Dear Lt Col Dr Carl Kutsche,

Enclosed, please find the final report of Contract 01WE039 “Enhancing Decision Perfromance” that I conducted with my research team: Dr. Wendy Smith, Helen Mander, Lorraine Murray, and Merrie Bonner.

Although all the 3 studies shared similar experimental paradigms, the report is written up as 3 separate studies, each presented as an independent and coherent mini-report. We hope in the future to expand on this work and to submit each study for publication in a scientific journal.

Please let me know if I can be of any further assistance.

Dr Itiel Dror
Senior Lecturer
E-mail: id@ecs.soton.ac.uk

CONTRACT 01WE039 “ENHANCING DECISION PERFROMANCE” -- FINAL REPORT:

Study 1: Flexiablity in Performance

Study 2: Use and Memory of Configural and Holistic Information

Study 3: Time Pressure

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
09-07-2002	Final Report	17 July 2001 - 17-Jul-02

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
Enhancing Decision Performance	F61775-01-WE039
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Dr. Itiel Dror	**5d. TASK NUMBER**
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
University of Southampton Shackleton Building, No. 44 Southampton SO17 1 BJ United Kingdom	N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
EOARD PSC 802 BOX 14 FPO 09499-0014	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** SPC 01-4039

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report results from a contract tasking University of Southampton as follows: The contractor will investigate the link between knowledge acquistion (via training) and usability (via testing), as well as between the stage of planning and decision-making and the stage in which plans & decisions are executed. Experiments will investigage how people internalize and represent information, and how this influences their subsequent ability to use that information. Results will focus on the ability to generalize beyond examples used during training, the ability to use and incorporate new information, the capacity to be flexible, and the ability to develop creative solutions dictated by changing task demands under high infornmation load and time pressure..

15. SUBJECT TERMS

EOARD, Human Factors, Psychology, Cognition

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18, NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Valerie Martindale, Maj, USAF
UNCLAS	UNCLAS	UNCLAS	UL	40	**19b. TELEPHONE NUMBER** *(Include area code)* +44 (0)20 7514 4437

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

STUDY 1: FLEXIABLITY IN PERFORMANCE

Abstract:
Previous research suggested that training would improve performance, but also induce functional fixedness, therefore decrease creativity in subsequent related tasks. This experiment was designed to identify training procedures that would maximise performance, and minimise reductions of creativity. Creativity is assessed by application of novel solutions to a problem-solving task. Evidence is compared for the alternative hypotheses that functional fixedness is due to interference; in the generation of novel solutions, or inhibiting the application of these solutions. This study explored the effects of specific (one suggested solution) and non-specific (several suggested solutions) training on later performance in isomorphic and homomorphic tangram puzzles. Results showed that effects of functional fixedness did not influence performance on homomorphic tasks. Performance on isomorphic tasks was inhibited after training, especially specific training. Findings of this research are that functional fixedness reduces creativity-dependent performance, only when stimuli in training and at test are similar. Observations suggest that functional fixedness is due to the over-application of previously learnt methods, as a result of inadequate training-task stimuli comparisons. Also, an unexpected finding is that functional fixedness may encourage generalisation of peripheral, and not central features of training stimuli, when inter-stimuli similarity direct attention towards these features.

Introduction:
Training is an activity in which participation will influence performance on subsequent tasks. This process of applying skills to another task may also be known as transfer. The value of training or transfer may be quantified by the extent of influence on later task performance. Research into the effects of training has been directed towards using training to efficiently achieve the greatest change in performance.

Early studies concerning the value of training included the theory of identical elements, first proposed by Thorndike and Woodworth (1901, cited in Kahney, 1993). The authors describe how the transfer of skills from one task to another is positively correlated to the similarity between the tasks. The influence of one task to another is reduced as the two tasks become increasingly dissimilar (Osgood, 1949, cited in Kahney, 1993) so absence of any improvement in task performance after training is an indication that there is no relationship between two tasks. Differences in problem solving strategies used by novices and experts have also been identified. For example, Simon and Chase (1973) showed that experts and novices differ in the level of processing they apply to a task. Experts search for a solution for a task by looking at the underlying structure, whereas novices attend primarily to the surface components of the task (Marsh, Lindau and Hicks, 1996).

The effects of training are usually observed when subsequent performance is enhanced. However, negative transfer is the term used to describe inhibition in task performance, due to the prior presentation of a similar task (Singley and Anderson, 1989). One type of negative transfer is functional fixedness. This relates to the inhibited ability to find novel uses for a particular object as a result of previous experience.

Functional fixedness was experimentally created by Dunker (1945, cited in Hampson and Morris, 1997) who presented a problem where solution depended on using a feature of an object, unrelated to the conventional use. In one example, the successful solving strategy involved using the weight of a wrench, instead of the conventional tool use of the object. Poor performance in this problem was attributed to the interference of prior knowledge regarding the function of the object on the ability to develop a novel use for the object. Functional fixedness can be related to recently learned, as well as well established concepts.

Luchins (1942, cited in Kahney, 1993) trained participants to complete five logical puzzles, with a common method of solution. When faced with similar problems, trained participants continued to use the method acquired in training and were consequently less likely than a control group to generate a more economic method of problem solving. This experiment shows how recent experience with one successful method of problem solving can increase difficulty in developing solutions based solely on the requirements of new tasks.

Ward (1994) studied interference of functional fixedness in creativity. Participants were required to invent and draw an imaginary creature that may live in outer space. The creatures produced bore significant similarity to animals found on earth, notably in terms of symmetry, existence and design of sense organs and modes of transportation. Ward explained that these creations were influenced by the category in which they are formed, in this case, animals. The subjective creativity was the result of merging components selected from different exemplars. Further research showed that this ‘conformity effect’ increases relative to the number of examples seen (Ward, 1994).

The above research shows that functional fixedness, a result of experience with related examples, can reduce ability to find solutions to novel problems, which is an important component of creativity. The effects of functional fixedness are a potential problem when training people to become proficient in one of many fields, such as medicine, engineering etc. as experience is necessary, but creativity and flexibility is also desirable. Is release from the detrimental effects of functional fixedness possible or is reduction of creativity an inevitable effect of training? Glucksberg and Danks (1968) showed that it is possible to reduce the inhibitory effects of functional fixedness on performance in a well-defined task. The task involved using the electrical conductive properties of a part of a familiar object to solve a puzzle. Glucksberg and Danks (1968) considered that inability to solve this problem was due to previous experience using the whole object therefore inability to mentally separate components of the object. When labels, even nonsense labels were applied to different parts of the object, the component required in the solution was easier to access. Glucksberg and Danks (1968) reasoned that the labels highlighted the fact that the item could be reduced into different components, hence, properties of the components were implied. This shows that the effects of functional fixedness can be reduced in relation to a specific task. This experiment may not be relevant to general creativity related tasks. Solution of the task used by Glucksberg and Danks (1968) was dependent on focussing on the properties of components of the objects: therefore, it had a specific goal, to which the authors found a clue, whereas when trying to increase creativity, there is not one specific goal that can be encouraged. However, this is an important piece of research as it shows that once the reason that solutions are blocked is identified, steps can be taken to eliminate this. In the same way as Glucksberg and Danks’ (1968) identified and reduced the problem; seeing the stimulus as a ‘whole’ object, discovery of why previous experience decreases creativity, can lead to avoidance of this. The aim of this study is to first discover exactly what mechanisms underlie functional fixedness. Once this has been achieved, training can be adapted to avoid inhibiting the creative potential.

Previous studies have shown that functional fixedness is due to trying to apply a previously learnt method to an inappropriate task. However, specifically why, in the presence of a new task, does experience make this inappropriate method more likely to be selected, compared to generation and application of other methods has not been ascertained. One possibility is that functional fixedness is the result of blocking the application of new ideas with those already present in memory. Smith, Ward and Schumacher (1993) suggest that when examples are seen, they are stored in memory. New tasks, judged to be sufficiently similar to the training task, trigger the retrieval of these prior examples. Therefore, memory serves as a schema or guideline, reducing cognitive effort. Once guidelines for the means to complete a task have been experienced, these ideas are transferred to a similar new task before attempts to apply a new, internally generated problem solving strategy are made. This hypothesis is supported by the work of Birch and Rabinowitz (1951) who showed that functional fixedness is more likely to inhibit performance when participants are shown one use for the object, but less likely when the participant is given no guidance. Thus, when there are no other examples of how the object can be used, no interference is made in the application of new ideas.
When competition exists between one or more self-generated methods and an established problem solving method, performance will be hindered by first attempting to apply the previously presented method.

Raaheim (1965, cited in Kahney, 1993) suggests that interference of past experience on performance is due, not only to the problems with applying new solutions, but to the inability to generate new problem solving strategies. In experiments demonstrating the role played by functional fixedness, participants are given examples relating to how to solve the task (i.e. Luchins, 1942, cited in Kahney, 1993) or they share common experience regarding an idea about how concepts should be used (Ward, 1994). These designs do not prove that when attempting to use an inappropriate method, participants are aware that another method of problem solving is possible. If unavailability of alternative methods were responsible, then functional fixedness would be the result of lack of varied experience that may be necessary to display creativity in this field. If this hypothesis were correct, then presenting several alternative strategies would encourage the awareness that alternative problem solving strategies are possible and make novel generation of alternative strategies more likely. Glick and Holyoak (1983) have theorised that presentation of many examples is the key to developing a general problem-solving schema.

The aim of this experiment is to distinguish between these two alternative explanations for functional fixedness and accordingly, determine the most appropriate method to train in order to limit the effects of functional fixedness. An experiment will test ability to develop and identify appropriate strategies to complete novel and related problem-solving tasks. The task involves completing tangrams; a visio-spatial puzzle that can be used as a valid measure of creativity (Domino, 1980). Training will involve chance to develop methods suitable to complete tasks, and then the ability to complete novel and related tasks will be tested. Training given will be either specific; all puzzles may be solved by one method, or non-specific; there will be three appropriate methods to solve the puzzles. The success in completing novel tasks will be dependant on formulating a different method to complete the puzzles, whereas the success in completing related tasks will be dependent on correctly identifying the appropriate method from those experienced in training.

If the interference hypothesis suggested by Smith, Ward and Schumacher (1993) is correct, then functional fixedness is a result of competition between methods that have been established during training and those generated by the participant. Therefore, with introduction of more possible solutions, performance on novel and related tasks will be inhibited to a greater extent, as there will be greater competition. Alternatively, the results may support the hypothesis generated on the grounds of Glick and Holyoak's (1983) work; that functional fixedness is a result of inappropriate use of one method due to not realising that alternative methods for problem solving are possible. This unavailability hypothesis would be supported if performance on a novel task is improved after several alternative strategies have been ; if performance in novel task is better after non-specific training compared to specific training.

The experiment will also include measures of performance regarding previously seen puzzles (old task) and isomorphic puzzles; puzzles that are unseen but based on the same structure as the previously seen problems (new task). The logic for testing the former is to prove that training has been effective by showing that performance on identical tasks where transfer should be optimum has been improved. Success in 'new task' is dependent in identifying the category dependent principles and applying these to appropriate tasks.

Appropriate methods to solve 'new test' will be introduced through training, so performance depends on the ability to access these methods. Results suggesting that performance on 'new task' is affected by functional fixedness would suggest that this type of negative transfer is due to inaccessibility of appropriate method, as, due to experience with similar tasks, method generation is not necessary. If results show that non-specific training leads to decrease in performance, compared to specific training, this would suggest that competition between alternative methods is responsible for the reduction of performance as non-specific training involves presentation of more possible methods so would encourage more competition.

As well as the difference, according to training group, in performance on 'novel task' information about the mechanisms behind functional fixedness will also be gathered through difference in performance between the test types. The use of several tasks allows comparison between relative performances on different test types. Also, gathering data from different types of test allows specific conclusions if the effects of functional fixedness are due to a combination of the two hypotheses.

Despite training specificity differences, amount of training examples must be kept constant to make the performance of the training groups comparable, as Ward (1994) showed that greater conformity effects are the result of exposure to a larger number of examples. For this reason, the number of examples seen from each category will be kept constant, and only the number of categories will be manipulated.

The results will show how specificity of training affects the ability to develop creative solutions to novel tasks, and will also add information to the limited body of knowledge about negative transfer, including what conditions, in terms of training and test are most likely to elicit negative transfer and also what parts of the puzzle are affected by this phenomenon. The knowledge about effects on creativity will indicate a method of training, which will maximise performance without hindering the ability to generate new methods.

Method:

Participants:24 Undergraduate Psychology students: including 7 male and 14 females, volunteered for the study as partial fulfilment of course requirements. The mean age of these participants was 22.08 years with a standard deviation of 5.98 years. No participant had any prior experience with computer generated tangram puzzles.

Design:A 2x3 mixed design was used. The independent variable; the training condition, was studied at 2 levels, specific training, non-specific training. The performance of a control group, given no training was also measured. The dependant variable tested was the time taken and the number of moves made to complete tangram tasks. The performance during training and 'old test' consisting of previously seen examples 'new test' including examples with the same structure, but different in appearance to training examples, and 'novel test'; examples different in structure from previously seen examples method were analysed.

Materials:

Stimuli:A tangram is a spatial puzzle, similar in principle to a jigsaw puzzle. A tangram is created from 7 components; two large, one medium and two small triangles, a parallelogram and a square. These components are arranged together, to represent a target. A computer generation module[1] was used by a group of undergraduate psychology students to create a catalogue of targets with these components. Relative configurations of the components were specified, and the resulting target was outlined. Marsh, Lindau and Hicks (1996) showed that the central structure is most susceptible to conformity effects; therefore, defining categories on the basis of these central components would increase the probability that conformity effects would relate to these categories. From the catalogue, targets were categorised according to the use of two large triangles, chosen as the determinants of category as they were the biggest therefore, generally the central components of the target. From all targets involving the two large triangles placed next to each other in a central position, four high frequency alternative uses were identified.

The type of representation created by each target was identified in a pilot study involving 21 students. Targets were individually displayed on an overhead projector for 30 seconds. Participants were asked to guess what the target represented, and rate the confidence with which they made this judgement, on a scale of 1-7. Based on this data, the targets were coded according to whether they represented a geometric form (i.e. random shapes, patterns, letters of the alphabet), a non-living object (i.e. a teapot, a saucepan) or a living object (i.e. a man, a swan).

Once targets were generated and coded, an experimental phase was used. In one trial of the experimental phase, two experimenter-specified targets were displayed on the screen. The left-hand target encompassed the components, and the right-hand target consisted of an outline only. The task involved moving the components from the left-hand target to fill and not overlap the outline of the right-hand target. To do this, controls consisted of: select the component by double clicking with the left-hand mouse button, positioning the mouse appropriately, rotating the component by rotating the wheel in the centre of the mouse and dropping it using the left-hand button. No limit was made on the number of moves possible with any component. The trial finished when all of the components were inside the right-hand target outline.

Forty college students took part in a second pilot study where trials of targets were completed and the time taken to complete each trial was recorded. The stimuli for the experiment were then chosen. All of these targets were asymmetric as pilot data suggested a difference in time taken to complete symmetric or asymmetric targets. Eight targets were chosen from each of the 4 categories (on the grounds of their core shape), All targets selected were completed between 3 and 7 minutes during the pilot study. This criterion was used to avoid floor and ceiling affects. Within each category, half of the chosen targets were geometric representations. Of the remaining 4 targets, 2 of these represented non-living objects and 2 represented living things. This criterion

[1] Software developed by Martin Hall and Jin Zhang, University of Southampton.

ensured that any effects of representation type on performance would be constant between categories.

Presentation: A 'start target' was generated, using all components, but not including a core shape; to ensure that presence of a similar core shape would not affect performance on any trial. One trial of the experimental phase involved moving the components from the start-target to fill the outline of another target. In the first trial of the experiment, the start-target was presented with an outline it's own 90 degree rotation. This was included in order to accustom the participant to the task, without providing any training, and performance on this trial was not analysed. Completion of one trial triggered the presentation of the next trial. Blocks consisted of 4 trials, balanced in terms of representation type. Presentation of individual targets and categories were counterbalanced to be included in each part of the experiment and for each training condition. Order of presentation within a block was randomised.

Distraction task: The aim of this tracking task was to expose the control group to the same amount of technical experience and fatigue as the experimental groups, without exposure to the experimental stimuli. Participants in the control group were required to use the mouse to keep a curser aligned with a shape on a computer screen. The controls involved rotating the curser by turning the wheel in the mouse and by moving the mouse to adjust the position of the curser. This task lasted for 30 minutes. After each 10-minute block, the shape of the target was changed to reduce boredom.

