Modelling the Relationship between Visual Short-Term Memory Capacity and Recall Ability

Richard Ll. Smith
School of Social Sciences
Brunel University
Uxbridge UB8 3PH, UK
rls@hwyl.org

Peter C. R. Lane
School of Computer Science
University of Hertfordshire
Hatfield AL10 9AB, UK
peter.lane@bcs.org.uk

Fernand Gobet
School of Social Sciences
Brunel University
Uxbridge UB8 3PH, UK
fernand.gobet@brunel.ac.uk

Abstract

Previous cognitive modelling work has suggested that the decline of short-term memory (STM) capacity is the dominant factor of age-related decline on recall ability. We report the results of a set of experiments investigating in further detail the effect of varying short-term memory capacity on the recall of chess positions using the CHREST cognitive architecture, and demonstrate a good qualitative match for human data. We then use these results to suggest a potential functional reason for the size of STM capacity in humans and go on to explore the relationships between STM capacity, ageing, knowledge, presentation time, and recall ability.

1 Introduction

Psychology and cognitive science study the workings of the human mind. Applications include many areas of human behaviour, both in trying to understand why people do things in certain ways, as well as trying to predict how people will react in new situations. The process of simulation and modelling is a challenge in these sciences, as the phenomena under study (people) are complex, and simple models typically fail to capture enough properties of the situation to provide adequate results. Since Newell [20], cognitive scientists have used *cognitive architectures* to collect the results from many different concrete simulations into single explanatory frameworks. The advantage of having a single architecture is that constructing a specific model in one area can rely on the results of models from other areas to constrain their parameters and focus attention on what is scientifically important.

In this paper, we are interested in what happens to human memory as a result of ageing, a topic of increasing importance in a society where people are living and working for longer. As a focus of our study, we use a popular memory task from psychology, the recall of chess positions; a participant in the experiment is presented with a chess position for a few seconds, and then must reconstruct the position from memory. The most complete model of chess expertise has been constructed in the CHREST (Chunk Hierarchy and REtrieval STructures) [10, 12] cognitive architecture, and we use this model to simulate the effects of ageing. In previous work [24], we repeated an experiment performed by Charness [3, 4, 5] to explore the interactions between chess skill, age, and presentation time on recall ability. Charness found a positive correlation between recall ability and chess skill. He also found a negative correlation with ageing, and that this difference between young and old players became greater as the presentation time increased.

One of the main advantages of using a cognitive model is that it becomes possible to manipulate the workings of the mind in a way not possible with human subjects. Consequently, when we repeated the experiment, we explored the combined qualitative effect of varying each of a set of six key parameters of CHREST. We found that only one of the key parameters, namely STM capacity, caused the necessary interaction between age and presentation time when manipulated.

Computational limitations meant that we were only able to explore high and low values for each of the parameters in [24], but with the identification of STM capacity as the key parameter, we here report our further investigations into the effect of changing this parameter specifically.

2 Background

The human mind can be considered to have three different types of memory, differentiated by the amount of time information can remain in that memory, as well as the source of that information. *Sensory memory* serves only to store sensory information from the senses, which is rapidly

lost unless selected by the attention system. *Short-term memory* (STM) is the mind's working memory; the information that one is directly conscious of. *Long-term memory* (LTM) is the mind's persistent information store, and the information thus stored mostly endures without being consciously accessed. It is STM with which we are primarily concerned in this paper.

It is still a matter of some debate as to how many forms of STM exist, and to the extent of their domains and independence. One influential model is that proposed by Baddeley and Hitch [2], in which STM is divided into the visuospatial sketchpad, the phonological loop, the central executive, and the episodic buffer [1]. It can be shown (through dual-task experiments) that there is at least a strong degree of independence between the component of STM that handles visuo-spatial information and the component that handles auditory information. In this paper we are primarily concerned with the part of STM that handles visuo-spatial operations (hereafter referred to simply as STM).

