Computational Modelling of Mental Imagery in Chess: A Sensitivity Analysis

Fernand.Gobet (fernand.gobet@brunel.ac.uk)
Centre for the Study of Expertise, Brunel University
Cleveland Road, Uxbridge UB8 3PH UK

Andrew J. Waters (andrew.waters@usuhs.mil)
Department of Medical and Clinical Psychology
Uniformed Services University of the Health Sciences
Bethesda, MD 20814 USA

Abstract

An important aim of cognitive science is to build computational models that account for a large number of phenomena but have few free parameters, and to obtain more veridical values for the models' parameters by successive approximations. A good example of this approach is the CHREST model (Gobet & Simon, 2000), which has simulated numerous phenomena on chess expertise and in other domains. In this paper, we are interested in the parameter the model uses for shifting chess pieces in its mind's eye (125 ms per piece), a parameter that had been estimated based on relatively sparse experimental evidence. Recently, Waters and Gobet (2008) tested the validity of this parameter in a memory experiment that required players to recall briefly presented positions in which the pieces were placed on the intersections between squares. Position types ranged from game positions to positions where both the piece distribution and location were randomised. CHREST, which assumed that pieces must be centred back to the middle of the squares in the mind's eye before chunks can be recognized, simulated the data fairly well using the default parameter for shifting pieces. The sensitivity analysis presented in the current paper shows that the fit was nearly optimal for all groups of players except the grandmaster group for which, counterintuitively, a slower shifting time gave a better fit. The implications for theory development are discussed.

Keywords: chess; computer modelling; expertise; mental imagery; learning; recall task; sensitivity analysis.

Introduction

As argued powerfully by Newell (1990) and others, one important aim of cognitive science is to develop computational models that account for an increasingly large number of phenomena; the number of free parameters in the models should be kept low, and their value should be made more precise by successive approximations. This aim has inspired the development of CHREST (Chunk Hierarchy and REtrieval STructures; see Gobet et al. 2001, for an overview). CHREST is a model of perception, learning, and expertise that explains the acquisition of knowledge by the growth of a discrimination net, where chunks and templates are stored. It also provides mechanisms explaining how long-term memory (LTM) knowledge directs eye movements. CHREST has accounted for data on chess perception, learning, and memory, the use of diagrammatic information in physics, the acquisition of vocabulary, and the acquisition of syntactic structures. The model has several capacity and time parameters that have been set using empirical data and similar parameters used in other computational models. While these parameters have turned out to be robust in the sense that they have enabled the simulation of numerous empirical data, little work has been done to establish the extent to which their value is optimal or near-optimal. In this paper, we address this question by considering a time parameter important in simulating chess players' mental imagery. This choice is justified not only by the theoretical importance of the parameter, but also by the long-term practical impact that a better understanding of expert mental imagery could have for training and education. We also note that CHREST is one of the very few computational models currently able to make quantitative predictions about expert behaviour in tasks requiring mental imagery.

We first review the experimental evidence on chess mental imagery, some of which was used by De Groot and Gobet (1996) for estimating our target parameter, and then provide an overview of CHREST. Next, we present in some detail a recent study by Waters and Gobet (2008), where CHREST's predictions about the role of chunking in mental imagery were studied through a recall task. One aim of their study was to indirectly test CHREST's time parameter for shifting a piece in the mind's eye by half a square (thereafter, *shifting parameter*), and we summarize how the model's predictions were met by the results of the experiment they carried out. The main part of the paper consists in a sensitivity analysis, where we examine to what extent CHREST's simulations can be improved by searching the optimal value of the shifting parameter, which was set *a priori* in the original simulations. The discussion section highlights how CHREST could be improved using the outcome of the sensitivity analysis.

Mental Imagery in Chess: Experimental Data

The available experimental evidence on mental imagery in chess comes from two main sources: Experiments on blindfold chess, and experiments that have attempted to measure the time needed to move a piece in the mind's eye.