Procedure: Participants were randomly assigned to a training condition. Each participant was seated in a cubicle including a window and a computer. After reading and signing a statement of consent, the participant was asked to read instructions presented on the screen. The experiment was presented on an IBM compatible personal computer with 19-inch monitor and a screen resolution of 1024 by 920.

The experiment consisted of; training, 'old task', 'new task' and 'novel task'. Training involved completing 3 blocks of targets. For participants in the specific training condition, training consisted of one block of 4 targets, selected from within one category, three times. For participants in the non-specific training condition, training involved completing three blocks of targets, each selected from a different category. During 'old task', participants completed one block, consisting of targets that had been experienced during training. For the specific training group, these targets were the same as those completed in training. For the non-specific training group, these targets were selected from within the three blocks completed during training.

In 'new task', participants completed one block of previously unseen targets selected from the categories encountered during training and 'old test'. For participants in the specific training condition, these targets were selected from the one condition experienced earlier and for the non-specific training condition, these targets were taken from the three categories previously experienced. For both training conditions, 'novel test' involved completing one block of targets selected from one category from which no targets had been previously presented for that participant. Counterbalancing was carried out so that targets and categories were equally represented in each stage of the experiment and for each training condition.

Throughout the experiment, participants completed one block and then raised their hand, at which point the experimenter loaded the next block. The delay was designed to reduce effects of fatigue on performance. The participants in the experimental conditions completed 6 blocks of 4 targets. The experiment took around 90 minutes to complete.

Participants allocated to the control condition completed the distraction task before carrying out 'novel test'. All participants were then debriefed.

Results:

Time taken in seconds to complete block, by training condition.

Training Condition	Specific [a]		Non-specific [a]		None [a]	
	M	SD	M	SD	M	SD
train block1	1194.40	693.14	1211.45	430.58		
train block2	461.79	276.45	844.39	474.25		
train block3	276.10	53.20	1030.92	596.76		
test old	312.38	74.28	496.42	318.65		
test new	971.82	695.28	703.42	578.82		
test novel	477.61	241.93	539.35	230.07	1148.77	911.46

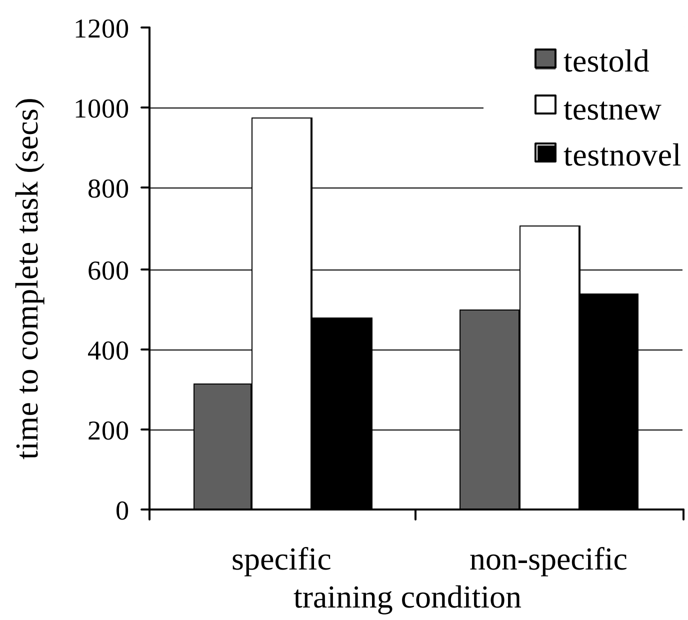
[a] $\underline{n} = 8$.

Observations of data regarding time taken and number of moves made to complete tasks yielded similar pattern of results. Therefore, only results relating to time taken are reported. All rejection levels for analysis are set at .05.

Effect of Training on Novel test performance

For descriptive data on novel test performance please refer to bottom row of the table above. Observations showed that performance on novel task was poorer under no training condition, than performance under specific or non-specific training conditions. A Repeated measures ANOVA showed a significant effect of condition on time taken to complete novel task: ($F_{,2,14}$ = 3.51, p = .49). No significant difference was found in time taken to complete novel task for specific and for non-specific training ($t_{,14}$ = -.523, p = .609). The difference between time taken to complete novel task was borderline significant between no training and specific training conditions ($t_{,14}$ = -2.01, p = .064) and between no training and novel training conditions ($t_{,14}$ = -1.83, p = .088).

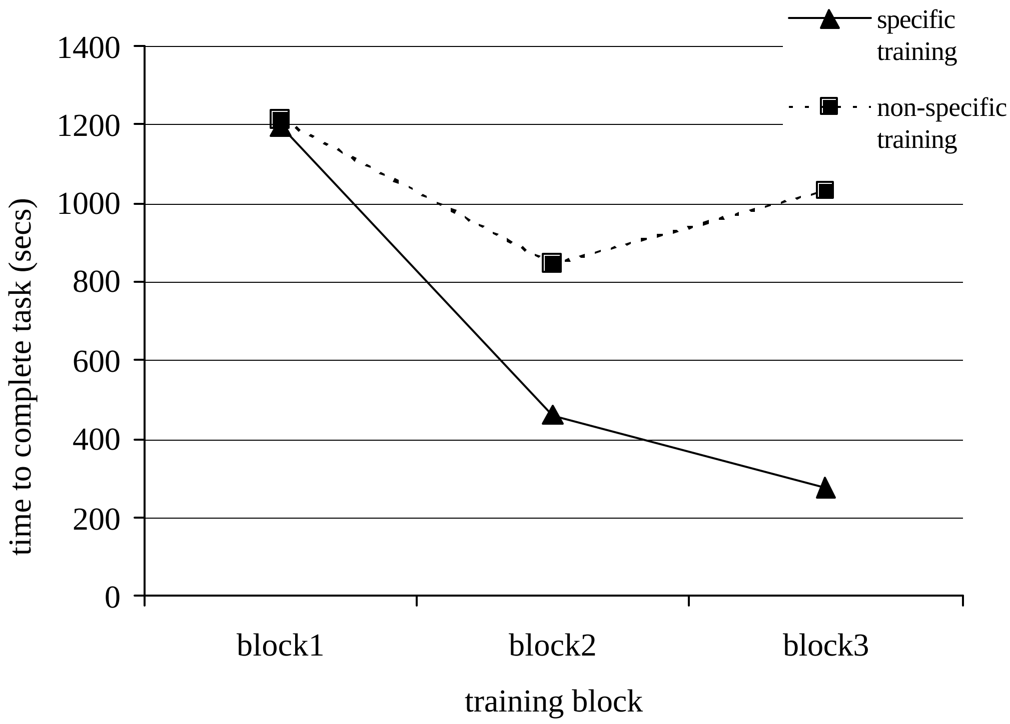
Effect of Training Condition on test performance



Performance by training condition according to test type.

Within each training condition, ‘test old’ was the quickest test to complete, followed by ‘test novel’. Greatest time was taken to complete ‘test new’. This pattern was more pronounced under specific training conditions (figure above). A repeated measures ANOVA revealed a significant effect of test type on task performance ($F_{,2,14}$ = 5.48, p = .01). The ANOVA also revealed no effect of training condition on test performance ($F_{2,14}$ = 0.03, p = .956). Further analysis revealed a significant difference between test old and test new for both specific training condition ($t_{,7}$ = 11.89, p <.001) and for non-specific training condition ($t_{,7}$ = 4.406, p = .003). Significant differences were also found between new and novel test for both specific training (t_{7} = 3.95, p = .006) and non-specific training (t_{7} = 3.44, p = .011). Differences were also significant between old and novel tests for specific training (t_{7} = 5.58, p = .001) and for non-specific training (t_{7} = 6.63, p < .001).

Effects of training condition on performance during training.



Performance on training blocks by training condition

The curve representing performance for specific training condition showed reduced time taken from block1 to block2 and to block3. For non-specific training condition, time taken to complete task decreased from block1 to block2, then increased for block3. Differences in performance between specific and non-specific training were slight for block1 moderate at block2 and increased further for block3 (figure above).

A Repeated measures ANOVA revealed significant effects of training block (F_{2} = 7.38, p = .003) and a main effect of condition ($F_{2,14}$ = 7.843, p = .014) but no significant interaction between training block and condition ($F_{2,14}$ = 2.49, p = .101).

For both specific and non-specific training conditions, differences were significant between block1 and block2: (t_{14} = 4.87, p = .002) (t_{14} = 7.96, p <.001) respectively. In both specific and non-specific training categories, significant differences were found between training block2 and block3 (t_{14} = 4.73, p = .002) (t_{14} = 5.04, p = .002) respectively. Differences between block1 and block3 were also significant for specific training (t_{14} = 14.69, p < .001) and non-specific training (t_{14} = 4.89, p = .002).

The difference between specific and non-specific training was non-significant for block1 (t_{14} = -.059, p = .954), borderline significant for block2 (t_{14} = -1.971, p = .069) and significant for block3 (t_{14} = -3.56, p = .003).

Discussion:
The aim of this study was to determine training conditions in which creativity is least inhibited by functional fixedness. To do this, the study tested whether unavailability of appropriate methods resulting in decreased performance in tasks after training is due to lack of awareness that alternative methods are possible, or due to competition from inappropriate skills derived from experience.

To distinguish between these alternatives, this study analysed performance on a novel task after specific or non-specific training. It was theorised that if the competition theory were correct, performance would be hindered to a greater extent after several alternative methods had been introduced, as in non-specific training. Alternatively, if the absence of method generation was due to assuming that only one method was possible, then non-specific training, where several possible methods for completing the task were suggested, would highlight the possibility of multiple methods, therefore reduce the negative effects of fixedness. The results showed that ‘novel task’ performance was improved after training, compared to performance with no training and was approached performance shown in practised tasks. This suggests that functional fixedness was not a factor in the performance of ‘novel task’. The reason that this novel task was not hindered may have been that the task was adequately different from the training task to be unaffected by negative transfer, or it may be that previous completion of ‘new task’ had eliminated detrimental effects. These alternatives will be discussed later. Further to the improvement in performance, there was no significant difference between performances after specific or non-specific training conditions, so ‘novel test’ performance does not provide support for either hypothesis regarding the mechanisms of functional fixedness. Therefore, further analysis of the results is needed to distinguish between these hypotheses.

Analysis of all test results showed that in ‘new task’ (a task where structure was based on the same core shape as the tasks used in training, but was not identical in presentation to training tasks), performance was poor in comparison to both ‘old task’ and ‘novel task’, and was only slightly better than performance with no training at all (control group). The decreased performance was more pronounced in the case of specific training; the difference between performances on ‘new task’ between training conditions approached significance despite a small sample: 8 participants in each group. This suggests that negative transfer plays a role in cases where the training and task are very similar, and especially with limited variability of training examples.

This finding that negative transfer affects performance on ‘new task’ disputes the hypothesis that functional fixedness is due to unavailability of alternative methods for task completion, as the detrimental effects are shown when the appropriate method for completion of task has been experienced during the training phase. However, if the alternative competition hypothesis was true, and inhibited performance was due to competition between alternative methods, it was expected that the performance in ‘new task’ would be worse after non-specific training, as there are a greater number of proposed methods, therefore, several possible sources of competition. The results showed that when the participant completed non-specific training, thus completing several methods in contrast to one method as experienced by the specific training group, the detrimental effect was less pronounced. The theory that competition of alternative methods would decrease task performance was based on the assumption that one method from those experienced previously will be chosen at random and replaced at random if it was not suitable to complete the task. These findings show that the application of previously experienced methods for problem solving is not random. In fact, when the appropriate method has been experienced previously, then provision of several alternative methods decreases the time taken to find the correct method.

Before interpretations of the results are discussed any further, two alternative accounts for the results must be acknowledged. In the design of this experiment, the primary concern was to allow comparison of performance after specific and non-specific training. To keep the strength of the conformity effect constant between training conditions, the same numbers of targets from each category were seen by each training condition. However, in order to also equalise the effects of fatigue between training conditions, participants in each each training condition completed the same number of trials. As the non-specific training group completed 12 trials altogether: 4 trials from each of 3 categories, the specific training group were required to repeat the 4 trials selected from one category three times. The result of this design was, for the specific training condition especially, learning may have been confined to these targets and not generalise to the category. This would explain the observed poor performance on ‘new task’ in comparison to previously seen targets in ‘old task’. However, if this explanation were valid, no difference in performance would be expected between tests, when none of the actual targets had been seen.

However, by looking at the results, it is clear that there is a significant difference between performance on the same category and performance on a different category (difference between time taken on ‘new task’ and on ‘novel task’). In other words, although there is a significant difference between performance on seen and unseen targets within one category: between performance on ‘test old’ and ‘test new’, the difference between previously experienced and inexperienced category is also significant. This shows that whether or not the category has been experienced is influential; therefore learning is not confined to the specific targets seen, but affects the category represented. Also, as the effects of test type have been observed to follow the same pattern for both specific and non-specific training conditions, the design used, involving repetition of fewer examples during specific training condition, was a valid method of comparing training conditions.

The other possible non-intentional explanation of the results is related to the order of presentation. Hammerton (1981, cited in Singley and Anderson, 1989) describes negative transfer as a short-lived effect. Further, Bilodeau and Bilodeau (1961, cited in Singley and Anderson, 1989) have shown that this negative effect may convert to positive transfer after a short time. The results of this experiment: that performance in ‘new task’ is poorer than performance in ‘novel task’ may be showing the morph of negative into positive transfer, which would happen irrelevant of the type of test being performed. There are two reasons why this interpretation is unlikely to be accurate. The first reason is that the results show a stable low level of performance shifting to a stable improved performance for ‘test novel’, unlike the gradual improvement in performance over the course of the trials, as described by Bilodeau and Bilodeau (1961). Also recovery is shown in every participant and this recovery coincides with the same trial as test type is changed. If the change in performance were due to spontaneous recovery, not test type, then this coincidence would be unlikely. This suggests that the improvement in performance is due to the type of test and not the order of the trials.

Functional fixedness can account for the decrease in performance on the ‘new test’ as performance was poor compared to the other test phases and almost comparable to performance on the task with no training. It may be suggested that if training were continued to asymptote, the performance would be worse for ‘test new’ after specific training than performance without any training.

It was anticipated that functional fixedness developed as a result of targets encountered during training would improve performance in ‘new task’, as the internal structure is the same in ‘new test’ as in training. The internal structure was considered most susceptible to the effects of functional fixedness as this structure was common for all training targets, and previous research has shown that the conformity effect is stronger when a greater number of examples have been seen (Ward, 1994). Also, functional fixedness has been shown to be most susceptible to central components of the task (Marsh, Lindau and Hicks, 1996), a position in which the core structure was found. However, the decrease in performance observed in ‘new task’ suggests that fixedness is related to other features of the target than the internal structure. A possible explanation is that fixedness is related to the peripheral details of the targets; participants noticed the non-core elements and attempted to use them in completing the new trials where these elements were not appropriate. Evidence from the results in conjunction with previous findings suggests that peripheral details were influential in the negative transfer of skills to ‘new test’.

Previous research, (i.e. Mayer, 1977) has shown that when stimuli are similar to one another, attention will be focussed on the peripheral details, as these features will be used to distinguish between the stimuli. Therefore, in the specific training condition, where in comparison to the non-specific training condition, the stimuli are collectively similar, the peripheral details would be noticed more. As a result of this, learning of these peripheral details and therefore, attempting to apply these to solving the tasks in ‘new test’ would be more likely in the specific training group. This would mean that, if performance in ‘new test’ were based on application of peripheral details, performance would be inhibited to a greater extent for those who had completed specific training compared to those who had completed non-specific training. This is the pattern observed in the results.

Previous training also inhibits the performance on ‘new task’ after non-specific training, but to a lesser extent. The stimuli in training was varied in core structure, therefore it is less likely that participants would focus on the peripheral details of the target in order to distinguish between the stimuli. The focus on peripheral as opposed to central details suggests that these participants were not behaving as experts in this task (Simon and Chase, 1973). This may be because the training was not adequate, and in fact performance did not reach asymptote. If the training had have been more comprehensive, involving targets from a greater number of categories, then the focus of attention may have been directed towards the central components of the targets and functional fixedness may have reflected the core structures as opposed to the surface details. Including more training trials in the experiment could test this. The finding that functional fixedness can be shown in relation to the peripheral details has not been highlighted in previous research. This research suggests that the focus of attention and therefore the focus of functional fixedness may be manipulated by changing extent of training and the type of training stimuli.