It is known that the capacity of STM is quite small. Despite some early evidence that human STM was 7 ± 2 [19], it is now understood that the limited STM capacity can be masked by the chunking effect [7, 12]. Coltheart [8] found that subjects could only retain an average of 4.3 items in STM, and later studies have broadly agreed with this figure; it is now generally agreed that young adults have a visual STM of around 4 chunks [9, 18].

It came as some surprise to find that the capacity of STM was so small, given the complex tasks encompassing many different pieces of information that humans are capable of performing. However, as shown by Simon [13, 22, 23], the human mind is able to act on pieces of information in combination, known as chunks. Chunks are associated pieces of information that have become familiar, and so collectively possess some meaning; a good example is how individual letters become associated as words, which are then stored in memory. Current research indicates that items in STM are actually references to the information stored in LTM [12].

It is not presently understood from either an evolutionary, or a neurological basis, why the capacity of STM should be around 4 chunks. What is clear, though, is that STM capacity does not remain constant. Abilities such as vision, hearing and memory [21] all decline with increasing age. An important question in the study of ageing is the extent to which expertise might moderate the effects of ageing [6]; does increasing knowledge mitigate the loss of STM capacity? To properly explore this question, we need a deeper understanding of the relationship between STM capacity and memory ability. We next introduce CHREST, the cognitive theory within which we shall explore this relationship.
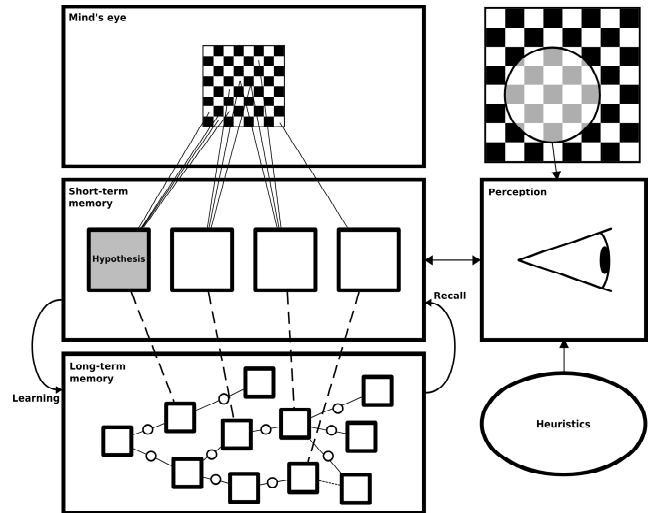

Figure 1. The main components of CHREST

3 CHREST

CHREST is a computer-based, general-purpose cognitive architecture that incorporates STM, LTM, and perception/attention systems [12]; see Figure 1. It has previously been shown to successfully model domains as diverse as chess expertise [10], physics representation [16], and language acquisition [11, 15]; in this paper, we are looking specifically at its ability to model the perception and memory of chess positions.

An important goal of CHREST is that it operates under the same constraints as humans do as far as possible. Specifically, humans take time to perform tasks, and an important aspect of the experiments in this investigation is keeping track of this. Consequently, for each operation that is performed by CHREST that is known to take time by humans, such as fixating a new square, time equivalent to the measured (or best estimated) human value is added to CHREST's internal clock.

The basic unit of the CHREST memory representation is the chunk. As described above, a chunk is a familiar pattern of features; in the case of chess, chunks consist of a subset of the piece-square combinations in a game position. These chunks are stored in a LTM structure represented as a discrimination network. CHREST learns chunks through repeated exposure to game positions. New chunks that are encountered are added to the discrimination network in a process called *discrimination*, and more information is stored about chunks that are re-encountered in a process called *familiarisation*; thus, discrimination increases the number of chunks potentially recognisable, and familiarisation in-

creases the amount of detail that can be recalled about each chunk.

The STM of the model is a set of references to chunks held in LTM. The STM structure can contain up to a fixed maximum number of chunks. This maximum is normally equal to the human STM capacity of around 4, but in this investigation we look into the effects of varying this figure.