Blindfold Chess

Blindfold chess is a spectacular form of chess where players play one or several games simultaneously without seeing the

board. In a series of ingenious experiments (Saariluoma, 1991; Saariluoma & Kalakoski, 1997), games were presented aurally or visually, with or without interfering tasks. With aural presentation, the games were dictated using a standard chess notation (the algebraic chess notation). With visual presentation, only the current move was displayed on a computer screen (the remainder of the pieces were absent). For the present purposes, the most significant finding was that while performance was not affected by the presentation mode (auditory or visual), the modality of interfering tasks (verbal or visual) had a significant effect. Blindfold chess does not appear to rely much on verbal working memory, but makes heavy use of visuo-spatial working memory, with the qualification that the importance of visuo-spatial working memory is limited to the early stages of encoding. Once stored in long-term memory (LTM), information about positions becomes insensitive to tasks interfering with working memory. Finally, by manipulating the randomness of positions or the location of groups of pieces, Saariluoma (1991) obtained additional support for Chase and Simon's (1973) hypothesis that perceptual chunks underpin skill in chess.

Imagery Time Tasks

The second group of experiments attempted to examine the variables influencing the time to carry out chess moves in the mind's eye. Church and Church (1977) required a single Class A player to report whether the black King was being attacked (or not) by a lone white piece. The decision time (for the check verification task) increased as a function of the distance between the two pieces for diagonal moves, but not for horizontal/vertical moves. Milojkovic (1982) instructed participants to mentally carry out a particular capture in a subsequently presented position (P1). The chessboard remained on the screen as the capture was mentally performed ("P2"). The task was to decide whether the (imagined) position after capture ("P2") was the same as another position (P3) (which appeared after P1 had been removed from view). The Master's reaction times were faster than those of the novices. With both skill levels, reaction times depended on the distance between the two pieces involved in the capture. Novices, but not the Master, took longer with diagonal moves than with horizontal/vertical moves. Gruber (1991), who used the largest sample size of the studies mentioned in this section (24 experts, 24 novices), obtained the same results as Milojkovic in a check verification task: A significant skill effect, a significant distance effect, and a significant interaction between skill and movement-type.

In Bachman and Oit's (1992) experiment, chess players and non-players were presented with either an 8 x 8 grid or a chessboard. They were then required to close their eyes, listen to a sequence of instructions about the moves of a spot or a chess piece (up, down, right or left), and imagine following the spot or piece at it moves. At the end of the sequence of moves, participants had to indicate the end position of the spot or the piece. There were no skill

differences in the moving-spot (8 x 8 grid) condition, but non-players made more errors than chess players in the moving-chess piece (chessboard) condition. Furthermore, in the moving-chess piece condition, skilled players tended to show Stroop-like interference when the piece was required to move in an unnatural fashion. For example, chess players found it difficult to imagine a Bishop moving horizontally (which is incongruent with its typical diagonal movement).

The CHREST Theory

CHREST consists of four main components: an LTM, where chunks are stored; a visual short-term memory (STM) with a capacity of 3 items; a mind's eye system; and a simulated eye. LTM chunks are accessed by traversing a discrimination net (Simon, 1979). A discrimination net is a treelike structure consisting of a set of nodes connected by links. The links have tests, which are applied to check features of the external stimuli. The outcome of each test determines which link will be taken below a node. When a new object is presented to the model, it is sorted through the discrimination net, starting from the root node, until no further test applies. When a node is reached at the end of this process, the object is compared with the image of the node, which is the internal representation of the object. Two learning mechanisms are used. If the image under-represents the object, new features are added to the image, by the process of *familiarization*, which takes 2 s. If the information in the image and the object differ on some feature or some sub-element, a new link and a new node are created below the current node, by the process of *discrimination*, which takes 8 s.

Chunks that are often recognized evolve into more complex data structures, known as templates, which have slots allowing variables to be instantiated rapidly (filling in information into a template slot takes 250 ms). In particular, information about piece location, piece type, or chunks can be (recursively) encoded into template slots. Slots are created at chunks where there is substantial variation in squares, pieces, or groups of pieces in the test links below. In addition to slots, templates contain a core, basically similar to the information stored in chunks. Chunks and templates can be linked to other information stored in LTM, such as (sequences of) moves.