As mentioned earlier, the effects of functional fixedness are not due to random application of any available method, but use of the most appropriate method taking into account the structures of training and test stimuli. This explains why, during ‘new test’, participants given non-specific training did not have more interference from a greater number of possible methods, as would be expected if any method experienced was randomly applied to a related task. It appears that the number of possible methods does not have a proportional detrimental effect on the time taken to find an appropriate method from within them. This shows that there must be some similarity between tasks for transfer of non-optimal methods to take place. The reason that performance was hindered in ‘new test’ is because the test stimuli were similar to the training stimuli; so non-optimal methods found to be appropriate in the training tasks were transferred. Why were these detrimental effects not transferred to ‘novel test’?

One explanation for why non-optimal methods were not transferred from training to ‘novel task’ is related to the earlier presentation of ‘new test’. Glucksberg and Danks (1968) suggested that optimum training involves developing many possible methods and then having an opportunity to identify the successful from the unsuccessful attempts. In this experiment, training may have resulted in development of many hypotheses regarding appropriate methods to use when solving tests: some accurate and some inaccurate. These hypotheses would be applied to solving ‘new test’ and, according to their influence on solving the puzzles, modified to incorporate only those hypotheses that are still useful.

In this experiment, core shape was the identifying feature of each category. Arranging the components according to the core shape would have led to success in completing targets during the training phase, during ‘test old’ and ‘test new’ but not in completing targets in ‘novel test’. Therefore, if the hypotheses formulated in the training phase were modified according to success in ‘old test’ and ‘new test’, these core shapes should remain the only relevant methods. When transferred onto ‘novel test’, these refined hypotheses are no longer appropriate, as the novel stimuli have different core shapes. Therefore, if this hypothesis modification hypothesis were correct, then it would lead to inappropriate methods in ‘novel test’, resulting in decreased performance. This pattern was not observed in the results, so the sequence of test phases is unlikely to be responsible for the pattern of results.

A more probable explanation to account for the finding that non-optimal methods were transferred from the training phase to ‘new test’, but not applied to ‘novel test’ is that the targets used in ‘novel test’ differed in structure enough compared to training targets for transference of exact methods to be judged not relevant. The improvement in performance in ‘novel task’ compared to performance in ‘novel task’ by the control group, suggests that benefits of training were transferred. This suggests that the difference in structure of the tasks allowed general problem solving principles gathered throughout training to be transferred, but the specific non-optimal methods relating to the peripheral features of the specific targets were not used in ‘novel test’. Verbal accounts from the participants verified the assumption that novel targets were identified as different in appearance compared to targets used in training, ‘old test’ or ‘new test’. This accounts for why completion of the novel targets was not inhibited by previous training.

However, during training, it has been shown that peripheral details of the targets, not core shapes, were the most influential in transfer to later training phases. The only difference between the ‘novel tasks’ and the previous trials were the core shapes involved. The verbal accounts and experimental data indicate that core structure influenced the decision whether or not to use specific methods gathered in previous trials. In other words, the internal structure was not noticed within categories, when it was constant, it was only noticed between categories, when the categories differed on the grounds of inner structure.

To summarise, specific and non-specific training both result in functional fixedness, in tasks judged similar to those encountered during training. In tasks similar to those used in training, the greatest effects of functional fixedness were shown after specific training. Experience with similar targets does not necessarily encourage functional fixedness in regards to features that are common to, or centrally located in all training examples. When the entire method is not transferred to a related task, the part that is susceptible to functional fixedness may be determined by the focus of attention during training; which is influenced by similarity within the examples experienced. This study has also shown that there may be a difference in the way that features of a stimulus are noticed between categories, compared to how these features are noticed within a category.

Specifically, when all stimuli are similar, peripheral differences between the stimuli are noticed. However, when stimuli are from a different category, the defining differences between the categories are noticed.

The aim of this study was to find why functional fixedness effects performance, with the aim of formulating a training method which eliminates this constraint on creativity. In regards to training tasks, the results suggest that functional fixedness influences performance in similar, and not novel tasks. Therefore, if training is sufficiently different from the required task, the negative affects of functional fixedness will be avoided, but the positive transfer of general skill will improve performance. If training needs to be similar to a later task, non-specific training will be most suitable. The benefits of improved performance from experience with specific training are also seen after non-specific training, though the disadvantages of decreased flexibility to novel tasks are less pronounced. Functional fixedness will reduce creativity in similar tasks due to the over-identification of situations where methods are appropriate. Non-specific training will optimise creativity potential, and will provide different features to be applied when appropriate, thus reducing inappropriate over-application.

References:

Birch, H.G. and Rabinowitz, H.S. (1951) The negative effect of previous experience on productive thinking. *Journal of Experimental Psychology*, 41. 121-125.

Domino, G. (1980) Chinese tangrams as a technique to assess creativity. *Journal of Creative Behaviour*, 14. 204-213.

Glick, M.L and Holyoak, K.J. (1980) Analogical Problem Solving. *Cognitive Psychology*, 12, 306-355.

Glucksberg, S. and Danks, J.L. (1968) effects of discriminative labels and of nonsense labels upon availability of novel function. *Journal of Verbal Learning and Verbal Behaviour*, 7. 72-

Hampton, P.J and Morris, P.E. (1997) *Understanding Cognition*. Oxford: Blackwell.

Kahney, H. (1993) *Problem solving current issues* (2nd Ed) London: Open University Press.

Marsch, R.L, Lindau, J.D. and Hicks, J.L (1996) How examples may and may not constrain creativity. *Memory and Cognition*, 24. 669-680

Mayer, R.E. (1977) *Thinking and problem solving*. California: Scott, Foresman and Company.

Simon, H.A. and Chase, W.G. (1973) Skill in Chess. *American Scientist*, 61, 394-403.

Singley, M.K. and Anderson, J.R. (1989) *The transfer of cognitive skill*. Harvard: Harvard University Press.

Smith, S.M., Ward, T.B. and Schumacher, J.S. (1993) Constraining effects of examples in a creative generation task. *Memory and Cognition*, 21 837-845

Ward, T.B. (1994) Structured Imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, 27. 1-40

STUDY 2: USE AND MEMORY OF CONFIGURAL AND HOLISTIC INFORMATION

Abstract:
Part and whole task training were investigated to determine which method of training produces more efficient task performance and enhances skill acquisition of component skills for a visual spatial task. With a mixed design the study compared part-stimuli training of the Tangram Game with whole-stimuli training, also different categories of tangrams were considered, transparent and opaque. With regards to skill acquisition participants' performance was analysed on an untrained task and untrained stimuli. The results support Stammers' (1982) preference for whole-task training, as there were no significant differences in response times for the different training conditions. Overall transparent tangrams were better for skill acquisition of this task. Participants constructing transparent tangrams had lower response time in training, better performance after training than the other groups, and were more able to transfer skills to new stimuli and similar untrained tasks, showing better transferability. The results of shape category influence are discussed in terms of skill acquisition theories. It has been suggested transparent shapes enhance skill acquisition of the tangram task as a result of the procedural reinstatement theory (Clawson, Healy, Ericsson and Bourne, 2001) and transparent shapes allow the development of strategic skills.

Introduction:
Previous research has demonstrated a difference in performance on specific tasks after part or whole-task training. "Part-task training is defined as practice on some set of components of a whole task as a prelude to practice of or performance on the whole task" (Wightman and Lintern, 1985, p. 267). There has been a long unresolved debate over which method is more efficient for overall performance. Part-task training is thought to allow individuals to develop component skills that can transfer to the whole task without obstructing cognitive demands needed for the whole-task, however, the advantages of part-task training have only received moderate empirical support and in numerous studies whole-task training has been more effective (Goettl and Shute, 1996). It was discussed in a review by Stammers (1982) that the range of tasks so far has been limited and it is unacceptable as yet to make any generalisations about part or whole task training. Hence, the main focus of this paper is to examine which method of training, part-task or whole-task, produces more efficient task performance and enhances skill acquisition of a visual spatial task.

The aims of part-task training are to reduce costs, for example time and cognitive resources, and improve learning to make performance of the whole-task more efficient. Wightman and Lintern (1985) review 3 ways in which tasks can be divided into separate parts. Segmentation is a method that divides a task by temporal dimensions, where a task has identifiable endpoints. A bombing task experiment by Bailey, Hughes and Jones (1980, cited in Wightman & Lintern, 1985) was reviewed as a segmentation method, it showed an advantage for part-task training, it permitted intensive practise of difficult parts without wasting time on easy parts. Fractionation is a part-task training method for whole tasks in which two or more subtasks are happening at the same time. Only one of the experiments reviewed by Wightman and Lintern (1985) showed an advantage for fractionation part training, where as all the others favoured whole-task training. This may be because within these tasks the parts have high interaction, i.e. they happen at the same time, which would make training the parts separately more difficult to later integrate. Simplification is a part-task procedure where a difficult task is made easier. Wightman and Lintern (1985) conclude simplification allows positive transfer of knowledge. Skills learned with an easier task can be applied to a similar difficult task. It is claimed simplification is more powerful than whole-task training, however, little is known about how to optimise simplification for training. This area of part-task training allows for further research.

Previous research has highlighted that part-task training is an effective method to train complex manual control and tracking tasks because training on individual components of a complex skill improves performance on the whole task (Goettl & Shute, 1996). The component fluency hypothesis by Carlson, Khoo and Elliot (1990) supports this idea. It believes that complex tasks consist of a hierarchy of basic component skills and organisation strategies, and that there are capacity limits placed on cognitive resources hence fluency of the component skills is critical to skilled performance on complex tasks. The hypothesis claims that practicing the critical component skills separately can increase the fluency of target task performance. However, for part-task training to be effective the critical component skills need to be identified, this is challenged by the fact that critical skills may change as expertise develops. According to Ackerman's Theory of Skill Acquisition (1992) early learning knowledge is declarative, it is based on what is in working memory, after practise knowledge becomes procedural and component skills become automated. For this reason there may be changes in the criticality of component skills as acquisition develops, and training the components separately may hinder task performance.

Skill acquisition is maintained high performance of a specific task. However, the time needed for skill acquisition can depend on the type of task and whether the same procedures can be used across the whole task. In a recent paper, Clawson, Healy, Ericsson and Bourne (2001) evaluated the Procedural Reinstatement Theory with regard to skill acquisition. This states procedural tasks explicitly require the use of the same procedures at retention as during training. Declarative tasks do not use procedural knowledge, they use different procedures at training and retention; i.e. they start with declarative knowledge and as they become skilled use procedural knowledge. The theory claims performance on procedural and declarative tasks are different at retention, procedural tasks perform more efficiently because their solutions are more direct and hence they do not need to rely on skills learnt during training. The learning procedures are more indirect for declarative tasks, and hence it could be thought that training tasks that use declarative information by parts will hinder performance of the whole task, as it will have low levels of retention of all the parts learned separately.

Skill acquisition is also influenced by comprehension of stimuli specific knowledge and strategic skills of the task (Doane, Sohn & Schreiber, 1999). It has been shown that strategic skills optimise performance of a task as they serve to minimize redundant information processing (Brown & Kane, 1988). This could be compared to the hypothetical positive effects of part-task training, there may be something in the method of training or the task being trained that allows strategic skills to develop above stimuli specific knowledge. One idea is that task difficulty of initial learning context influences strategic skill acquisition (Doane, Alderton, Sohn, & Pellegrino, 1996).

Part-task training effectiveness depends on the interactions among components (Goettl & Shute, 1996). Blum and Naylor (1968) previously addressed the issue of interrelatedness of components tasks. They believed the characteristics of a task such as complexity, organisation, and interactions of sub-components determine the relative efficiency of whole-task or part-task training. With high organisation tasks (interrelatedness of tasks), as complexity increases whole-task training should be relatively more efficient, with low organisation tasks as complexity increases part-task training methods should be more efficient than whole. In Stammers review (1982) there is supportive evidence of this hypothesis by Naylor and Briggs (1963).

Exposure to the criterion task (i.e. the whole task) is thought to increase the effectiveness of component practice (Carlson et al., 1990). However there is a contentious issue on when to expose the trainee to the task. Carlson et al. (1990) concludes there is little or no transfer of knowledge if the criterion task is not presented before skill acquisition. They put forward that trainees should be introduced to the criterion task after some part-task training but before they have developed procedural knowledge for the task. They are suggesting a mixture of part and whole task training achieves the best final problem solving performance as it allows the trainee to develop component procedure that may be effectively practiced outside the criterion context. Singley and Anderson (1989) suggest exposure to the whole task will be most effective as early on in the training as possible so it can provide a template for the development of whole task specific skills in part-task training. These findings could almost suggest that it is whole-task training that influences skill acquisition of the component tasks and hence influences overall task performance.

As it happens on many occasions whole-task training has been more effective than part-task training. In the Goettl and Shute (1996) experiments an initial advantage for whole-task training was found. Whole-task training was more efficient than part-task training with regard to speed measures. In Stammers (1982) review of part and whole methods of training of the four experiments considered, part-task training did not have any advantage over whole-task training and in two of the experiments advantages were found for whole-task training. Stammers (1982) provides a very cautious favourable use of whole-task training, claiming whole-task training gives the most effective approach to training in the long term. His argument is supported by observations from the reviewed studies. He maintains part-task training does offer some savings, however, because part-task training places greater demands on the trainee and as there are only small differences between the groups of part and whole training it is better to use whole-task training; it avoids isolating key skill elements from context, which could be distorted in part-task training.

Whaley and Fisk (1993) also found no benefit for part-task training for a memory dependent skilled search task. However, when retention of task was assessed 30 days later part-task training was slightly superior to whole-task training. The article suggests part-task training may be beneficial in refresher courses and in the long term for tasks involving memory or visual search tasks.

This aim of this study is to discover which method of training, part-task training or whole-task training, is better for skill acquisition of a visual spatial task and whether the category of stimuli has an effect on performance. Specifically the study focuses on performance during training, to examine whether one training group improves significantly more on their task than the other. This study also considers performance after training, to see whether one type of training group performs consistently better in a testing condition than the other. With regard to skill acquisition the experiment considers transferability, whether the participants are able to transfer their knowledge; for example, which training group performs better on untrained stimuli, on an untrained similar task and on an untrained task with novel stimuli. Lastly the experiment considers the effect of stimuli type and whether different stimulus affects the efficiency of part or whole-task training in training, testing and transfer conditions.

The Tangram Game is the visual spatial task used in this experiment. This was chosen as very little research has been done on spatial puzzles, and hence little is known about the processes involved in solving them (Butler, 1994). The tangram consists of a square tile cut into seven pieces: two large triangles, two small triangles, one medium triangle, a square and a parallelogram. The whole-task form of the game is to rearrange all seven pieces from a square position to form outline geometric shapes or living and non-living objects. Once constructed you move from that given shape to form another given outline, the previously completed tangram shape becomes the new starting position. The part-task training method is a simplification procedure. Instead of the seven parts of the game always arranged together, whether it is a square or a previously completed tangram, the seven parts are placed separately before each tangram shape is constructed. This study concentrates on the training of the actual stimuli structured in different ways and hence for this study part-task and whole-task training will be substituted with part-stimuli and whole-stimuli training. There are thousands of possible tangram shapes, however, this study compares only two distinct categories (opaque tangrams are those that have less than 4 identifiable component parts in the outline and transparent tangrams are those that have 4 or more) to keep the complexity and similarity at a similar level throughout. Each group (part-stimuli training and whole-stimuli training) are split into two sub-groups to compare the affect of different shape categorisation, i.e. part-stimuli training with opaque tangrams and part-stimuli training with transparent tangrams.

