternative moves had an impact in counterintuitive moves, but not when moves led to a position that was perceptually more similar to the goal state.

Our results support some aspects of Atwood and Polson's (1976) theory about human problem solving in Water Jug problems. They argued that normal participants used a hill climbing strategy to solve the Water Jug task in which all possible moves forward from the current position were considered, and the move that was preferred was that which took the solver closest to the goal state in perceptual terms. Our evidence suggested that participants took longer to make decisions when there were more alternative moves to choose between, supporting their position. However, we have shown that the mere presence of alternative moves does not automatically evoke an evaluation as Atwood and Polson (1976) suggested. Rather planners seem to be able to screen out redundant routes forward. Furthermore they can more actively consider chosen routes when desired, such as when move presented is correct but counterintuitive. Atwood and Polson (1976) also argued that counterintuitive moves were selected as a possible new route forward only after other alternative moves had been tried. Our evidence showed that a counterintuitive move was accepted on its first presentation, but participants considered it for longer than they would a perceptually consistent move.

According to Atwood and Polson's model, forward planning was not necessary to solve the task. Whilst the task can be solved by trial and error, accuracy in our participants was good and we would argue that this was because participants used forward planning strategies and evaluated alternative moves in the context of the goal. This suggests that participants did not just engage a hill climbing strategy, but also explored other strategies. Although our use of easier problems that were within the capacity of human processing resources may have made forward planning a viable strategy, the findings nevertheless suggest that humans can be more flexible in their problem solving approach than Atwood and Polson's suggested.

The results highlight one limitation of the information processing approach to understanding human cognition and so assist our understanding of the brain-behaviour relationship. Whilst process models can simulate human performance well, they might sometimes use different processes. For example, a computer will use the same problem solving process on every occurrence regardless of how easy or hard a problem is, but here our participants seem to have varied their strategy according to the task's demands; they use forward planning and a hill climbing strategy. They also seem to have made conscious strategy adjustments, such as choosing whether or not to search alternative solutions.

Colvin et al. (2001) argued that counterintuitive moves were particularly difficult for frontal patients, and they argued that an inhibition deficit was the source of this difficulty. On closer inspection Colvin et al.'s data provided only weak evidence for this claim. Recall that patients only performed significantly differently from normal controls on two of the six counterintuitive moves. However, there was a more general deficit amongst the patient group in which they performed more poorly on all move types including perceptually consistent moves.

In previous research with TOL we argued that the search of alternative moves was related to the inhibitory demands of that task (Carder et al., 2004). This seems logical since both tasks require choosing between alternative responses. We have certainly presented evidence to support the fact that the consideration of alternative moves is served by conscious and reflective cognitive processes. Our evidence that the number of alternatives impacts performance on counterintuitive moves is a novel and interesting finding. It is quite possible that this task feature elicits the deficit observed amongst patient groups and might provide useful subject matter for further research.

The findings of these experiments are also interesting on a more general level. They suggest that when a presented option is inconsistent with the application of a simple strategy, as is the case with a counterintuitive move, participants abandon this strategy in favour of a more effortful and demanding explicit consideration of alternatives. That is when a move is inconsistent with a simple perceptual or intuitive strategy; additional effort is engaged on the task in the form of alternative search. This is consistent with a range of findings in the field of motivated cognition whereby arguments are evaluated at greater depth (see, for example, Kunda, 1990), or flaws in experimental demands are identified (see for example, Lord, Ross, & Lepper, 1979) when a given conclusion is incongruent with an individual's belief. These findings show that additional effort is engaged when an argument cannot be evaluated or rated simply based upon its congruence with beliefs. This effect is observed in other research into higher level cognition, most notably in the field of deductive reasoning. In these paradigms conclusions are given that are either valid or invalid and believable or unbelievable. The typical finding is an interaction between belief and logic, such that the effect of beliefs is only apparent on invalid arguments (e.g. Evans, Barston, & Pollard, 1983; Goel & Dolan, 2003). A typical claim is that when presented with unbelievable conclusions participants evaluate their veracity by searching for alternative representations of the premises in which the conclusion does not hold. In the case of a valid conclusion there are no such models. In contrast, in the case of invalid conclusions there are by definition alternative models in which the conclusion does not hold. When the conclusion is believable no search for alternatives is carried out and the conclusion is accepted resulting in the interaction observed. These effects parallel our finding that the presentation of a counterintuitive move instigates an effortful search for alternative possibilities.