The mind's eye stores perceptual structures, both from external inputs and from memory stores, for a short time (cf. Chase & Simon, 1973). The visuo-spatial information stored there can be subjected to visuo-spatial mental operations; in chess, it is the place where, for example, the trajectories of pieces are computed. The information stored in the mind's eye decays rapidly and needs to be updated either by inputs from the external world or by inputs from memory structures. This assumption is in line with Kosslyn's influential work on mental images (e.g., Kosslyn, 1994).

CHREST makes several assumptions about the operations that are carried out in the mind's eye. For chess, these operations include the time to move a piece mentally. These mental operations are assumed to take a definite amount of
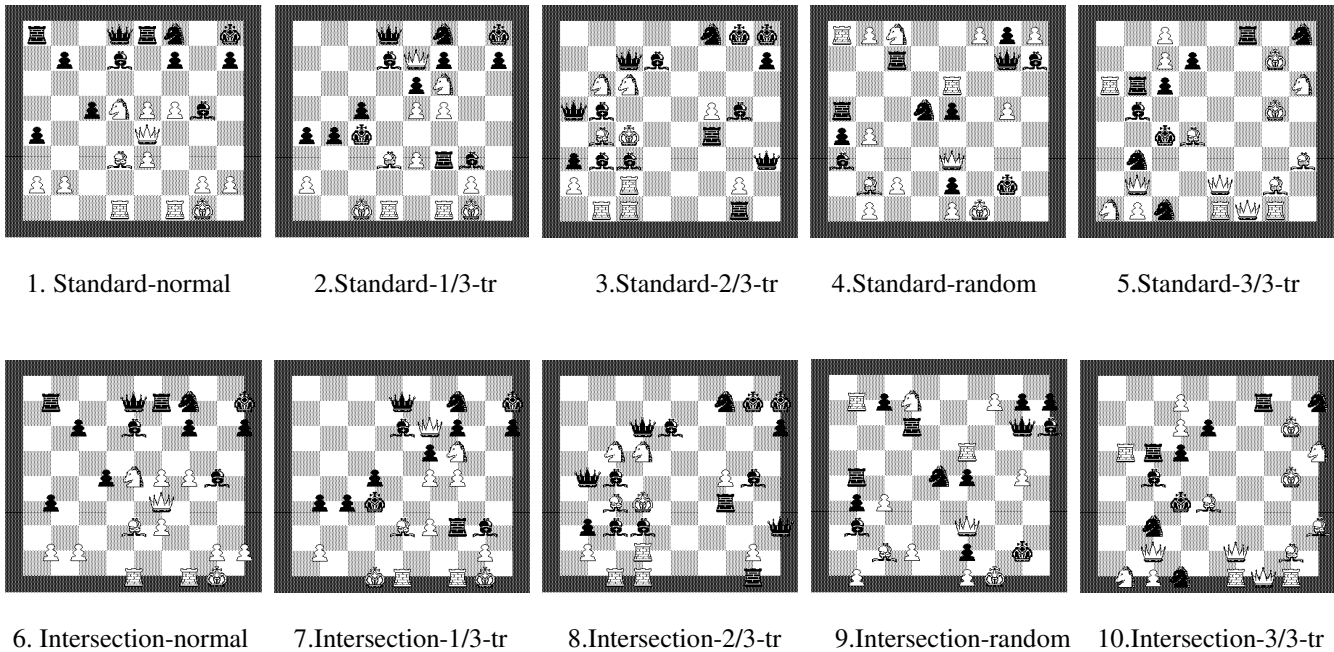
Figure 1: Examples of the 5 types of stimuli used in the Gobet and Waters (2003; positions 1 to 5) and Waters and Gobet (2008; positions 6 to 10). "tr" stands for "truly randomized." (To make comparison easier, the same original position is used in the 10 conditions; different positions were actually used in the experiments.)

time and are carried out serially (see Kosslyn, Cave, Provost, & Von Gierke, 1988, for data supporting the assumption that mental images are generated serially). In addition, the theory postulates definite mechanisms linking LTM, short-term memory (STM), and the mind's eye (see Waters & Gobet, 2008, for details). The eye movements are directed from a combination of acquired knowledge, mediated by the structure of the discrimination net, and heuristics, such as fixating a part of the board about which nothing is known yet (De Groot & Gobet, 1996).