Participants are timed on how long it takes and how many moves they make to complete a tangram. To start both groups are trained and then tested on old and novel shapes of their category. Whole-stimuli training could be more efficient for skill acquisition as it gives an example of how the seven components fit together, components of the task may be highly interactive and hence need to be practiced as a whole. If whole-stimuli training were better for skill acquisition then participants would have a significant reduction in response time or number of moves to complete a tangram across learning blocks. Also if their training gave better performance during testing participants would have lower response times for old shapes than part-stimuli training and if they had a small difference in response times between old and novel tangrams this may suggest they can transfer their skills better than part-stimuli training. However, these predictions could be vice versa for part-stimuli training. Part-stimuli training may be more efficient for skill acquisition as presenting the components separately should not restrict participants perceptions of where components can be placed within the given outline. After initial training and testing both groups are tested on the opposite task, an untrained task, with old and novel stimuli of their category. If whole-stimuli training has skill acquisition participants would have lower response times for old shapes than part-stimuli training and if they had a small difference in response times between old and novel stimuli they can transfer their skills better than part-stimuli training. This effect may be because participants have acquired the procedural knowledge during training and the first test, and hence now benefit from placing the components separately so they can try other possibilities. However, part-stimuli training may be superior to whole-stimuli training in transferring skills to an untrained task, if this were true opposite results would be expected. All these predictions may be influenced by the category of tangrams used. For example the transparent tangrams may be a procedural task, as it is always clear from the outline where component parts may go, and hence skill acquisition and transferability may be easier for part-stimuli or whole-stimuli training. If this were the case transparent tangrams would show lower response times. Opaque tangrams may also affect the performance of part and whole-stimuli training.

Method;

Design: There were four experimental groups; part-stimuli training with opaque shapes, part-stimuli training with transparent shapes, whole-stimuli training with opaque shapes and whole-stimuli training with transparent shapes. The experiment was a mixed design. The between-subject independent variables were the experimental condition (part-stimuli training or whole-stimuli training) and the tangram shape classification (opaque tangrams or transparent tangrams). Within-subject independent variables were the novelty of tangram shape constructed (old and novel) and the task testing condition (part-stimuli test and whole-stimuli test). Table 1 shows the basic design for the experiment.

Training	Testing Condition	
Part-stimuli training		
Opaque shapes	part-stimuli test; old and novel opaque shapes	whole-stimuli test; old and novel opaque shapes
Transparent shapes	part-stimuli test; old and novel trans. shapes	whole-stimuli test; old and novel trans. shapes
Whole-stimuli training		
Opaque shapes	whole-stimuli test; old and novel opaque shapes	part-stimuli test; old and novel opaque shapes
Transparent shapes	whole-stimuli test; old and novel trans. shapes	part-stimuli test; old and novel trans. shapes

Experimental design: The dependent variables were the response times and number of moves to complete a tangram shape. In addition participants performed two recognition tests, after each testing condition, for tangram outline shapes and tangram part components. The independent variable was condition of shape or component (seen or unseen) and the dependent variable was a correct response.

Participants: Thirty-two participants were recruited through the research participant scheme at the University of Southampton; participants were credited participation points towards their degree for their partaking in the experiment. There were 25 women and 7 men, with ages ranging from 18-22 years, the mean age being 20. There were 8 participants in each experimental condition. Participants were assigned randomly. All participants were asked a preliminary question to confirm they had no previous experience with the Tangram Game.

Materials: The main stimulus used was the Tangram Game. Tangrams are spatial puzzles consisting of triangles, a square, and a parallelogram that can be arranged into larger simple shapes or other more complex and creative figures. The seven component parts of the tangram game that fit together to form a square, based on a 16 square grid. The basic idea of tangrams is to use all 7 components to construct given outlines of geometric shapes or pictures.

For the experiment a library of tangram shapes, living, nonliving, and geometric, was stored with a computer tangram development programme, designed by Mr. Martin Hall and Dr. Jin Zhang with a C++ builder for Windows. The outline of the shape was then saved as a target shape. The figure below shows examples of constructed tangrams that were saved in the development programme.

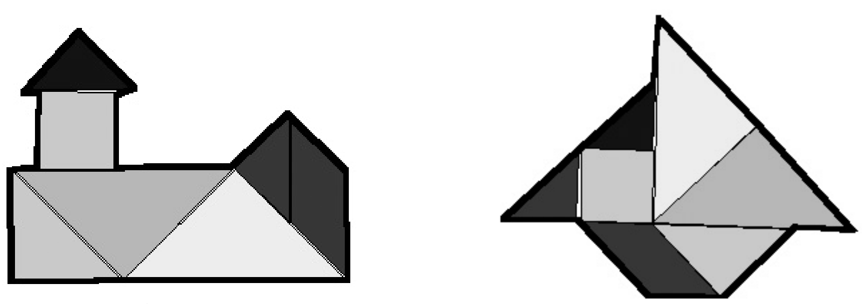


Figure. Examples of completed tangrams

As there is very little known about visual spatial puzzles a series of pre-experimental stimuli assessments were done. They were carried out to compose categories of tangrams that were similar in structure and complexity so that when interpreting the data any differences found could only be attributed to experimental manipulation. Pilot studies were done to validate all the target tangram shapes. Volunteers, not later involved in the experiment, assisted to compare similarity and complexity of the tangrams.

Tangrams were placed into two categories depending on how many of the 7 individual component parts could be identified from the outline of the tangram shape. Experimenters individually categorised each tangram into two classifications, opaque tangrams (where less than 4 individual component parts of the tangram could be identified) and transparent tangrams (where 4 or more individual component parts could be identified). Afterwards the five experimenters compared their ratings.

The tangram shapes were also placed in categories, defined by underlying structure i.e. the orientation of the two large triangles. After grouping the stimulus the first pilot study asked participants to name each of the shapes and rate their confidence of the decision they made on a scale of 1 to 7, where 7 is highly confident and 1 is not confident at all. This was done to check the experimenters' categorisation of living, non-living and geometric with people that had not had any contact with tangram shapes. The results indicated a high confidence value (above 4), thus high meaningfulness value, of shapes previously categorised as living and non-living by the experimenters' and low confidence value (below 4) of names given to shapes the experimenters' categorised as geometric.

Another pilot study investigated the time it took to complete each tangram. The response times varied a great deal and therefore tangrams that took longer than 7 minutes and less than 1 minute were removed from the available tangram library, this was to keep difficulty level fairly equal.

Participants of the pilot study were also asked to complete two similarity questionnaires, one testing the similarity of the 7 individual tangram component parts with 40 other non-tangram shapes of similar size, the other testing similarity of tangrams with other tangrams within the same category, i.e. geometric/opaque compared with other geometric/opaque. Each was rated on a scale of 1 to 7, 1 being very dissimilar and 7 being very similar.

Fourteen of these extra non-tangram shapes, from the similarity pilot study, were chosen to be the novel components in two recognition tests. These components were selected by taking an average of the similarity scores between the same categories of tangram shapes, then non-tangram components with the same similarity average were chosen to use in the recognition test. This methodology increased validity of the recognition test, as results would be determined by participants' representation of the outline shapes or component shapes not whether parts or shapes were more or less familiar to stimuli used in test phase, as they were of the same similarity.

Stimuli: There were 32 tangram target shapes chosen for the experiment, 16 transparent tangrams and 16 opaque tangrams. Of the 16 tangrams in each group 8 target shapes were used for the computer tangram tasks and the other 8 were used as unseen stimuli in the recognition tests. The 8 recognition tangram target shapes were used as a foil, however the other 8 computer tangram target shapes were counterbalanced throughout the experiment. Within both stimuli groups of 16 there was an equal mixture of non-living and geometric outlines from groups W, X, and Y. The experiment only used three of the underlying structure groups (W, X, Y), as they were all based on a new underlying geometric shape where as Z was two separate triangles, it was felt this outsized difference may disrupt any possible patterns of results. Also the tangram shapes chosen for the experiment were either geometric or nonliving, this was done as it was apparent from the pilot studies that there was substantially more familiarity towards living tangram shapes.

Procedure: Participants were tested in one session that took approximately 1½ hour. They were tested in a single computer cubicle, all instructions were given via the computer, and the experimenter was also present at the beginning to clarify any questions. The experiment was administered on an IBM compatible personal computer with a 19-inch monitor and a screen resolution of 1024 by 920. Participants were instructed to move component parts from the left side of the screen to the tangram outline on the right side of the screen by only using the mouse; double clicking a component with the mouse from the left side and dragging it into the outline of the target shape on the right side moved the components. Moving the wheel on the mouse could also rotate the components.

The experiment had four conditions where training, whole -stimuli training or part -stimuli training, and tangram shape classification, opaque or transparent, were manipulated. With whole-stimuli training the 7 component parts were arranged first in the square and then for each subsequent trial the previous completed shape became the starting position on the left side of the screen, components were black (see figure below, left panel). With part-stimuli training the 7 component parts were presented in different colours separately in a line down the left of the screen (see figure below, right panel.). Opaque tangrams were defined as having less than 4 of the components identifiable and transparent tangrams were defined as having 4 or more components visible. The four conditions were part-stimuli training with opaque tangrams, part-stimuli training with transparent tangrams, whole-stimuli training with opaque tangrams, and whole-stimuli training with transparent tangrams.

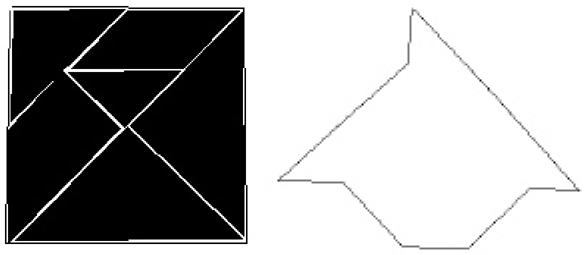


Figure. Whole-stimuli training presentation with an opaque shape.

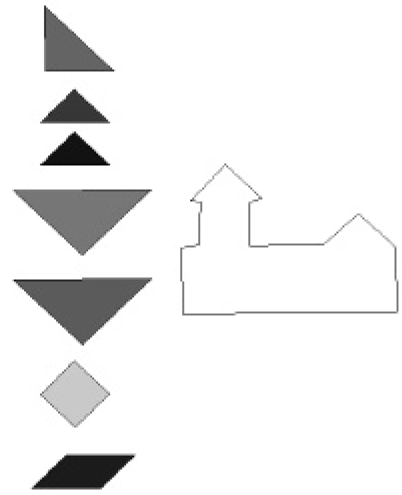


Figure. Part-stimuli training position with an opaque shape.

There were 5 stages in the experiment; training, testing of the groups initial trained task with old and novel stimuli, testing of a similar untrained task with old and novel stimuli, and two recognition tests placed after each main test. The stages of the experiment are described in more detail below.

Training Phase: The training phase was the same for each group; however, the part-stimuli training groups started each trial with the components displayed individually and the whole-stimuli training groups started with the components in the square position and then each subsequent trial the previously completed shape was the new starting position. Training contained two training blocks, block 1 and block 2, and 4 tangrams. The four tangram target shapes were presented in random order within each training block. Participants were instructed that this was just a training phase and they were asked to construct the target shapes using all 7 component shapes not overlapping by moving the components from the left side starting position to the target shape outline on the right side of the screen. The purpose of the training phase was to see how well participants learnt a tangram task depending on their training method and the category of shape used.

Test 1a: This was administered under the same conditions as the training phase. Participants were instructed that they would be tested on 4 tangram target shapes and they were to construct the target shapes as quickly as possible while maintaining high accuracy. They had previously seen 2 of the shapes in the training condition and the other 2 were novel shapes of the same category. The 4 target shapes were presented with the previously trained shapes first then the untrained stimuli. The purpose of this test was to examine transfer of knowledge of training shapes to similar problems, and whether there is a difference in performance between the two training groups.

Test 1b: A recognition test was then administered. Eight tangram shapes (4 previously seen from test 1a and 4 unseen) and 14 individual components (the 7 tangram individual components and 7 similar unseen non-tangram components) were individually presented in random order. The participant was asked whether they had seen the shape or component before. The objective for the recognition test was to establish the representation the participants had formed, whether it was different for the part-stimuli or whole-stimuli training groups.

Test 2a: During this testing phase the initial presentation of the tangram components was changed over, each training condition was tested on an untrained task. The part-stimuli training groups did the task as a whole; they now had the components presented in the basic square position. The whole-stimuli training groups were tested on the task by parts; they now had the components presented individually in a line down the left side of the screen. The basic instructions were still the same, participants were instructed that they would be tested on 4 tangram target shapes and they were to construct the target shapes as quickly as possible while maintaining high accuracy. They had previously seen 2 of the tangram shapes in the training condition; the other 2 were novel tangram shapes of the same category. The 4 target tangram shapes were presented with the previously trained shapes first then the untrained stimuli. The purpose of this test was to examine how participants perform with an untrained similar task on shapes they were trained on and more novel shapes to assess which group developed skill acquisition.

Test 2b: Another recognition test was administered the same as test 1b. Eight tangram shapes (4 previously seen from test 2a and 4 unseen) and 14 individual components (the 7 tangram individual components and 7 additional similar unseen non-tangram components) were individually presented in random order and the participant was asked whether they had seen the shape or component before. The objective for the recognition test was to establish the representation the participants have formed, whether it was different for the part-stimuli or whole-stimuli training groups.

Results:
All analyses of performance on the tangram task were completed using response times, the average time (s) it took to complete a tangram shape for each section. The average number of moves was also considered as during data collection a small number of participants had problems placing the components within the tangram outline. On various occasions even when they knew where the piece went it bounced back to the original place on the left side of the screen. For this reason it was necessary to make sure the average response time showed no obvious difference in pattern to number of moves for each experimental group. As there seemed to be the same pattern of results when considering response times or number of moves for all groups the bouncing back effect should not disrupt the analysis. Unless otherwise stated, an alpha level of .05 was used as the criterion for the analyses of results.

Training: The data were analysed using a mixed analysis of variance (ANOVA). Experimental condition (part-stimuli training or whole-stimuli training) and shape category (opaque tangrams or transparent tangrams) were used as between factors and the training block (block 1 and block 2) was the within factor in the analyses for the training phase.

All the groups' performance improved during training; there was a reduction in time and number of moves to complete a tangram shape. There appeared to be some differences in improvement rates of training for the four experimental groups. Both the whole-stimuli training groups, opaque tangrams and transparent tangrams, improved over the training phase by 199.12 s and 138.98 s respectively. The part-stimuli training groups improved slightly faster than the whole-stimuli training groups, transparent tangrams improved by 252.63 s from block 1 to block 2 and opaque tangrams improved by 468.36 s. Participants within the part-stimuli training condition with opaque shapes took the longest time in the first training block, (M = 651.54, SE = 81.4), followed by whole-stimuli training with opaque shapes, (M = 413.74, SE = 95.3), then part-stimuli training with transparent shapes, (M = 380.88, SE = 87.0). The most efficient group to begin with was the whole-stimuli training with transparent shapes, (M = 289.81, SE = 39.8). By block 2 of the training phase all groups seemed to be performing at the same level.

The mixed ANOVA supported all these observations. There was an overall main effect of shape category, $F_{(1, 28)} = 8.129$, $p = .008$. Participants constructing opaque shapes performed significantly slower than those constructing transparent shapes. The effect of experimental condition was non-significant. However, there appeared to be a trend $F_{(1, 28)} = 2.331$, $p = .138$ participants in the part-stimuli training condition performing slightly slower than the whole-stimuli training condition. There was no interaction.

Participants performed better in the second training block compared to the first. There was a significant improvement in response time across learning, reflected in a main effect of training block, $F_{(1, 28)} = 42.942$, $p < .001$. There was also a significant interaction between training block and experimental condition, $F_{(1, 28)} = 5.613$, $p = .025$. Further analysis was needed to see where the differences were, as 4 t-tests were needed to determine which groups were causing the interaction the alpha level was corrected to .0125, using the Bon Ferroni correction. Paired t-tests revealed a significant improvement for the part-stimuli training groups from training block 1 to block 2, $t(15) = 5.253$, $p < .001$ and a significant improvement across training blocks for the whole-stimuli training groups, $t(15) = 3.588$, $p = .003$. There was a bigger difference for the part-stimuli training group from block 1 to block 2 than the whole-stimuli training groups, hence the part-stimuli training group improved their performance significantly more than the whole-stimuli training group. Independent t-tests showed there were no significant differences between the part-stimuli training and the whole-stimuli training experimental conditions at training block 1 or 2. Both groups means were very similar in block 2 of the training phase, part-stimuli training M = 155.72 SE = 20.6 and whole-stimuli training M = 182.72 SE = 26.9, however for block 1 there seems to be a much larger differences between means, part-stimuli training M = 516.21 SE = 67.3 and whole-stimuli training M = 351.77 SE = 52.3. This supports the previous finding that part-stimuli training improved more than whole-stimuli training.

Testing: The response times for the testing phase were also submitted to a mixed analysis of variance (ANOVA), again experimental condition (part-stimuli training or whole-stimuli training) and shape category (opaque tangrams or transparent tangrams) were used as between factors and the task testing condition (part-stimuli test and whole-stimuli test) and the novelty of the testing tangram (old and new) were used as within factors of the analyses. The pattern of response time and number of moves for constructing the tangrams in each stage of the testing phase is illustrated in the figure below.