The learning process is directed by the perception system (see [17] for a recent overview). Simulated eye movements (approximately corresponding to attention) are carried out between squares on the board. The primary method of determining the next square to fixate utilises the chunks currently held in STM. The perception system uses links previously established between chunks in LTM to attempt to determine if the chunk is (or is not) part of a previously learnt, more specific chunk. When this strategy is not possible (for example, when first looking at the board, or with a novice to the domain with few chunks in LTM), one of a number of domain-independent heuristics is used, such as "look at the centre of the board", or "select a random occupied square". As can be seen above, the perception system is both responsible for, and affected by, information in memory, resulting in complex emergent behaviour.

More information on CHREST is available from:

`http://people.brunel.ac.uk/˜hssrrls/chrest/index.php`

4   Experiment

The CHREST system was set up as a model of the expertise of chess players. A set of discrimination networks were trained to represent the acquired knowledge of chess players of a range of abilities. The data used to train the networks were a set of 76,420 chess positions (mostly taken from Master games after around 20 moves, but with some earlier game states also included).

Charness's experiment on position memorisation [5] was repeated using CHREST to simulate the subjects. In the original experiment, a group of human subjects were each presented with a chess position for a short time, and were then required to attempt to recall the pieces present and their locations. With CHREST, the system was exposed to a test position for a length of time as defined by the system's internal clock. During this simulated period of time, CHREST modelled an attempt by a human subject to memorise the position: eye movements were simulated, as directed by the contents of STM and heuristics, data were perceived, and STM was updated with chunks recalled from LTM. After the allotted time was up, CHREST made an attempt to reconstruct the position using the contents of its memory. The results of comparing this reconstruction to the original position were then collected for evaluation. The dependent variable was the proportion of pieces that CHREST correctly replaced.

The independent variables were as follows:

- CHREST networks of {100, 1,000, 10,000, 100,000, 200,000, 300,000} nodes were used. 300,000 nodes represents the estimated number of nodes acquired by a chess Master [14].

- Presentation times of {1, 2, 4} seconds. These were the values used by Charness in the human-based experiments; the same values were used in the experiment for ease of comparison.

- STM Capacity was modified in the range [1, 10]. As discussed above, it has been shown that young adult humans have a visual STM capacity of around 4. Previous modelling with CHREST suggests that STM capacity decreases with age, whereas older research suggested STM capacity was larger. We have chosen a set of values to investigate this full range.

For each combination of parameters as described above, this process was repeated ten times to simulate the ten subjects in the Charness experiment: this meant 1,800 experiments in total. Each simulated subject was tested on 50 separate positions independent of the training data.

5   Results

Figure 2 shows graphs of the interactions between STM capacity and presentation time by network size.

5.1   General Trends

In general, the results of the experiment show the expected correlation with the independent variables: recall tends to improve as the maximum capacity of STM is increased, when the time allocated to the system for inspecting the board is lengthened, and with larger discrimination networks (simulating increased knowledge). The explanations are straightforward: a larger STM capacity allows for more chunks (and hence more information about the position) to be stored; a longer presentation time allows the system to perceive more of the board and spend more time searching its memory for chunks; and a more comprehensive discrimination network makes it more likely that a chunk relevant to the position will be available.