Waters and Gobet's (2008) Study

De Groot and Gobet (1996, p. 236) proposed definite parameters for the time to move pieces in the mind's eye. These parameters were derived from the few experiments available, which sometimes led to inconsistent results (see above). Based on this admittedly imperfect evidence, De Groot and Gobet opted for two parameters: first, a *base* parameter, set to 100 ms both for masters and novices; second, a *square* parameter, set to 50 ms for the masters and 100 ms for the novices. The base parameter refers to the time needed to start the process of generating a move, while the square parameter estimates the time needed to move a piece over one square in the mind's eye. For example, when a strong player imagines a bishop moving from the square "a1" (left bottom corner of the board) to the square "h8" (right top corner of the board), CHREST predicts that this takes 450 ms (100 ms to start the process and 7 x 50 ms per square). Although these parameters are plausible—they

were derived from empirical data—they had not been tested directly, and the aim of Waters and Gobet's (2008) study was to directly test their validity. (As there were no novices in their experiment, Waters and Gobet used only the "masters" parameters.)

Waters and Gobet created "intersection positions," where the pieces were placed at the intersection of squares, rather than being placed in the middle of the squares (see Figure 1). If chunks are recognized without the need to re-center pieces, then recall on the intersection positions should not differ from that on the standard positions. If, on the other hand, pieces need to be re-centered before chunks can be recognized, then there should be a decrease in performance. This decrease can be predicted by CHREST, and the parameter to shift pieces diagonally is crucial for these predictions.

In addition, the ease by which chunks could be accessed in LTM was manipulated using positions with different levels of structure. These positions, which contained 25 pieces on average, varied from game positions to fully randomized positions (see Figure 1, and see Gobet & Waters, 2003, for the detail of how these positions were constructed). Game positions were taken from master games without any change. Random positions were constructed by randomly reassigning the pieces of a game position to new squares. In "truly" random positions, not only the location of the pieces was randomized, but also the distribution of pieces (e.g., there could be 12 white kings in a position, contrary to the standard chess rules). One-third and two-

third truly random positions were positions where 1/3 and 2/3 of the pieces were truly randomized. On intersection positions, the end product was manipulated by shifting all pieces to the south-east corners of the squares.

The same nets as those used by Gobet and Waters (2003) in their simulation of standard positions were selected. These nets, defined by the number of LTM chunks, were selected as they fairly closely matched the mean recall of the four groups of human subjects on standard game positions. They were created by letting the program scan a large number of positions, so that chunks and templates can be acquired. The result of the simulations, both for standard and intersection positions, is shown in Figure 2 (left panel).

These predictions were tested with a sample of 36 players: a "grandmasters" group, a "masters/experts" group, a "Class A/B players" group (consisting of moderate to strong club players), and a "Class C/D players" group (consisting of weak club players). On each trial, a position was presented for 5 s. The screen then was blank for 2 s, and then an empty chess board appeared. The participants were instructed to try to recall the positions as completely and as accurately as possible. Further details of the participants and experimental procedures are available in Waters and Gobet (2008).

Figure 2 (right panel) shows the results. Recall was impaired on the intersection positions compared to the standard positions. This impairment was especially pronounced on the intersection game positions. Skill effects were present on the intersection game positions, but not on the other intersection positions.

Interestingly, participants were better at recalling bishops than knights on the intersection positions (but not the standard positions), which is in line with Bachman and Oit's (1992) study reviewed above and also supports the hypothesis that mental imagery played a role in this task: the mental transformations were easier for the bishops than the knights.

These results suggest that, as predicted by CHREST, human information processing is slowed down by the processes of carrying out mental transformations to re-centre pieces, which impairs the ability to access chunks/templates in the intersection game positions. However, while the fit for the simulation with intersection positions is good (all $r^2 >= .90$; see Table 1, top), there is still room for improvement. Given that the re-centring process occurs often, and given the importance of the 125 ms in this process, examining this parameter is a natural place to start for improving the model's fit to the data.