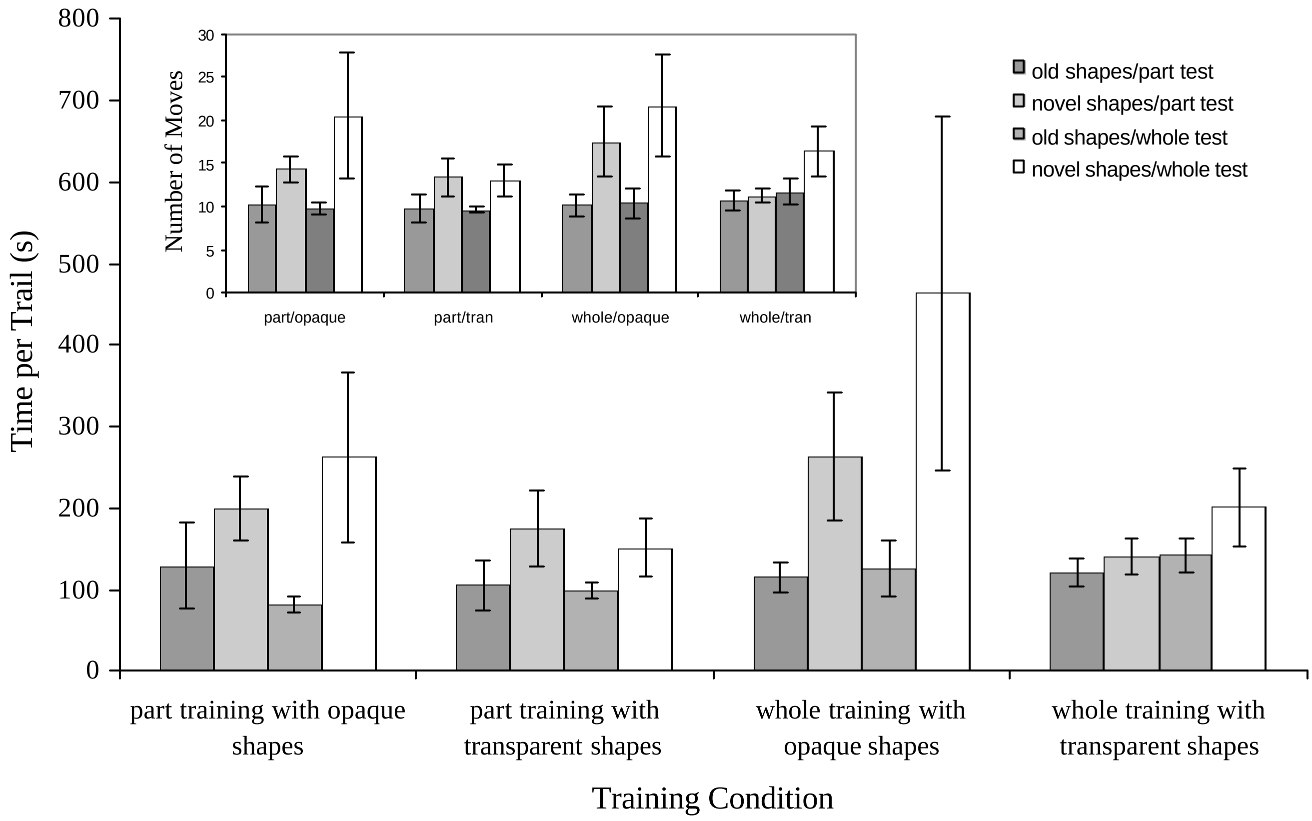


Figure. Mean response times (s) and number of moves (+SE) for each training group showing tangram performance at each stage of the testing phase.

Participants of the part-stimuli training group with transparent tangram shapes had the best performance after training; they performed most efficiently in the testing phase, i.e. had the lowest average response time (M = 265.08, SE = 40.46). The second most efficient group in the testing phase was the whole-stimuli training group with transparent tangram shapes (M = 312.48, SE = 24.33) closely followed by the part-stimuli training group with opaque tangram shapes (M = 337.39, SE = 68.86). The participants trained on whole-stimuli training with opaque shapes were the least efficient during the testing phase, having the highest average response time (M = 484.78, SE = 98.61). The results suggest there is not much difference in response time between the part-stimuli and whole-stimuli training groups.

The figure suggests that participants using transparent shapes were able to construct old and novel tangrams at a similar rate. Participants using opaque shapes seem to be less able to construct novel stimuli as quickly as those they had previously seen.

Overall performance: The mixed ANOVA supported all these observations. There was a main effect of shape representation which was marginally significant with participants constructing transparent tangram shapes demonstrating better performance than those constructing opaque tangram shapes, $F_{(1, 28)} = 4.134$, $p = .052$. The effect of experimental condition was non-significant, participants in the part-stimuli training condition performed similarly to the whole-stimuli training condition. There was no interaction.

Transfer: The ANOVA was also used to determine whether there were differences in transferability to untrained stimuli and untrained similar tasks.

There was a main effect of novelty, indicating participants performed better on tangrams they had completed before in training than novel tangrams, $F_{(1, 28)} = 26.916$, $p < 0.001$. There was also a near significant interaction between novelty of tangram shape and shape category, $F_{(1, 28)} = 3.384$, $p = 0.076$. Further analysis was needed to see where the differences were, as 4 t-tests were needed to determine what was causing the interaction the alpha level was corrected to .0125, using the Bon Ferroni correction. Paired t-tests revealed a significant difference between old and novel stimuli for the transparent tangram shape groups, $t(15) = -4.471$, $p < .001$, with lower response times and a better performance with old tangram shapes. Also there was a significant difference between old and novel stimuli for the opaque tangram shape groups, $t(15) = -2.861$, $p = .012$, again with lower response times and a better performance with old tangram shapes. Furthermore, there was a bigger difference across opaque shapes than transparent shapes with regard to old and novel tangrams, which indicate participants constructing transparent shapes were more able to transfer skills to new stimuli. Independent t-tests showed there was not a significant difference between transparent and opaque shapes with regard to old shapes and their means were very similar ($M = 116.66$, $SE = 10.54$ and $M = 113.24$, $SE = 14.56$ respectively). Again there was not a significant difference between transparent and opaque shapes with regard to new shapes but their means were very different ($M = 172.12$, $SE = 15.72$ and $M = 297.83$, $SE = 61.14$ respectively), it may just be a case of reduced power that has hidden a significant result.

There was no main effect of task; participants did not perform significantly worse or better on the part-stimuli test, where components were presented separately, or the whole-stimuli test, where the components were arranged in the previous completed tangram shape, even though participants had had specific training in one of either testing condition. However, a significant interaction was found between novelty of tangram shape and task being tested, i.e. part-stimuli testing and whole-stimuli testing, $F_{(1, 28)} = 6.633$, $p = .016$. Further analysis was needed to see where the differences were; again the alpha level was corrected to .0125, using Bon Ferroni correction. Paired t-tests revealed a near significant difference for the part-stimuli testing stage between old and novel stimuli, $t(31) = -2.537$, $p = .016$, with lower response time and a better performance with old tangram shapes. Also there was another near significant difference between old and novel stimuli for the whole-stimuli testing stage, $t(31) = -2.592$, $p = .014$, again with lower response times and a better performance with old tangram shapes. In addition, there was a bigger difference between old and new tangrams across the whole-stimuli testing stage than the part-stimuli testing stage, which indicate all participants performed better on the part-stimuli test stage. However because the part-stimuli test stage was only a novel untrained task for the whole-stimuli training group this could indicate that they are more able to transfer skills learnt during training to untrained similar tasks. Nevertheless this comment is not backed up by statistics, there was no interaction between task testing and training group.

Knowledge: D prime (d’) measures of sensitivity were calculated from the results of the two recognition tests giving four d’, components in recognition test 1b, tangram outlines for recognition 1b, components for recognition test 2b and tangram outlines for recognition test 2b. The four d’ were analysed using analyses of variance (ANOVA). Experimental condition (part-stimuli training or whole-stimuli training) and shape category (opaque tangrams or transparent tangrams) were used as between factors. There were no significant effects between training group and tangram shapes constructed, therefore this suggests there were no differences in knowledge of the parts used and the tangram outline shapes constructed. Consequently, it is assumed that all differences found are due to skills and strategies learnt, not knowledge that they have acquired specific to their training condition.

Overall summary: The figure below shows the progressive performance of the four experimental groups across all the different stages within the experiment. This graph is supportive to all the significant results found. All groups showed a major improvement during training, which extended to the testing phase. The part-stimuli training group with opaque shapes improved during training significantly more than the other groups. Overall response times decreased even more for all groups during test 1 and test 2 of old shapes. Even though the initial representation of the 7 component parts had changed for both part-stimuli training and whole-stimuli training in test 2 they were still able to improve performance by reducing their average response time. Therefore both training groups can be acknowledged to transfer their skills to an untrained task. When transferability was considered for untrained stimuli the only group to have a significant increase in response time was the whole-stimuli training group with opaque tangram shapes, the other groups performance increased somewhat but not by a substantial amount.

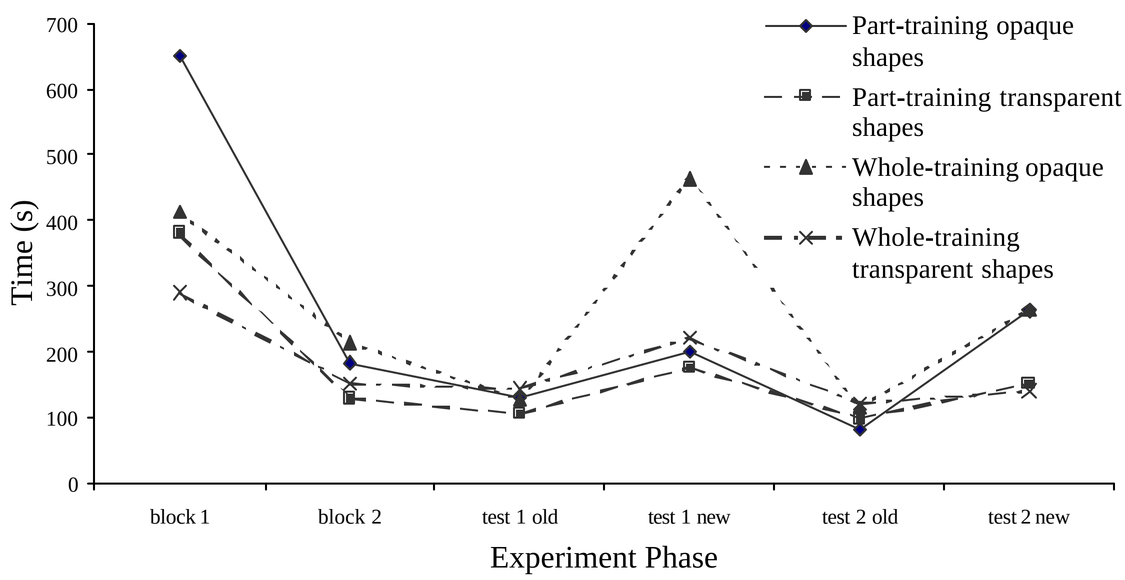


Figure. A learning curve to show the response times (s) for each training group over the experiment phases.

Finally when transferability of skills to an untrained task with untrained stimuli was considered a significant difference was not found between training conditions but there was a significant difference between groups that constructed opaque shapes and those that constructed transparent shapes. Participants constructing transparent shapes were more able to transfer skills to a similar task with novel stimuli than the other groups, where as participants constructing opaque shapes showed a larger increase in response times on transfer to a novel task with novel stimuli; this advantage was not dependant on training condition.

Overall it was the transparent tangrams that were better for skill acquisition of this visual spatial task. Participants constructing transparent tangrams had a lower response time in the training phase. They also had better performance after training than the other groups, i.e. the lowest response time for the testing phase. Also transparent tangram experimental condition groups were more able to transfer skills to new stimuli and similar untrained tasks, showing better transferability. These advantages were not dependent on training condition.

Discussion:

This experiment examined which method of training, part-stimuli training or whole-stimuli training, was better for skill acquisition of a visual spatial task, the Tangram Game, and whether the category of stimuli had an effect on performance. In relation to skill acquisition the experiment considered performance during training, performance in a testing condition and transferability to an untrained task and untrained stimuli to determine which experimental groups' performance was optimal. A reduction in response time to complete a tangram indicates that participants are more efficient at the task. Also smaller discrepancies between trained and novel stimuli indicate more transferability, and hence better skill acquisition.

During training, part-stimuli training improved significantly more than whole-stimuli training across learning blocks. This may give the impression that part-stimuli training is more efficient than whole-stimuli training. However, both training groups ended the training session with similar response times; hence it could also reflect that part-stimuli training was a more difficult task to begin with than its counterpart. Whole-stimuli training participants are more exposed to the criterion task making the task easier, during each tangram trial they are exposed to different possible ways of how the 7 component parts fit together. This result supports the findings of Singley and Anderson (1989), for this visual spatial task exposure to the whole-stimuli provides a template for the development of whole-task specific skills and hence is most efficient early on in training.

It has also been suggested that the component parts of the tangram game are highly interrelated. There are very few ways that the component parts can fit together, all of which are first determined by the three ways the two large triangles fit together. Organisation of the tangram component parts is high, therefore according to Blum and Naylor (1968) whole-task training should be relatively more efficient. This is true for this study, participants of whole-stimuli training do perform relatively quicker in completing the tangrams early on in training. This may be because participants are exposed to the high interrelatedness hence they perform more efficiently to begin with and as trials go on participants of part-stimuli training learn this interrelatedness and their performance improves significantly.

For the most part, there were no significant differences between part-stimuli and whole-stimuli training over the entire experiment. Neither group performed significantly worse or better, because of their training, during the testing phase. It is evident from the results that both training groups were able to transfer knowledge to an untrained task and to untrained stimuli. The findings of this experiment suggest the method of training a visual spatial task has no effect on improved performance or skill acquisition. These findings are supported by previous research of other tasks that have also found no advantage for part-stimuli training (Cox & Boren, 1965, cited in Stammers, 1982, Whaley and Fisk, 1993). Goettl and Shute (1996) also found no initial advantage for part-stimuli training. However, they felt their components for part-stimuli training did not adequately represent the critical skills underlying the criterion task. The critical skills in efficiently completing a tangram trial were either placing the two large triangles in one of three possible combinations or making sure the odd parallelogram was placed early on in construction. However it seems evident that neither the whole-stimuli training condition, by giving an example of how the parts could fit together, or the part-stimuli training condition, by not restricting possible combination of parts, gave way for better skill acquisition to develop. In actual fact it seems evident that the shape category influenced acquisition of these critical skills more.

Overall during training, as well as during the testing phase, there was a significant difference between the response times of the transparent tangrams and the opaque tangrams. Throughout the experiment participants with the transparent tangrams out performed the participants with opaque tangrams, as they continuously had the lowest response times. This may be due to simplification. The tangram task was split by shape category; instead of constructing a variety of shapes participants of this experiment only constructed a similar set of tangram shapes. It could be assumed that only completing transparent tangrams is a simplification of the whole tangram task, as in all trials participants are able to identify 4 of the individual component parts from the target outline making the task easier. In line with Wightman and Lintern’s (1985) idea of simplification completing transparent tangrams allowed positive transfer of knowledge to similar tasks. The positive transfer of knowledge may have also produced enhanced skill acquisition.

Skill acquisition is influenced by comprehension of stimuli specific knowledge and strategic skills of the task (Doane et al., 1999). According to Brown and Kane (1988) strategic skills optimise performance of a task, it may be that because transparent tangrams are easier they allow participants to identify the strategic skills, as they see the possible answers very clearly from the tangram outline. On the other hand participants using opaque tangrams see very little in their tangram outlines hence they may use more stimuli specific knowledge they have from completing other tangrams. Again this idea supports the experiment’s results, practice with transparent tangrams is easier than opaque tangrams, which allows development of strategic skills and stimuli specific knowledge and hence improves skill acquisition.

It is evident throughout the experiment that transparent tangrams have developed better skill acquisition than opaque tangrams. The results show an interaction between shape category and novelty of tangram. Participants who constructed transparent tangrams had smaller discrepancies between trained and novel stimuli than the participants of opaque shapes. This indicates participants of transparent tangram were more able to transfer skills than those of opaque tangrams. This finding can also be explained by theories of skill acquisition. The component fluency hypothesis by Carlson et al. (1990) suggests that there are capacity limits placed on cognitive resources and that the fluency of the component skills is critical to skilled performance on complex tasks. As transparent tangrams have been described as being easier than opaque tangrams, participants may have more cognitive resources to identify and develop the critical component skills, which are either placing the two large triangles in one of three possible combinations or making sure the odd parallelogram was placed early on in construction, and hence have the skills to perform better on tasks they have not been trained on.