However, diminishing returns are seen as the values of the independent variables approach their maxima. As the STM capacity of the system increases, the selected chunks cover more of the board, until it becomes increasingly difficult and/or less likely to find a separate chunk in LTM that incorporates the uncaptured information. Similarly, as additional time for inspecting the board is made available to the system, it becomes increasingly unlikely that further novel
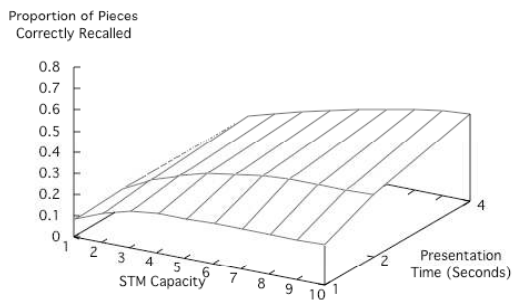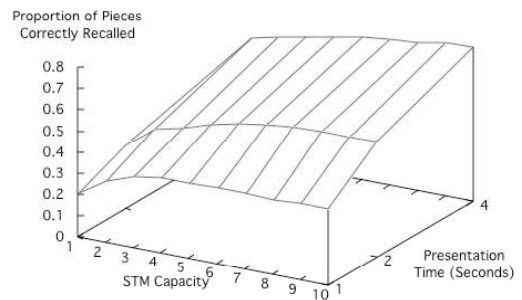
(a) Network size: 100
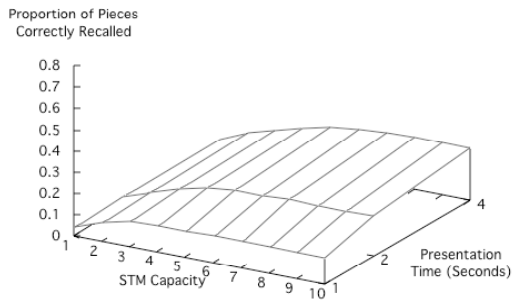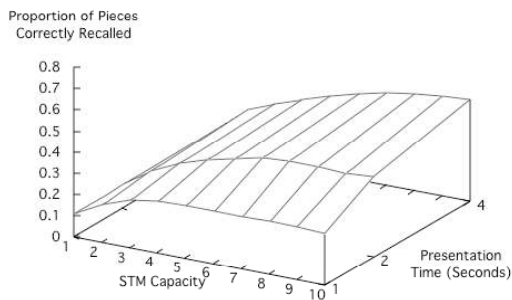
(d) Network size: 100,000
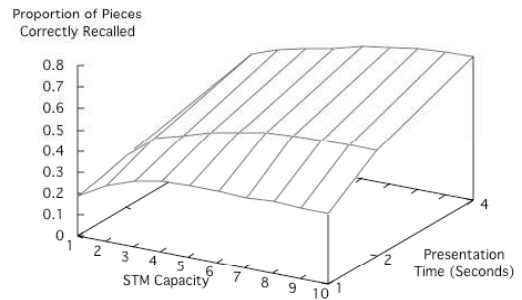
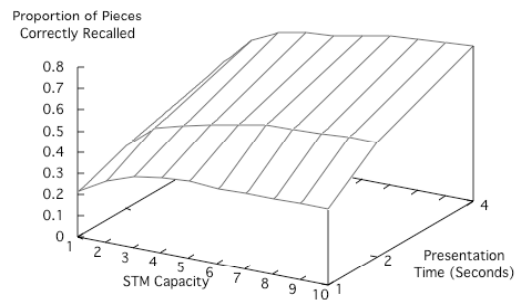
(b) Network size: 1,000

(e) Network size: 200,000

(c) Network size: 10,000

(f) Network size: 300,000

Figure 2. Interactions between STM capacity and presentation time, by network size.

information is perceived from the board, and less likely that a chunk more specific to the position exists in LTM. As more chunks are added to the LTM store, the new chunks that are added are less commonly-occurring and more specific, so each additional chunk is less likely to be relevant to any specific position.

There are some specific interactions between the variables. Increasing the presentation time only improves the recall accuracy if the system has sufficient resources to take advantage of the extra time; with a network size of only 100 nodes, there is little improvement when the STM capacity and/or the presentation time is increased.

5.2 Comparison to Experimental Data

Figure 3 shows the results of the human subjects in the Charness experiments compared to a subset of the CHREST results limited to the same approximate parameters of the human subjects. Chess players of the standard used in the experiments (class B/C) would be expected to have acquired between 100 and 1,000 chunks [25]. The STM capacity of a young adult is around 4 chunks as has been already discussed; results of the simulations with a lower STM capacity (arbitrarily selected as 3: the human data are not clear) as would be expected due to ageing are also shown.