Sensitivity Analysis

The transition-time of 125 ms was chosen for theoretical reasons, based on the estimate provided by De Groot and Gobet (1996) for moving a piece diagonally in the mind's eye. To investigate the role of this parameter, we systematically varied it from 0 ms to 450 ms. The simulations otherwise followed the same procedure as that used in Waters and Gobet (2008).

Methods

We used the same version of CHREST and the same networks as those used by Gobet and Waters (2003) and Waters and Gobet (2008). During the simulations, the model moves its simulated eye around the board, and attempts to recognize chunks (or templates). The presentation time for each position was 5 s. Given this relatively short presentation time, the model could add information to LTM by familiarization and by filling in information into a template slot, but not by discrimination.
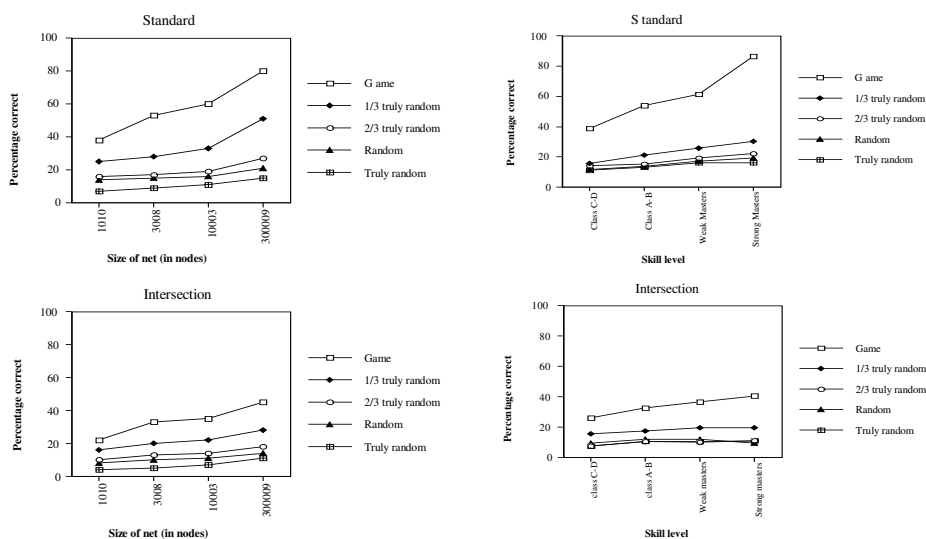

Figure 2: Percentage of correct pieces as a function of type of position for standard positions (upper panel) and intersection positions (lower panel). Left: computer simulations; right: human data. Human data are from Gobet and Waters (2003) for the standard positions, and from Waters and Gobet (2008) for the intersection positions.

The model attempts to memorize the intersection positions by carrying out the following steps: (a) in the mind's eye, up to three pieces within the external visual field are moved serially to the centre of the square (as with previous versions of CHREST, the external visual field is defined as the set of squares +/- 2 squares away from the fixation point); (b) the (shifted) pattern of pieces is sorted through the discrimination net; and (c) if a chunk is recognized for this pattern, it is handled in the same way as with previous simulations using standard positions (Gobet & Simon, 2000; Gobet & Waters, 2003); that is, when an external pattern leads to the successful recognition of an LTM chunk, a pointer to the chunk is placed in visual STM. During recall, the model shifts the pieces back to their intersection location. With the exception of these mechanisms for handling intersection positions, CHREST's mechanisms are the same as in previous simulations (i.e., simulations with standard positions). In addition, they are the same in all position types (from game to truly random). Thus, differences in recall performance reflect the probability that patterns present in the positions will elicit chunks or templates in LTM.

To simplify the simulations, the two parameters defined by De Groot and Gobet (1996) were considered as a single parameter, which is the time needed to shift a piece across half a square. This time was systematically varied from 0 ms to 450 ms, by steps of 25 ms. Thus, there were 380 different conditions: 4 (network sizes) x 5 (position types) x 19 (shifting time values). For each condition, 500 positions were used.