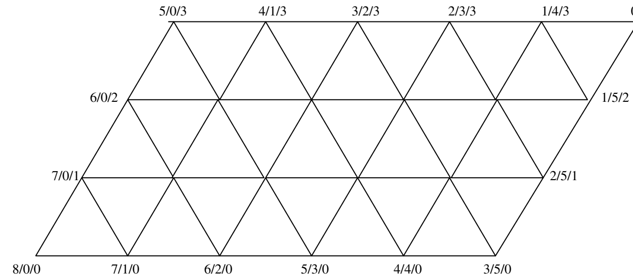
In summary the present research has clearly demonstrated that the Water Jug task shares a number of common features with other MOVE problems. Specifically the difficulty of the Water Jug task resides in two specific characteristics, the presence of moves that appear to take the problem solver away from the goal and the fact that there are multiple possible alternative moves at every stage within a given problem space. Through the use of a verification paradigm the present work presents a more structured method for manipulating and experimentally evaluating the relationship between the structural characteristics of the task and problem difficulty. In the future this method may be usefully applied in investigating the specific aspect of the task that contributes to the difficulty experienced by frontal patients on executive tasks of this kind.

Acknowledgments

We thank Simon Paice for writing the programs and Ian Dennis for help with the problem space. We thank our anonymous reviewers for their comments on an earlier draft of this paper.

Appendix A

The legal problem space of the complete 8/5/3 problem mapped onto barycentric co-ordinates. The solution can be found by tracing the route from 8/0/0 to 4/4/0. In order that the pouring rules are adhered to lines must be followed to the edges of the parallelogram. Hence the shortest route is 8/8/0, 3/5/0, 3/2/3, 6/2/0, 6/0/2, 1/5/2, 1/4/3, 4/4/0. Solution 2 is: 8/8/0, 5/0/3, 5/3/0, 2/3/3, 2/5/1, 7/0/1, 7/1/0, 4/1/3, 4/4/0.



Appendix B. The evaluation function (EVF) formulae

$$EVF_i = [C_i(A) - G(A)] + [C_i(B) - G(B)]$$

C_i is the number of units of water in the jug for state i. G is the goal state of a denoted jug. A = Jug A, B = Jug B. Thus the evaluation function (EVF) is difference between the current and goal contents of the largest and middle sized jugs, summed.

Example: Fig. 1: first two moves of 8/0/0 to 4/4/0.

8/0/0 : EVFi=8=[8 – 4] + [0 – 4]

EVF = reduces = perceptually consistent

3/5/0 : EVFi=2=[3 – 4] + [5 – 4]

EVF = increases = counterintuitive

3/2/3 : EVFi=3=[3 – 4] + [2 – 4]

References

Atwood, M. E., & Polson, P. G. (1976). A process model for Water Jug problems. *Cognitive Psychology, 8*, 191–216.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.). *The psychology of learning and motivation: Advances in theory and research* (Vol. 8, pp. 47–89). New York: Academic Press.

Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S., et al. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia, 34*, 515–526.

Bogomolny, A. (2000). Barycentric Coordinates. http://www.cut-the-knot.org/triangle/barycenter.shtml.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove, UK: Psychology Press.

Campbell, J. I. D., & Thompson, V. A. (2002). More power to you: Simple power calculations for treatment effects with one degree of freedom. *Behavior Research Methods, Instruments, and Computers, 34*, 332–337.

Carder, H.P. (2004) Deconstructing the Tower of London: Resolution, alternatives and executive function. Unpublished Thesis.