These results can also be explained more simply by the procedural reinstatement theory (Clawson et al., 2001). It may be that transparent tangrams are procedural tasks; participants are able to see where most of the component tangram parts go, hence the same procedures are used at retention as during training. Opaque tangrams are said to be more difficult and hence use predominately declarative knowledge, these learning procedures are more indirect and result in poor performance at retention.

During testing all participants, regardless of their training condition, were tested on the task as a whole-stimuli and also as part-stimuli. Participants did not perform significantly worse or better on the part-stimuli test or the whole test, even though they had had specific training in one of either testing condition. There was, however, an interaction between the task being tested and novelty of tangram. All participants performed significantly better on the part-stimuli test; there was less of a discrepancy between trained and novel stimuli than for the whole-stimuli test. This indicates more transfer of knowledge during the part-stimuli test. However because the part-stimuli testing stage was only a novel untrained task for the whole-stimuli training group this could indicate that they are slightly more able to transfer skills learnt during training to an untrained similar task. It may be the fact that as whole-stimuli training is exposed to the criterion task early in training they have learnt the critical skills; hence they are more effective in component practice (Carlson et al., 1990). This result may also be influenced by the fact that it is easier for the part-stimuli training group to do the part-stimuli test, as they are not influenced by the tangram example given in the starting position of the whole-stimuli test. However, there are only minor differences between training groups and there is no statistical evidence to suggest whole-stimuli training is significantly better for skill acquisition than part-stimuli training. Therefore, it must be assumed that this result is just an effect of chance. If the statistical power of the experiment is increased there might be an increased possibility of a three-way interaction between training condition, task being tested and novelty of tangram to determine skill acquisition.

It is proposed that the findings of this experiment support Stammers' (1982) favourable use of whole-task training. To start part-stimuli training did offer some savings, as there was a greater improvement during the training phase for this condition. Nevertheless, just as Stammers commented, part-task training also places greater demands on the trainee, in this experiment participants had to learn the high interrelatedness of the component parts whereas with whole-stimuli training they were able to see possibilities of the components' organisation. Finally even though there are only small differences between the groups of part and whole stimuli training it is better to use whole-stimuli training as it avoids isolating key skill elements from context that could be distorted in part-stimuli training.

Overall it is the transparent tangrams that are better for skill acquisition of this visual spatial task. Participants constructing transparent tangrams have a lower response time in the training phase. They also have better performance after training than the other groups, i.e. the lowest response time for the testing phase. Also transparent tangram experimental condition groups are more able to transfer skills to new stimuli and similar tasks, showing better transferability.

The results of this experiment may prove to be more significant if the statistical power of the design was increased. The small differences between the part-stimuli and whole-stimuli training condition might be due to the low number of participants and some high variations in response times within each experimental group. It is thought for further investigation if the size of experimental groups were increased a more evident pattern would emerge. On the other hand it could be just that the visual spatial task used, the tangram game, is just too short or simple to warrant being broken into parts. Research on other visual spatial tasks could be used to support or even contest these findings. One final point that Stammers (1982) addresses in his review is that with the limited research on training methods it is not yet clear what is the optimal point to change from part-task to whole-task training. It may be that in this experiment the change over to a different testing condition was introduced quite late. Further research could examine different combinations of part and whole task training to see if this influences the acquisition of critical component skills.

To conclude the research reported here suggests there are no advantage for either part-stimuli or whole-stimuli training, if anything, it should be whole-stimuli training that is advised for the Tangram Game purely on the basis of keeping all the component skills together. What is of greater importance is the significant effect that category of tangram has on overall performance and skill acquisition of the task. It is this area that opens up possibilities of further research, for example can trainees of transparent tangrams transfer their strategic skills and stimuli specific knowledge to perform efficiently with opaque tangrams.

References:

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, *77*, 598-614.

Blum, M. L. & Naylor, J. C. (1968). *Industrial Psychology*. New York: Harper and Row.

Brown, A. L. & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, *20*, 493-523.

Butler, B. E. (1994). Spatial puzzles: A guide for researchers. *Canadian Psychology, 35,* 47-65.

Clawson, D. M., Healy, A. F., Ericsson, K. A., & Bourne Jr. L. E. (2001). Retention and transfer of Morse code reception skill by novices: part-whole training. *Journal of Experimental Psychology: Applied, 7 (2)*, 129-142.

Carlson, R. A., Khoo, H. & Elliot, R. G. (1990). Component practice and exposure to a problem-solving context. *Human Factors, 32*, 267-286.

Doane, S. M., Alderton, D. L., Sohn, Y. W. & Pellegrino, J. W. (1996). Acquisition and transfer of skilled performance: Are visual discrimination skills stimulus specific? *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1218-1248.

Doane, S. M., Sohn, Y. W., & Schreiber, B. (1999). The Role of Processing Strategies in the Acquisition and Transfer of a Cognitive Skill. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1390-1410.

Goettl, B. P., & Shute, V. J. (1996). Analysis of Part-Task Training Using the Backward-Transfer Technique. *Journal of Experimental Psychology: Applied*, *2*, 227-249.

Singley, M. K. & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Stammers, R. B. (1982). Part and whole practice in training for procedural tasks. *Human Learning, 1*, 185-207.

Whaley, C. J. & Fisk, A. D. (1993). Effects of part-time training on memory set unitisation and retention of memory-dependent skilled search. *Human Factors, 35*, 639-652.

Wightman, D. C., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors*, *27*, 267-283.

STUDY 3: TIME PRESSURE

Abstract:

Much research has been done on decision making under time pressure, but how to train for these situations has so far been ignored. In the following study, it was investigated whether the best method is to be trained on the task to be tested on under time pressure or without, or whether training under time pressure in an alternate task with time pressure followed by training on the tested task would be most useful. A new paradigm was used in test call tangrams (puzzles made up of seven geometric shapes). Results were largely non-significant in comparison between training groups, but analysis of the groups separately showed that when trained under time pressure, there is a large degree of specificity for what is learnt, whereas in the other conditions more general rules are acquired. This rule acquisition is on a continuum, where stimuli most similar to those in training are the best performed. Additionally, it appears that giving separate exposure to time pressure and to the task to be tested on is the most advantageous method of training. The conclusion is made that training for decision making under time pressure is indeed useful to enhance performance in test, and does not merely represent a learning curve from exposure to the same stimuli. Suggestions for progress of this research are given.

Introduction:

Increasing in today's society is the number of decisions that must be made under a limited amount of time. Some situations many of us experience on a regular basis, such as choosing a route when driving, or what to order in a busy restaurant, that have little consequence if the wrong choice is made. However, there are also a number of important political and business and even personal decisions made when time is limited that can have serious consequences if a unsuitable judgment is made (see Neck, & Moorhead, 1995). The effects that time pressure has on decision making have been studied widely by a number of researchers (Bowden, 1985; Franklin, & Hunt, 1993; Johnston, Driskell, & Salas, 1997; Joslyn, & Hunt, 1998; Lerch, & Harter, 2001; Romano, & Brna, 2001; Verplanken, 1993). There is a large gap in the research however with regards to how to train people for time pressure situations. With an estimated $60 billion being spent by U.S. companies on training programs in 1997 (Bartel, 2000), it is remarkable that there is no formal evidence that can be used to train people for the time pressure they are so often exposed to.

To understand what might be an appropriate method to train people for time pressure situations it is important first to discuss what is already known about the effects of time pressure on decision making. Three strategies have been identified as being used when trying to cope with time pressure, first proposed by Miller in 1960. One strategy is filtration (Ben Zur, & Breznitz, 1981) whereby what is subjectively viewed as the most important information is processed first. This is a compromising strategy as a certain amount of information is ignored to reduce the limitations created by the time pressure. A second strategy is to process information at a faster rate (Benson, & Beach, 1996; Kerstholt, 1994). Thus, the decision process happens in an abbreviated form but the task is completed in the time required. The third possible strategy adopted is avoidance (Benson, & Beach). This results in random choices being made or decisions made on momentarily prominent aspects of the situation. Therefore, time pressure alters the strategies used to make decisions from those used under normal conditions.

The above information on strategy alterations refers to both time pressure and time constraints. There is however an important difference between the two (Ordóñez, & Benson, 1997). When there is a time deadline that a task has to be completed by this is referred to as *time constraint*. *Time pressure* refers to situations in which either an awareness of the amount of time spent completing a task or the impending deadline create a feeling of stress. Thus, theories explaining the effects of stress must be considered in this situation.

A robust finding when arousal (e.g. stress) is related to performance is called the Yerkes-Dodson law (Hardy, & Hayes, 1996).

Arousal, i.e. stress, increases performance increases until it reaches a peak. After this point, any further increase in arousal will be detrimental to performance. Therefore overall an inverted U-shape is formed. Easterbrook (1959) explains this relationship in his cue-utilisation hypothesis.

According to the theory, as arousal increases the number or range of cues attended to progressively decreases. To an extent, this allows an improvement in performance as the most essential information is attended to, while information that is more peripheral is ignored. There comes a point however where by attention becomes excessively focussed and even relevant cues are excluded from information processed.

It can be seen how theories relating to decision-making under time pressure can be related to this. Consider a faster processing strategy: by trying to cope with extreme time pressure, processing may be skipping steps that are in fact important for effectively making a particular decision. This would be reflected by performance being beyond the optimum point, resulting in a decrease in performance represented on the right had portion of the inverted-U. This proposal would also apply to filtration and avoidance were they to occur to an extent that is detrimental. Hence, the degree to which the time pressure is experienced as arousing and stressful will alter how it affects performance.

As time pressure alters attention to information, it will therefore interact with working memory. Working memory is defined as a system that both processes information and the temporary storage of that information, and can be engaged in a wide range of cognitive tasks (Eysenck, 1990). The importance of working memory in decision-making has been supported by neurological studies. In a neuroimaging study (Paulus, Hozack, Zauscher, McDowell, Frank, Brown, & Braff, 2001) where participants were required to make predictions about stimulus occurrence, it was found that participants in a decision making task had a more active prefrontal cortex than those in a control condition. Other neuroimaging studies have shown working memory to be centred in the prefrontal cortex (Cohen, Perlstien, Brave, Nystrom, Noll, Jonides, & Smith, 1997), therefore suggesting the two are related. Furthermore, the study of damage to the ventromedial prefrontal cortex has shown that decision making suffers as a result (Bechara, Dolan, Denburg, Hindes, Anderson, & Nathan, 2001), thus supporting the relationship between working memory and decision making.

There are three demands made on working memory during the completion of any complex task, such as decision making (Carlson, Khoo, Yaure, & Schneider, 1990). The first is temporary storage of information, as proposed by Atkinson and Shiffrin (1968) in their model of short term memory. The second demand is active manipulation of information that restricts resources through reducing the processing resources available (Navon, 1984). The third is integration of processes or representations where information needed to complete the task is maintained in working memory at the same time as new information is combined with it (Carlson, Sullivan, & Schneider, 1989). It is widely recognised that working memory has a limited capacity (Baddeley, 1999) such that any of the above demands can overload the resources available, and performance can suffer as a result.

One method of coping with this limited resource is to "chunk" information (Hockley, & Lewandowsky, 1991). A chunk is defined as "a collection of elements having strong associations with one another, but weak associations with elements within other chunks" (p.236, Gobet, Lane, Croker, Cheng, Jones, Oliver, & Pine, 2001). As working memory has a fixed capacity, chunking allows a number of information points, which would otherwise take up a large portion of the working memory span, to instead act as only one piece of information. An example of this is letters being chunked as meaningful acronyms or abbreviations (Bower, & Springston, 1970). Therefore when presented with such letter combinations as "IBM", "CIA" or "BBC" they are encoded not as the separate letters that they are composed of but as the trigram of letters. When considering a decision making task, to chunk information it may be that only the solution in the task is remembered as opposed to the series of steps that lead to that decision. For example, when making a decision about which direction to take, it may only be remembered that turning left was the correct direction, as opposed to remembering that the left turn was chosen for a multiple set of reasons, such as it was the longest route, but the most scenic etc. Thus, the decision process may be encoded in different ways according to the resources available at that time.

As an excessive amount of information can affect performance, time pressure may be an interfering factor. When information is required to be processed in a lesser amount of time, more information must be manipulated in a shorter period of time (Kellogg, Hopko, Ashcraft, 1999). In addition, attending to the passage of time in the desire to try and complete the task as quickly as possible will also put a strain on working memory. That is, concentration on the passage of time is another piece of information being maintained in working memory, putting further strain on the resources available. Therefore, time pressure will put strain on working memory. This may cause strategies that allow more information to be processed concurrently, such as chunking, to take place in addition to the three previously mentioned changes made to processing when decisions are made under time pressure.

It is important to consider other aspects that will effect decision making that are of ecological importance, such as affect. Zajonc (1980) argued that emotional responses are often the most automatic and consequently guide behaviour and cognitions. Even small events can generate a transient affective state (Mittal, & Ross Jr., 1998), such that although experiments such as this one will not generate high levels of arousal, decision-making tasks may well induce a level of stress that will affect performance. Furthermore, as partly reflected by the Yerkes-Dodson law (Hardy, & Hayes, 1996), emotions can change cognitions, often for longer than the event that provoked the emotion (Lerner, & Keltner, 2001).

As previously mentioned, time pressure induces a feeling of stress and it is considered by some that stress is an inherent aspect of time pressure if it is truly going to be perceived: "Time pressure is the subjective perception of stress" (p. 243, Benson 3rd, Groth, Beach, 1998). It is proposed by Maule and Hockey (1993, as cited in Maule, Hockey, & Bdzola, 2000) that by imposing time pressure, in fact a number of affective states can be induced.

In turn, emotions can affect decision making, as has been noted by a number of authors (for a review see Loewenstein, Weber, Hsee, & Welch, 2001). In a study of risk taking in a strategic decision making situation, positive affect was found to make less risky decisions than those with negative affect (Mittal, & Ross Jr., 1998). This is interpreted using the information processing perspective, saying positive affect leads to information being processed more heuristically or strategically, whereas negative affect causes information to be processed systematically, i.e. less efficiently. To explain these kinds of results, a theory has been developed by Antonio Damasio called the somatic-marker hypothesis (1994). This hypothesis proposes that impaired decision-making will in part be due to a defect in emotion and feeling and that there is a somatosensory pattern that marks a scenario as good or bad. The marker then has a non-conscious influence creating a bias in later cognitive processes. Evidence for his theory has come from various studies involving patients with ventromedial prefrontal cortex (VM) damage. Firstly, in studying their deficits in decision making have been studies using a task which involves four desks of cards, named A, B, C, and D (Bechara, Damasio, A.R., Damasio, H., & Anderson, 1994). In this task, in which the goal is to maximise profit of play money, decks A and B give the greatest reward ($100 per card) but also the greatest loss when a punishment is given, such that a total loss of $250 is incurred every 10 cards. Decks C and D, on the other hand, give a lesser reward ($50 per card) but the punishment is such that after ten cards there is a gain of $250. Therefore, although initially it appears that decks A and B are better choices, over time decks C and D are actually more advantageous. Normal controls when given this task gradually learn to choose cards from C and D (the good decks) where as VM patients persist in picking more cards from A and B (the bad decks). This demonstrates that the VM patients are unable to learn from their mistakes and their decisions remain disadvantageous to them (Bechara et al., 1994). Thus, decision-making processes are thought to originate in prefrontal cortex as well.

To study the if biases are formed as a result of exposure to reward and punishment through the decisions made, normal control subjects and VM patients were given the same task again but this time skin conductance response (SCR) activity was recorded in the 5s before a card was chosen (Bechara, Tranel, Damasio, H., & Damasio, A.R., 1996). For normal controls, it was found that the anticipatory SCRs increased as the task progressed, i.e. with more exposure to the good deck/ bad deck pattern, and became more pronounced before choosing cards from the bad decks. The anticipatory SCRs were absent from VM patients. This indicates that the outcomes of decisions become learnt and that there is a concept of positive or negative affect related to them which effects what decision will be made. This affective bias will allow normal subjects to avoid disadvantageous decisions being made in the future (Bechara, Damasio, A.R., & Damasio, H., 2000), and occurs even before subjects are aware of the "goodness" or "badness" of the various decisions.