As can be seen from Figure 3, CHREST performs less well than the human subjects in an absolute sense but the data do match some qualities of the human data. Like humans, the model displays improved recall accuracy with longer presentation time, and a smaller increase per second between 2 and 4 seconds than between 1 and 2 seconds (interestingly, this second feature is only observed when STM capacity is set between the approximate human norms of 3 and 4). When STM capacity is reduced (believed to be the dominant factor in age-related degradation in the domain, as already discussed), absolute accuracy decreases, but the shape of the plot remains the same again, consistent with the human results.

6 Discussion

6.1 Accuracy

CHREST has qualitatively reproduced the effects of skill and STM capacity, but specific models perform less well than humans of the same estimated skill level. Since CHREST has previously been shown to accurately reproduce human ability to recall chess positions [10], we believe that the difference in accuracy is due to the different positions CHREST was tested with (it has been previously shown [14] that the accuracy of position recall is strongly dependent on the type of position).
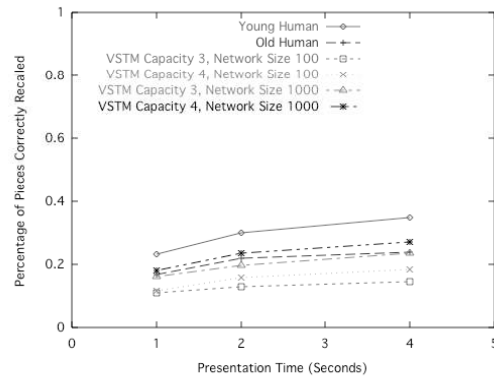

Figure 3. Plot of performance against presentation time

6.2 STM Capacity

One feature of the results is the diminishing returns obtained with a STM capacity of more than 4 chunks. In all experiments, most of the maximum performance had been obtained when STM capacity was set to 4. This is most obvious with an exposure time of 1 second. Accuracy as a proportion of the maximum accuracy achieved increases from 0.914 to 0.985 between a STM capacity of 3 and 4.

Thus, we suggest that the capacity of human STM is limited to around 4 chunks because a larger capacity is not particularly useful. Natural Selection predicts that traits of limited value to the organism's survival will not be preserved due to genetic drift; also it is likely that extra STM capacity requires some extra expenditure of energy to maintain.

One significant objection that may be raised against this hypothesis is the limitations of chess as a domain for such a general conclusion. Chess is an artificial and limited problem space, and though a STM of 4 may work well in perceiving the game, it is not clear if these results will generalise. (It may even be the case that chess has become such a popular game because the perception of the board is so well matched to human STM). This is certainly a good reason to be cautious of this interpretation, yet we believe that this is an intriguing result nevertheless and warrants further study.

6.3 Methodology

As described above, a major advantage of using a cognitive model is that it makes it possible for the cognitive scientist to explore the effects of setting parameters to specific values. In this experiment we have varied two parameters of the model: STM capacity, and skill level, and assigned each of them a fine-grained series of values to investigate

their effects and interactions. This systematic approach to exploring the parameters of the model is one we believe to be valuable, and which we hope to explore in further experiments. We are developing tools to automate this process and apply these experiments to larger sets of variables.

This experiment is but one part in a project to not only understand CHREST more fully, but also to obtain a deeper understanding of the parameters of human minds. Understanding the limits of human memory is likely to have implications beyond the cognitive and social sciences. Humans still clearly outperform computers in many AI applications, in spite of a relatively limited ability to process problem-solving states. A cognitive model of the domain knowledge acquired by humans will likely be of immense importance in improving the kinds of problem-solving heuristics which can be applied to computationally intensive tasks. As the wider CHREST project continues, we plan to give CHREST an ability to play chess well whilst retaining the 'limitations' of human memory, and previous CHREST work suggests that these results will transfer to many other domains.