Results

As expected, recall was better across all position types and net sizes when transition-time was briefer (all correlations larger than -.87 in absolute value, and all $p < .001$). As measures of goodness of fit with the human data, we computed r^2, the average absolute deviation (AAD), and the sum of squared errors (SSE). R square indicates how well a model captures the pattern of means of the empirical data. SSE and AAD provide information about the deviation of the model data from the empirical data. Higher r^2, and lower AAD and SSE, indicate a better fit. As r^2 was consistently above .93, .92, .90, and .85, for the 300k, 15k, 3k, and 1k nets, respectively, this measure was of little value for discriminating the effect of the parameter change. AAD and SSE produced more differentiated results. We focus on AAD here (With few exceptions, SSE produced similar results).

Figure 3 shows how AAD varies as a function of the transition-time and Net Size. One can see that, for all nets except the 300k net, the minimum AAD value is close to that obtained with 125 ms. However, a better approximation is obtained when the transition-time is optimised for each Net Size. Goodness of fit with the optimal transition-time for each Net Size (325 ms, 175 ms, 100 ms, and 75 ms, for the 300k, 15k, 3k, and 1k nets respectively) is reported in Table 1. As can be seen, there is actually a strong positive

relationship between optimal transition-time and the Net Size, meaning that the fit improves if the stronger simulated players use a longer translation time.
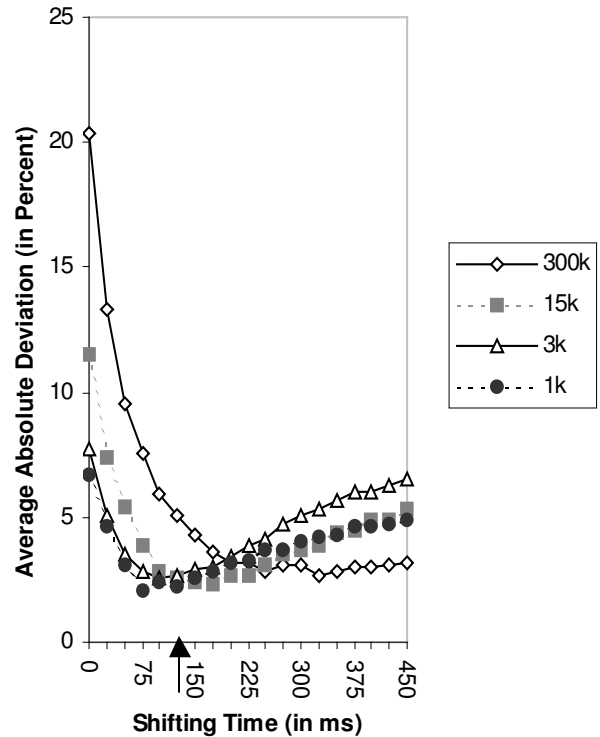

Figure 3: Absolute Average Deviation as a function of size of net and shifting time. The vertical arrow is placed at 125 ms, the value used in the main simulations of this paper.

Table 1: Goodness of fit for (a) the shifting value being equal to 125 ms for the four nets, and (b) the shifting values optimising AAD for each skill level.

net size	shift time	r^2	AAD	SSE
		(a) Fit with 125ms for all nets		
300k	125	0.94	5.1	165.7
15k	125	0.94	2.6	42.2
3k	125	0.91	2.7	44.4
1k	125	0.90	2.2	30.5
Average		0.92	3.1	70.7
		(b) Fit with time values minimizing AAD		
300k	325	0.95	2.7	52.6
15k	175	0.94	2.3	32.0
3k	100	0.90	2.6	44.0
1k	75	0.89	2.1	35.5
Average		0.92	2.4	41.0

Discussion

The simulations carried out by Waters and Gobet (2008) on the recall of intersection positions showed that CHREST made several correct predictions with respect to the mechanisms putatively carried out in the mind's eye. The predictions were made with parameters that were set *a priori* and independently from the data. Thus, the time parameters of CHREST were also well supported, in particular the shifting time for re-centring pieces. This value (125 ms) provided a fairly good fit for the data for all nets (see Figure 2 and Table 1). In general, the data supported the idea that pieces must be re-centred in the mind's eye before pattern recognition can happen, and they also provided support for the assumption that the shifting time would be 125 ms.