Carder, H.P., Perfect, T.J. & Handley, S.J. (2003). Ageing and Tower of London: Alternative moves, resolution gap and executive function as predictors of task performance. Poster presented at the BPS workshop on the Neuropsychology of Ageing, Holland House, UK.

Carder, H. P., Handley, S. J., & Perfect, T. J. (2004). Deconstructing the Tower of London: Alternative moves and conflict resolution as predictors of task performance. *The Quarterly Journal of Experimental Psychology, 57a*(8), 1459–1483.

Colvin, M. K., Dunbar, K., & Grafman, J. (2001). The effects of frontal lobe lesions on goal achievement in the Water Jug task. *Journal of Cognitive Neuroscience, 13*, 1129–1147.

Dagher, A., Owen, A. M., Boecker, H., & Brooks, D. J. (1999). Mapping the neural network for planning: A correlational PET activation study with the Tower of London task. *Brain, 122*, 1973–1987.

Ernst, G. W., & Newell, A. (1969). GPS: A case study in generality and problem solving. New York: Academic Press.

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition, 11*, 295–306.

Gilhooly, K. J., Wynn, V., Phillips, L. H., Logie, R. H., & Della Sala, S. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. *Thinking and Reasoning, 8*, 165–178.

Goel, V. (2002). *Planning: Neural and psychological in cognitive & neural basis of planning*. *Encyclopedia of cognitive science*. Macmillan, pp 697–703.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition., 87*, B11–B22.

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia, 33*, 623–642.

Handley, S. J., Capon, A., Copp, C., & Harper, C. (2002). reasoning and the Tower of Hanoi: The role of spatial and verbal working memory. *British Journal of Psychology, 93*, 501–518.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effectsof prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098–2109.

Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs, 54*(No. 248.).

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Morris, R. G., Ahmed, S., Syed, G. M., & Toone, B. K. (1994). Neural correlates of planning ability: Frontal lobe activation during the Tower of London test. *Neuropsycholgia, 31*, 1367–1378.

Morris, R. G., Miotto, E. C., Feigenbaum, J. D., Bullock, P., & Polkey, C. E. (1997). The effect of goal subgoal conflict on planning ability after frontal and temporal lobe lesions in humans. *Neuropsychologia, 8*, 1147–1157.

Morris, R. G., & Ward, G. (2005). *The Cognitive Psychology of Planning*. Hove, UK: Psychology Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, New Jersey: Prentice-Hall.

Norman, D.A. and Shallice, T. (1980/1986). Attention to action: Willed and automatic control of behaviour. Centre for Human Information Processing (Technical Report #99). Reprinted in revised form in Davidson, R.J., Schwartz, G.E., and Shapiro, D. (Eds.) (1986), Consciousness and Self-Regulation (Volume 4), New York: Plenum.

Owen, A. M., Downes, J. D., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia, 28*, 1021–1034.

Owen, A. M., Doyon, J., Petrides, M., & Evans, A. C. (1996). Planning and spatial working memory: A positron emission tomography study in humans. *European Journal of Neuroscience, 8*, 353–364.

Phillips, L. H., Gilhooly, K. J., Logie, R. H., Della-Sala, S., & Wynn, V. E. (2003). Age, working memory, and the Tower of London task. *European Journal of Cognitive Psychology, 15*, 291–312.

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory, 7*, 209–231.

Shallice, T. (1982). Specific impairments of planning. *Philosophical transactions of the Royal Society of London, B., 298*, 199–209.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

van den Heuvel, O. A., Groenewegen, H. J., Barkhof, F., Lazeron, R. H. C., van Dyck, R., & Deltman, D. J. (2003). Frontostriatal system in planning complexity: A parametric functional magnetic resonance version of Tower of London task. *NeuroImage, 18*, 367–374.

Ward, G., & Allport, A. (1997). Planning and problem solving using the five-disc Tower of London task. *The Quarterly Journal of Experimental Psychology, 50a*, 49–78.

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and Cognition, 41*, 231–242.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive ageing. *Psychological Bulletin, 120*, 272–292.