7 Conclusion

The CHREST cognitive architecture supports the computational simulation of many aspects of human memory. We have shown that CHREST qualitatively simulates human performance in recalling chess positions depending on their skill and STM capacity; the effect of STM capacity is particularly interesting, as this is believed to be key to understanding the effect of age on human performance. As part of a growing body of evidence, this result further supports the theory of human memory embodied in CHREST, and further work is aiming to utilise this memory in complex problem-solving tasks.

Acknowledgements

References

[1] A. D. Baddeley. The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4:417–423, 2000.

[2] A. D. Baddeley and G. Hitch. Working memory. In G. H. Bower, editor, *The psychology of learning and motivation: Advances in research and theory*, volume 8, pages 47–89. New York: Academic Press, 1974.

[3] N. Charness. Aging and skilled problem solving. *Journal of Experimental Psychology: General*, 110:21–38, 1981.

[4] N. Charness. Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, 7:467–476, 1981.

[5] N. Charness. Visual short-term memory and aging in chess players. *Journal of Gerontology*, 117:615–619, 1981.

[6] N. Charness and J. I. D. Campbell. Acquiring skill at mental calculation in adulthood: A task decomposition. *Journal of Experimental Psychology: General*, 117:115–129, 1988.

[7] W. G. Chase and H. A. Simon. Perception in chess. *Cognitive Psychology*, 4:55–81, 1973.

[8] M. Coltheart. *New Horizons in Psychology*, volume 2, chapter Visual information-processing. Penguin Books, 1972.

[9] N. Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24:87–114, 2001.

[10] A. D. de Groot and F. Gobet. *Perception and Memory in Chess: Heuristics of the Professional Eye*. Assen: Van Gorcum, 1996.

[11] D. Freudenthal, J. M. Pine, J. Aguado-Orea, and F. Gobet. Modelling the developmental patterning of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science*, 31:311–341, 2007.

[12] F. Gobet, P. C. R. Lane, S. J. Croker, P. C.-H. Cheng, G. Jones, I. Oliver, and J. M. Pine. Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5:236–243, 2001.

[13] F. Gobet and H. A. Simon. Expert chess memory: Revisiting the chunking hypothesis. *Memory*, 6:225–255, 1998.

[14] F. Gobet and H. A. Simon. Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, 24:651–682, 2000.

[15] G. A. Jones, F. Gobet, and J. M. Pine. Modelling vocabulary acquisition: An explanation of the link between the phonological loop and long-term memory. *Journal of Artificial Intelligence and Simulation of Behaviour*, 1:509–522, 2005.

[16] P. C. R. Lane, P. C.-H. Cheng, and F. Gobet. CHREST+: Investigating how humans learn to solve problems with diagrams. *AISB Quarterly*, 103:24–30, 2000.

[17] P. C. R. Lane, F. Gobet, and R. Ll. Smith. Attention mechanisms in the CHREST cognitive architecture. In *Proceedings of the Fifth International Workshop on Attention in Cognitive Science*, 2008.

[18] S. J. Luck and E. K. Vogel. The capacity of visual working memory for features and conjunctions. *Nature*, 390:279–281, 1997.

[19] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 1956.

[20] A. Newell. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press, 1990.

[21] R. Schulz and T. A. Salthouse. *Adult development and ageing*. Upper Saddle River, NJ: Prentice Hall, 1999.

[22] H. A. Simon. How big is a chunk? *Science*, 183:482–88, 1974.

[23] H. A. Simon and K. J. Gilmartin. A simulation of memory for chess positions. *Cognitive Psychology*, 5:29–46, 1973.

[24] R. Ll. Smith, F. Gobet, and P. C. R. Lane. An investigation into the effect of ageing on expert memory with CHREST. In *Proceedings of The Seventh UK Workshop on Computational Intelligence*, 2007.

[25] A. J. Waters and F. Gobet. Mental imagery and chunks: Empirical and computational findings. *Memory & Cognition*, 2008. (In press).