Therefore, the relationship of affective markers in training for decision making under time pressure may be the following: due to the stress caused by time pressure, the methods learnt in training to make the decision, regardless of how successful they are, will be associate with a negative marker. Therefore, it will be hard to access this method in subsequent testing trials where the skills would be advantageous if recalled.

Thus, with all the above information in consideration, it is difficult to make any concrete predictions about what the outcome of various training methods will be when in comes to testing decision making under time pressure. But aims can be identified. A main aim of this study is to see if there is an optimum method in which to train participants for decision making under time pressure. Secondly, to see if there is any difference in what participants learn given the different training conditions. And third, to see if there is in fact any need for training to deal with decision making under time pressure.

To study the effects of different training conditions on later performance in decision-making under time pressure, a paradigm was used that has rarely been employed in this kind of context. The task involved solving a puzzle known as a tangram. Tangrams are spatial puzzles composed by seven pieces (triangles, squares, diamonds etc.,) that are arranged into larger figures. These seven pieces can be used to make a wide range of other shapes, and indeed only a small portion of the tangrams available were piloted in this experiment, which totalled around 150. Although spatial puzzles, such as tangrams are popular as toys, they have rarely been used empirically (Butler, 1994). Rather, when studying problem solving or decision making conceptual puzzles have been more commonly used. Tangrams do not appear to have been used in decision making studies previously, although one study (Henderlong, & Paris, 1996) unrelated to this field has employed them, thus giving justification to their use here.

Three different training conditions were used: (a) training under time pressure with the tangrams, (a) training without time pressure with the tangrams, and (c) training under time pressure with a different task in addition to training with tangrams without time pressure. First, by comparing the two groups trained only on the tangrams, it can be seen the degree to which it is important to have exposure to the task to be tested on in the same conditions as in test, therefore allowing specific skills to be learnt that will be most advantageous at test. Alternatively, specific training may in fact be a detrimental method, as insufficient information will be extracted during training due to the constraints of working memory, and that somatic markers will make learnt skills less accessible. Therefore, by training people under time pressure in an alternate task, one that also involves decision making, it can be seen whether simple exposure to time pressure itself, without any association to the task to be tested on, allows skills to be learnt concerning how working memory should cope with time pressure. Then, by experiencing the tangrams without time pressure, a sufficient amount to time is given to learn more about the task, but will have no negative associations with the methods that have proven successful.

An aspect of the tangrams in specific that will be useful in this experiment is that there are different degrees of understanding of the puzzles. This is to say that they can either be solved by considering the tangrams holistically, or an underlying structure can be identified that aids placement of the remaining shapes. Of the shapes that make up the tangrams, the two largest triangles can be placed in only limited number of ways. In training, only one of the underlying structures is used, thus exposing the participants to only one way of how they should be placed. In test, the participants are then given tangrams they have already solved, novel tangrams with the same underlying structure, and novel tangrams with a new underlying structure. Therefore, the methods learnt to complete the tangrams can be seen by the length of time it takes to complete the novel tangrams. And this too can allow study of the necessity of training itself, in that if any skills that can be applied to novel situations are learnt, it will represent more than a simple learning curve, in which case repeated exposure to the task would be sufficient to enhance performance. Therefore, if it is found that training under time pressure with the tangrams leads to improvement in the amount of time taken to complete tangrams seen repeatedly but there is no difference when given novel tangrams, then there is no use in training. However, if one of the other methods of training gives better results, or if training under time pressure allows a significant improvement in completing novel tangrams in comparison with first exposures to the task, then training is an important aspect to consider with decision making under time pressure.

Method:

Design: This experiment was of mixed design. The independent variables were: (1) training type, with three levels of (a) training with time pressure using tangrams, (b) without time pressure using time pressure, or (c) with an alternate task with time pressure (between-subjects variable); and (2) stimulus type in test, having three types being old stimuli, new stimuli of the same category and new stimuli of a novel category (within-subjects variable). The dependent variable was amount of time taken to complete the tangrams, measured in seconds.

Participants: The 34 participants were recruited from the psychology department at the University of Southampton and received credits for participating. Of the 34, 20 were female and 14 male. Their ages ranged from 18 to 33, with a mean age of 20 years 10 months. No participants reported familiarity with tangrams and hence were regarded as naïve to the task.
Materials: Twelve tangrams were selected according to criteria identified through pilot studies outlined below. Tangrams could be completed using a mouse and the pieces could be rotated as desired using a scrolling button on the mouse. The stimuli were programmed in both training and test phases such that they were fully counterbalanced.

The alternate task used a sample of multiple-choice questions from the standard GRE examinations sampled from a publication by the Educational Testing Services, 2001. The GRE was chosen as the alternate task as it is also a decision making task but does not involve any spatial ability, unlike the tangrams. Therefore, it is similar to the tangrams enough to give participants experience with decision making but not specifically with a spatial task. Both verbal and mathematical questions from the GRE were used, 16 of each . This number was chosen, as it would take around the same amount of time to complete as the 12 tangrams.

An electronic clock was placed on the screens of both the tangram trials and the GRE questions. In the time pressure conditions, the clock was 20% faster than normal time (i.e. when the clock showed 10 seconds had passed only 8 real seconds had passed). In the no time pressure, condition the clock was slowed by 20%.

Additional tangrams and pieces were selected through piloting, again outlined below, for a recognition test requiring simple yes or no responses.

The tangram trials, GRE trials and the recognition were all program using C++ Builder for Win 32 and the program shell was made by Martin Hall and Jin Zhang. The computer used was and IBM compatible personal computer with a 19-inch monitor and screen resolution of 1024 by 920.

Pilot Studies: To validate the tangram task a series of pilot studies were undertaken.
1. Over 100 tangrams were tested in piloting to determine the approximate time required to complete them. From this those taking between two and seven minutes to complete were identified for use.
2. The tangrams were categorised by the underlying structure formed the two largest triangles that make up the tangram. Two variations on the underlying structure were used.
3. The tangrams were then categorised as being symmetric or asymmetric. Only those that were asymmetric were used. Using these criteria, 12 tangrams were chosen, six from each of the categories.
4. For use in the recognition tasks, it was necessary to pair each tangram with a tangram not to be used in any of the training or testing trials. These were to be paired on the basis of similarity, so each tangram to be used for the trials was compared with six other tangrams in a pilot study. These ratings were compared with ratings for the pieces.
5. The pieces that make up the tangrams were also compared with other shapes, in this case with 36 other shapes. Rating values were then used to find pieces and tangrams with analogous ratings to be used in the recognition test. In all cases, the highest possible values of similarity were used. Therefore, all the shapes used in the recognition task were rated as highly similar to those used in training.

Procedure: All participants were first asked to complete two tangrams to acquaint them with the task. Once this was completed, the participants were asked if they had ever had any experience with tangrams and that they were confident how the task worked. Were anyone to have responded that they did have previous experience with tangrams they would have to be dismissed, but in this experiment, it was not necessary.

Participants were randomly assigned to one of three training conditions: (a) training with tangrams under time pressure, (b) training with tangrams with no time pressure or (c) training under time pressure in an alternate task. In this third condition, the participants were first given the GRE test of 32 questions under time pressure followed by the tangram training without time pressure. Therefore all participants had equal exposure to tangrams in training. This consisted of 4 tangrams from one category only being presented twice (i.e. 8 trials total). Before all of the training conditions, a script appropriate to the condition was read to the participants to either encourage completing the tasks as quickly as possible (i.e. to help induce time pressure), or to give a sense of freedom of time.

After tangram training was completed, a recognition task requiring the participants to answer simply yes or no as to whether they had seen the object in training. There was no mention of time in this phase.

In testing the participants were given 6 tangrams. All participants were tested under time pressure and relevant script was read to them. The first two tangrams presented had been seen in training. The following two tangrams were novel tangrams of the same underlying structure. The last two were novel tangrams from a novel category.

The measurement taken for all tangram tasks was time taken to complete the tangrams. This was the data used in analysis.

Following the test phase, all participants were debriefed and told who they could contract should they have any questions.

Results:

Before analysis began, one participant's data was excluded, as they did not complete the task fully. Outliers were identified using box plots of the data and any data points more than 2 ½ standard deviations from the mean once those outliers had been removed were also deleted from the data set. From the first exposure in training, an average of 1.58 data points were removed per trial per group. From the second exposure, an average of 1.75 data points were removed per trial per group. At test, an average of 1 data points were removed per group from old shape completion times, and also from data from novel tangrams of the same category. For novel category data, an average of 1.67 points were removed per group.

To examine completion times during training, learning curves were created for both stimuli exposures. The averages for both exposures were calculated for comparison. For the first exposure, those under time pressure had a mean of 150.67s (SD = 60.57s, N = 12), those without time pressure had a mean of 217.16s (SD = 110.01s, N = 13), and those trained in the alternate task had a mean of 210.86s (SD = 91.68s, N = 9). For the second exposure, the time pressure group had a mean of 111.89s (SD = 38.59s, N = 12), the without time pressure group a mean of 101.99s (SD = 32.52s, N = 13) and the alternate task group a mean of 108.69s (SD = 46.89s, N = 9).

Means were compared to see if times differed for each exposure in training with a 3 x 2 ANOVA. As with all the analyses to follow, calculations were made at the .05 significance level. It was found that the main effect of exposure was significantly faster in the second than the first, $F_{(1,31)} = 26.53$, $p < .001$, but that the main effect of training type was non-significant, $F_{(2, 31)} = 1.23$, p = ns, with no interaction, $F_{(1, 31)} = 2.20$, p = ns. Analysis was then broken down to look at each training condition to see if the significance varied among them. The repeated measures analysis for the group trained under time pressure found that the difference in time of completion approached significance, $F_{(1, 11)} = 4.49$, $p = .06$. Significant differences were found in the group trained without time pressure, $F_{(1, 12)} = 11.342$, $p < .01$, and the alternate task group, $F_{(1, 8)} = 13.46$, $p < .01$.

Comparisons were then made between data from the second exposure at training and data from the first phase of testing, that used stimulus seen previously in training. Overall, times from the second exposure in training (M = 107.26, SD = 37.85) were significantly faster than those in testing (M = 82.22, SD = 28.06), $F_{(1,28)} = 15.05$, $p < .001$. However there was no significant effect of training type, $F_{(2, 28)} = 0.60$, p = ns, or any interaction, $F_{(2, 28)} = 0.43$, p = ns). Again, analysis was made for each of the training groups. The time pressure trained group had significant differences between the second exposure in training and in test (M = 88.72s, SD = 25.48s), $F_{(1, 11)} = 5.83$, $p < .05$, as did the group trained in the alternate task (test: M = 67.30, SD =

23.88), $F_{(1, 7)} = 6.41$, $p < .05$ There was no significant differences for the group trained without time pressure between training and test ($M = 85.98$s, $SD = 23.88$s), $F_{(1, 10)} = 2.27$, $p < .05$

The means and standard deviations for the trials using stimuli seen in test are listed above. For the trials using novel stimuli from the same category as seen in test, the mean for the time pressure trained group was 154.23s ($SD = 80.36$s), the no time pressure group 135.81s ($SD = 93.55$s) and the alternate task group 115.99s ($SD = 71.37$s). In the trials that had novel stimuli from a novel category, the time pressure group had a mean of 175.45s ($SD = 68.76$s), the no time pressure group a mean of 162.01s ($SD = 114.65$s) and alternate task group a mean of 147.19s ($SD = 83.01$s).

Repeated measures analysis of composite means for data from all three training types was made first to see the differences between the times for the stimuli types. The decrease in speed of completion from old stimuli ($M = 82.22$s, $SD = 28.06$s) to novel stimuli from the same category ($M = 81.48$s, $SD = 14.88$s) was significant, $F_{(1, 27)} = 14.46$, $p < 0.001$, as was the decrease in speed in novel stimuli from a novel category ($M = 162.56$s, $SD = 89.80$s), $F_{(1, 26)} = 18.64$, p < .001. However there was no significant increase in speed between novel stimuli from the same category and from a different category, $F_{(1, 24)} = 0.95$, p = ns.

The data was then split up by the training conditions. For those trained under time pressure there was a significant difference between the three stimuli types, $F_{(2, 18)} = 5.36$, $p < .01$. There was a significant decrease in speed of completion from old stimuli ($M = 88.72$s, $SD = 25.48$s) to novel stimuli of the same category ($M = 154.23$s, $SD = 80.36$s), $F_{(1, 9)} = 17.60$, $p < .01$, and again when compared with novel stimuli from a different category ($M = 175.45$s, $SD = 68.76$s), $F_{(1, 11)} = 9.26$, $p < .01$. There were no significant differences between novel stimuli from the same category and from a novel category, $F_{(1,9)} = 1.07$, p = ns.

For the data from the group trained without time pressure, there was no significant difference in time between the old stimuli completion times ($M = 85.98$s, $SD = 23.88$s), completion times for novel stimuli from the same category ($M = 135.81$s, $SD = 93.55$s) and completion times for novel stimuli from a novel category ($M = 162.01$s, $SD = 114.65$s), $F_{(2, 12)} = 0.88$, p = ns.

For those trained in the alternate task, there was also no significant difference between old stimuli ($M = 67.30$s, $SD = 34.50$s), and novel stimuli from the same category ($M = 115.99$s, $SD = 71.37$s), $F_{(1, 6)} = 2.54$, p = ns, but significant differences when compared to novel stimuli from a novel category ($M = 147.19$s, $SD = 83.01$s), $F_{(1, 7)} = 8.62$, $p < .05$. There was no significant difference between novel stimuli from the same category and from a novel category, $F_{(1, 6)} = 0.21$, p = ns.

Analysis was made of all first exposures to the specific tangrams, i.e. data from the first trial of training and from the two sets of novel stimuli in test. Although there will be some effect of experience with the task, the results show that this may not have a large effect as will be seen. A 3 x 2 ANOVA comparing results from training and novel stimuli from the same category found there to be no significant differences between the training conditions, $F_{(2, 27)} = 0.34$, p = ns, but a significant difference between the trials, $F_{(1, 27)} = 8.71$, $p < .01$. Interaction was non-significant, $F_{(2, 27)} = 2.71$, p = ns. In the comparison between training and novel stimuli from a different category there was no significant differences between training conditions, $F_{(2, 26)} = 0.53$, p = ns, or between trials, $F_{(1, 26)} = 0.82$, p = ns, and interaction was also non-significant, $F_{(2, 26)} = 1.63$, p = ns.

For those trained under time pressure, there was no significant differences between training completion times and those for the novel stimuli of the same category, $F_{(1, 11)} = 0.02$, p = ns, as was also the case for those trained without time pressure, $F_{(1,9)} = 3.93$, p = ns. There were significant differences between the trials for those trained in the alternate task, $F_{(1, 7)} = 12.77$, $p < .01$.

A d’ analysis was made on the results from the recognition test. The recognition test was completed with simple <u>yes</u> and <u>no</u> responses and from this the ratios of hits and false alarms were calculated. A one-way ANOVA was performed on both the results for tangram recognition and piece recognition. There was no significant difference between recognition of tangrams by those in the time pressure trained group ($M = 5.08$, $SD = 1.98$), those without time pressure ($M = 4.83$, $SD = 1.64$) and those in the alternate task ($M = 5.88$, $SD = 0.85$), $F_{(2, 31)} = 1.52$, p = ns. There was also no significant difference for piece recognition between the time pressure group ($M = 4.93$, SD

= 1.57), no time pressure group (M = 5.25, SD = 1.34) and the alternate task group (M = 5.17, SD = 1.53), $F_{(2,31)} = .187$, p = ns.

Discussion:

In this experiment, participants were trained in one of three conditions to see if there was any effect of training when tested on a decision making task under time pressure. One of the main aims of this study was to find the optimum method of training. The training conditions were (a) under time pressure in the tangram task, (b) without time pressure in the tangram task, or (c) an alternate task under time pressure followed by the tangram task without time pressure. When the time taken to complete the tangrams in test was analysed it was found there was no significant difference between the different training conditions. This suggests that there is no need to train people in a manner specific to the task such that they can gain experience with completing the task under time pressure, or that any pressure put on them by time pressure is more detrimental than not having pressure put on them. However, when the analysis was broken down into each of the groups, some differences were found. The data from the group trained under time pressure had significant differences between their completion times between shapes previously seen in training and novel shapes, regardless of the category they came from. This was not found in the other groups. This suggests that during training, when put under time pressure, the participants were only able to learn the specific exemplars that they were exposed to and did not learn any rules that they could generalise to novel stimuli. This is also shown by the fact that although their speed of completion increased with exposure to the same stimuli, upon exposure to novel stimuli, the speed returned to that of the tangrams seen in the first phase of testing. This is to say that there was no effect of exposure to the task in general when given novel stimuli, and it was as if they were doing tangrams for the first time again. This suggests that when trained under time pressure, the load on working memory is too great for the tangrams to be analysed at any level greater than their surface characteristics. Therefore, training under time pressure may be an optimum approach if in is the skill to be learnt only requires familiarity with a fixed set of stimuli or procedures. However, it would not allow flexibility in application of skills learnt and would hinder performance in a task that involved an array demands.