These results were satisfactory, considering that CHREST was not developed *ad hoc* to account for the results of the intersection experiment and that it makes absolute predictions about performance. However, an important scientific question is to know the extent to which the simulations could be improved by optimising some of the model's parameters. This was the aim of this paper, which we addressed by carrying out a sensitivity analysis on the shifting time parameter.

As expected, increasing the transition time from 0 to 450 ms led to poorer recall over all position types. In addition, the sensitivity analysis revealed that the best fit was obtained with a shifting time (325 ms) for the 300k-chunk net that was of longer duration than the values for the smaller nets (Table 2). In general, the shifting times producing the best fit were positively correlated with skill (as estimated by the number of chunks), which is somewhat counter-intuitive. One would have expected that strong players should be faster in moving pieces in their mind's eye. A likely explanation for this unexpected result is that stronger players keep more pieces in the mind's eye, and that this produces an overhead affecting piece translation. That is, in our simulations, shifting pieces to the centre of the square was the only process that was assumed to be new in the intersection condition, and the impact of this process was a direct function of the shifting time. However, it is plausible that, as more pieces are held in the mind's eye – and the model assumes that chunks held in visual STM are automatically unpacked in the mind's eye, with the consequence that stronger players hold more pieces there – additional processes must happen to refresh the mental images. For example, Kosslyn (1994, p. 322) assumes that "the amount of material one can hold in an image is limited by the number of stored units that can be activated at the same time, for the following reasons: Each unit is activated individually, and time is required for each operation. And as soon as a unit has been activated, the image begins to fade […]." Thus, the fact that slow values give a better fit with stronger players may be an artefact of the fact that we did not simulate in detail how information in the mind's eye is maintained to counteract decay.

Waters and Gobet (2008) carried out the simulations this way in order to keep the model as close as possible to earlier simulations, but it now appears that in this instance the model may have been too simple: while the simulations presented in that paper accounted for the data reasonably well, the sensitivity analysis pinpointed one aspect of the model that needs further development. Later versions of CHREST will have to take into consideration the detail of how the mind's eye generates and maintains visual images. This will also make it possible to simulate the Stroop-like difference in recall of bishops and knights found by Waters and Gobet (2008), which the current version of the model cannot account for.

References

Bachman, T., & Oit, M. (1992). Stroop-like interference in chess players' imagery. *Psychological Research, 54*, 27-31.

Chase, W.G., & Simon, H.A. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

Church, R.M., & Church, K.W. (1977). Plans, goals, and search strategies for the selection of a move in chess. In P.W. Frey (Ed.), *Chess skill in man and machine* (pp. 131-156). New York: Springer Verlag.

De Groot, A.D., & Gobet, F. (1996). *Perception and memory in chess.* Assen: Van Gorcum.

Simon, H. A. (1979). *Models of thought* (Vol. 1). New Haven, CT: Yale University Press.

Gobet, F., Lane, P.C.R., Croker, S., Cheng, P.C., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*, 236-243.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24*, 651-682.

Gobet, F., & Waters, A.J. (2003). The role of constraints in expert memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 29*, 1082-1094.

Gruber, H. (1991). *Qualitative Aspekte von Expertise im Schach*. Aachen: Feenschach.

Kosslyn, S. M. (1994). *Images and brain: The resolution of the imagery debate*. Cambridge, MA: Bradford.

Kosslyn, S.M., Cave, C.B., Provost, D.A., & Von Gierke, S.M. (1988). Sequential processes in image generation. *Cognitive Psychology, 20*, 319-343.

Milojkovic, J.D. (1982). Chess imagery in novice and master. *Journal of Mental Imagery, 6*, 125-144.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Saariluoma, P. (1991). Aspects of skilled imagery in blinfold chess. *Acta Psychologica, 77*, 65-89.

Saariluoma, P., & Kalakoski, V. (1997). Skilled imagery and long-term working memory. *American Journal of Psychology, 110*, 177-201.

Waters, A.J., & Gobet, F. (2008). Mental imagery and chunks: Empirical and computational findings. *Memory & Cognition*, *36,* 505-517.