The specificity of learning is concurrent with such factors as the Yerkes-Dodson law, where time pressure may increase arousal but decrease the amount of attention available to the task. With the addition of the requirement to attend to the amount of time passing, there are more demands on working memory than under conditions where those demands are absent. This is reflected by the results of those who were trained without time pressure. During the first exposure to the tangrams in training, they took significantly more time to complete the tangrams than in the second exposure, but then around the same amount of time in test as in the second exposure. Therefore, in the first phase they were able to use the ample time available to learn about the task and had no distracters to avert their attention. In comparison, the group trained under time pressure continued to improve their time with each exposure, suggesting they required the further exposures to allow them to extract more information. The above postulations also answer the question of how learning will differ given the training condition. Time pressure leads to specificity of learning due to the constraints put on them where as when trained without time pressure more general methods are extracted. This is also reflected by the lack of differences between the speeds of completion in the various testing conditions by the group trained without time pressure.

That the results between the groups are in general non-significant suggests that in training for a decision making paradigm somatic markers are not greatly relevant. However, with the group trained in the alternate task showing a pattern of being the fastest in completing the task, suggests that perhaps that by the positive feelings associated with training for the task to be tested being compared with the stress experienced in training under time pressure in the previous trials, they will be experienced more positively than if they had no stressful situation to compare it to. This is to say that owing to the stressful experienced of being subjected to time pressure in the first part of training, any positive feelings about completing the tangrams after that phase will be experienced as more positive. Therefore, due to the increasingly negative feelings associated with training from training in the alternate task to training under time pressure in the task to be tested in, performance will decline in a manner reflecting these affective markers, as is reflected in the pattern found.

The results from those trained in the alternate task followed by training with the tangrams without time pressure give the most interesting results and tie in a lot of the previously mentioned theories. With each exposure to the same stimuli, the speed of completion increased significantly. This reflects a simple learning curve, where they are learning more about those specific stimuli. However, when given stimuli in test of the same underlying structure as those previously seen, performance stays at the same level as that for the stimuli seen three times previously and, unlike the other groups, does not return to a level similar to that in the first exposure to the tangrams. In addition, the speed of completion of the novel tangrams from a novel category are similar to that seen in the first exposure in training, and although not significantly different to the time taken for novel tangrams from the same category as in training, is different to the old tangrams used in testing. Therefore, the times may be somewhere on a continuum, where degree of similarity of novel stimuli to those the participants were exposed to in training will effect the subsequent time taken to complete the tangrams. That is, the amount of time to complete the tangram will decrease the more similar it is to that in which participants were trained, such that tangrams with the same underlying structure will be completed faster than those with a novel underlying structure. This shows that this group is learning the underlying rules that increase efficiency in completing the tangrams, e.g. the underlying structure formed by the two large triangles. This is also seen to some degree with those trained just on the tangrams without time pressure. This is not seen in those trained under time pressure. But the actual significant results found in the group trained initially in the alternate task unearth interesting implications to training. By exposing these participants to time pressure in a task unrelated to the tangrams, they are able to learn how to train or reorganise their working memory to deal with juggling both the task at hand and attending to the passage of time. Then by exposing them to the task they will be tested in, i.e. tangrams, without time pressure, they have the advantage of having residual arousal, thus focusing their attention in the task, but without the need to attend to anything other than the task being currently completed. Thus, upon testing, they have the advantage of both groups. They have had the time to learn the task at perhaps more depth than those trained under time pressure, but also know how to cope with time pressure. Thus by combining these two skills, they are more able than the other groups.

It was hoped that the d-prime analysis would uncover the differences in what the participants were learning. However, the analysis showed there to be no significant differences in the different groups being able to identify the tangrams or the pieces they saw in training. In retrospect, this test was greatly insufficient. What would have been more useful would have been to test them on not only the tangrams and their pieces but also on various combinations that the pieces can make up. This would allow an analysis of any rules the participants had learnt or what more precisely they were focussing on given the different conditions. With the test given, pieces were unlikely to be misidentified because the whole set of seven pieces were seen in every exposure. Therefore, with this high degree of exposure, chances of mistakes being made were low regardless of the condition in which they were experienced.

Regardless of this, it does appear there are differences in what participants attend to in training, as can be seen from the above explanations. Therefore, in discussing the degree of specificity of learning that takes place given different training conditions the following seems to be the case: when trained under time pressure, there is a great degree of specificity of training, such that only knowledge of the stimuli they are exposed to are learnt. In contrast, when given time to learn the task, general rules or aspects of the task are learnt that can be applied when novel stimuli are experienced. To answer the question of which method of training is optimal cannot be deduced from this experiment due to the non-significant results, but given the seemingly robust pattern found in the test results, it appears that training under time pressure is the least advantageous method. In contrast, the more lengthy method of training first in an alternate task requiring similar skill, here decision making, and then training in the task to be tested on without time pressure may be the most favourable method.

Therefore, it appears that training may well be a useful thing to do to improve performance when tested on making a decision under time pressure. However, the most advantageous method of training is most likely to be have exposure to time pressure in a task other than that which will be tested. This being followed by ample time to learn the task that will be tested allows freedom to learn the task without any constraints on working memory or any negative markers becoming associated with effective methods of completing the task.

Although the results in this experiment are non-significant, there are a number of criticisms of the method here that may allow a more extensive experiment to better expose these factors. One such criticism is that the number of participants was quite small. This is one of several difficulties with third year projects in general. Firstly, other students in the department that, although required to participate in a specific number of experiments, often fail to attend the experiments, and it is difficult to get the require number of people to sign up in the first place. Secondly, this project was one of a group using tangrams as the basis of the research. Therefore, as participants were required to be naïve to the stimuli, anyone to have taken part in one of the other experiments was ineligible to participate in this experiment. This again limited the available pool of participants available to take part. Hence, the minimum number of people required to fulfil counterbalancing of stimuli were recruited. Ideally, each stimuli combination would have been tested twice to give results that are more representative. Therefore, although some *F*-values reflect that an increase in participants would be fruitless, there are enough values to suggest repetition of the experiment with more participants may give more significant results.

Another constraint of such a project as this is that there is only a limited amount of time available both to the experimenter and the participants. Therefore, both training and testing are not as lengthy as desired. In this case, training used only eight stimuli exposures, which is much less than what would be ideal. Testing is even briefer, which doesn't allow much to be revealed. Therefore, it would useful for the experiment to be repeated with an increased number of trials, both in training and test.

A major problem that will be associated with the kind of time pressure imposed on these participants is their motivation to comply. Given that they are simply asked to complete the tangrams as quickly as possible, it may not be enough to actually cause any change in their behaviour. Indeed, there was no significant difference between the groups in the amount of time the spent completing the tangrams in training. Therefore, the pure subjective experience of time pressure here is hard to measure and indeed impose upon participants. Furthermore, it is not known whether a feeling of time pressure was actually created in this experiment or not. A more appropriate method therefore would be to calculate what sufficient time to complete the tangrams is and then give a count down of time, which would then impose time constraints. This would impose a greater sense of pressure, as the limit of time available would be more perceptible than in this case. Had more piloting been plausible given the restrictions of this project, this could have been measured previous to designing the experiment. Sadly, as a number of pilots were undertaken for various other factors of greater importance, such as balancing for meaning of the tangrams, any more pilots would have been implausible due to the increasing limitations on the participant pool. Data from experiments such as this one can be analysed and applied as such in any future experiments using tangrams. Conversely, to increase motivation to comply, such measures as monetary reward could be used, as is seen in other experiments, or make the completion somewhat competitive, such that whomever finished fastest out of the participant pool would receive a reward (e.g. Smith, & Rogers, 1994). This would increase the plausibility of participants completing the task as quickly as possible while allowing individual variance in ability to cause problems should the task not be finished in the time available, thus removing a data point.

In fact, the tangrams themselves are a source of criticism themselves. As mentioned previously, tangrams and puzzles like them have rarely been used in experiments. As there are varying degrees to which people are capable of performing spatial puzzles and also degrees of confidence (many participants here confessed that they often had problems with these kinds of tasks), there may be too much personal variance to validate tangrams as a method of testing. It may indeed be more appropriate to sample questions from standardised tests such as the GRE, which was used here as the alternate task, as the paradigm to study decision-making.

Additionally, the program in which the trials were presented had a number of glitches. As a new program, there was not much time to test it and iron out any problems associated with it. Therefore, with some stimuli participants had some difficulty fitting the pieces for the program to recognise that they had been placed in the correct position. Therefore, if the tangrams were to be used in future research, such difficulties would have to be rectified.

In summary, training does seem to be a worthwhile measure to increase performance in a test involving decision making under time pressure. However, the most useful method is to separate the two factors, i.e. the task and the time pressure. By training participants in these factors separately, underlying rules that allow improved performance in task completion can be found, and at the same time, conditions are provided for skills associated with coping with time pressure to be learnt. Sadly, the differences between the training groups were largely non-significant but repetition of this experiment with more participants and more time to train and test may well result in some very significant and interesting results about training for decision making under time pressure.

Reference:

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory Vol. 2* (pp. 89-195). New York: Academic Press.

Baddeley, A. (1999). *Human memory: Theory and practice* (Rev. ed.). Hove, England: Psychology Press Ltd.

Bartel, A. (2000). Measuring the employer's return on investments in training: Evidence from the literature. Industrial Relations, 39, 502-524.

Bechara, A., Damasio, A.R., & Damasio, H. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, 10 (3), 295-307.

Bechara, A., Damasio, A.R., Damasio, H., & Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *15*, 7-15.

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S.W., & Nathan, P.E. (2001). Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia*, 39 (4), 376-389.

Bechara, A., Tranel, D., Damasio, H, & Damasio, A.R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex,* 6, 215-225.

Benson 3rd, L. & Beach, L.R. (1996). The effects of time constraints on the prechoice screening of decision options. *Organizational Behaviour and Human Decision Processes*, 67 (2), 222-228.

Benson 3rd, L., Groth, M., & Beach, L. (1998). The relationship between time constraint and time pressure. Americas Conference on Information Systems.

Ben Zur, H., & Brenznitz, J. (1981). The effect of time pressure on risky choice behaviour. *Acta Psychologica*, 47, 89-104.

Bowden, E.M. (1985). Accessing relevant information during problem solving: Time constraints on search in the problem space. *Memory & Cognition*, 13 (3), 280-286.

Bower, G.H., & Springston, F. (1970). Pauses as recoding points in letter series. *Journal of Experimental Psychology*, 83 (3), 421-430.

Butler, B.E. (1994). Spatial puzzles: A guide for researchers. *Canadian Psychology*, 35, 47-65.

Carlson, R.A., Khoo, B.H., Yaure, R.G., & Schneider, W. (1990). Acquisition of a problem-solving skill: Levels of organisation and use of working memory. *Journal of Experimental Psychology: General*, 119 (2), 193-214.

Carlson, R.A., Sullivan, M.A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 517-526.

Cohen, J.D., Perlstein, W.M., Braver, T.S., Nystrom, L.E., Noll, D.C., Jonides, J., & Smith, E.E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature,* 386, 604-608.

Damasio, A.R. (1994). *Descartes' error: emotion, reason, and the human brain*. New York: Grosset/ Putnam.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behaviour. Psychological Review, 66, 183-201.

Educational Testing Services (2001). *GRE: Practicing to take the general test* (9th ed.). Princeton, NJ: Graduate Record Examinations Board.

Eysenck, M.W. (Ed.) (1990). *The Blackwell dictionary of cognitive psychology*. Oxford: Basil Blackwell Ltd.

Franklin, A.L., & Hunt, E. (1993). An emergency simulator for examining time-pressured decision making.

Gobet, F., Lane, P.C.R., Croker, S., Cheng, P. C-H., Jones, G., Oliver, I., & Pine, J.M. (2001). Chunking mechanisms in human leaning. *Trends in Cognitive Sciences*, 5 (6), 236-243.

Hardy, M., & Hayes, S. (1996). *Beginning psychology* (4th ed.). Oxford: Oxford University Press.

Henderlong, J., & Paris, S.G. (1996). Children's motivation to explore partially completed exhibits in hands-on museums. *Contemporary Educational Psychology*, 21, 111-128.

Johnston, J.H., Driskell, J.E., & Salas, E. (1997). Vigilant and hypervigilant decision making. Journal of Applied Psychology, 82 (4), 614-622,

Joslyn, S., & Hunt, E. (1998). Evaluating individual differences in response to time-pressure situations. *Journal of Experimental Psychology: Applied*, 4 (1), 16-43.

Kellogg, J.S., Hopko, D.R., Ashcraft, M.H. (1999). The effects of time pressure on arithmetic performance. *Journal of Anxiety Disorders*, 13 (6), 591-600.

Kerstholt, J.E. (1994). The effect of time pressure on decision-making behaviour in a dynamic task environment. *Acta Psychologica*, 86, 89-104.

Lerch, F.J., & Harter, D.E. (2001). Cognitive support for real-time dynamic decision making. *Information Systems Research*, 12 (1), 63-82.

Lerner, J.S., & Keltner. Fear, anger, and risk. *Journal of Personality and Social Psychology*, 81 (1), 146-159.

Loewenstein, G.F, Weber, E.U., Hsee, C.K., Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127 (2), 267-286.

Maule, A.J., Hockey, R.J., & Bdzola, L. (2000). Effects of time-pressure on decision –making under uncertainty: Changes in affective state and information processing strategy. *Acta Psychologica*, 104 (4), 283-301.

Miller, J.G., Information input and overload and psychopathology. *American Journal of Psychiatry*, 116, 695-704.

Mittal, V., & Ross Jr., W.T. (1998). The impact of positive and negative affect and issue framing on issue interpretation and risk taking. *Organisational Behaviour and Human Decision Processes*, 76 (3), 298-324.

Navon, D. (1984). Resources- A theoretical soup stone? *Psychological Review*, 91, 216-234.

Neck, C.P., & Moorhead, G. (1995). Groupthink remodelled: The importance of leadership, time pressure, and methodical decision-making procedures. *Human Relations*, 48 (5), 537-557.

Ordonez, L., & Benson 3rd, L. (1997). Decision under time pressure: How time constraint affects risky decision making. *Organizational Behaviour and Human Decision Processes*, 71 (2), 121-140.

Paulus, M.P., Hozack, N., Zauscher, B., McDowell, J.E., Frank, L., Brown, G.G., & Braff, D.L. (2001). Prefrontal, parietal, and temporal cortex networks underlie decision-making in the presence of uncertainty. *NeuroImage*, 13, 91-100.

Smith, K.H., & Rogers, M. (1994). Effectiveness of subliminal messages in television commercials: Two experiments. *Journal of Applied Psychology*, 79 (6), 866-874.

Romano, D.M., & Brna, P. (2001). Presence and reflection in training: Support for learning to improve quality decision making skills under time limitations. *CyberPsychology and Behaviour*, 4 (2), 265-277.

Verplanken, B. (1993). Need for cognition and external information search: Responses to time pressure during decision-making. *Journal of Research in Personality*, 27, 238-252.

Zajonc, R.B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.

Contract F61775-01-WE039
Title: ENHANCING DECISION PERFORMANCE

This material is based upon work supported by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under Contract No. **F61775-01-WE039**.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory.

"The Contractor, UNIVERSITY OF SOUTHAMPTON, hereby delcares that, to the best of its knowledge and belief, the technical data delivered herewith under Contract No. **F61775-01-WE039** is complete, accurate, and complies with all requirements of the contract.

DATE: 18 Jan. 2003

Name and Title of Authorized Official:

Dr Itiel Dror

Senior Lecturer

I certify that there were no subject inventions to declare as defined in FAR 52.227-13, during the performance of this contract.

DATE: 18 Jan. 2003

Name and Title of Authorized Official:

Dr. Itiel Dror

Senior Lecturer