

Attention Mechanisms in the CHREST Cognitive Architecture

Peter C. R. Lane¹, Fernand Gobet², and Richard Ll. Smith³

¹ School of Computer Science, University of Hertfordshire, peter.lane@bcs.org.uk

² School of Social Sciences, Brunel University, fernand.gobet@brunel.ac.uk

³ School of Social Sciences, Brunel University, richard.smith@brunel.ac.uk

Abstract. In this paper, we describe the attention mechanisms in CHREST, a computational architecture of human visual expertise. CHREST organises information acquired by direct experience from the world in the form of *chunks*. These chunks are searched for, and verified, by a unique set of heuristics, comprising the attention mechanism. We explain how the attention mechanism combines bottom-up and top-down heuristics from internal and external sources of information. We describe some experimental evidence demonstrating the correspondence of CHREST's perceptual mechanisms with those of human subjects. Finally, we discuss how visual attention can play an important role in actions carried out by human experts in domains such as chess.

1 Introduction

Cognitive science studies the processes by which humans develop and manifest intelligent behaviour. The study of visual perception has been widely recognised as an important component of many areas of expertise. A seminal experiment by de Groot [1] uncovered a central component of human expertise: the ability to identify the important features of a stimulus in the domain of expertise. De Groot's experiments were performed on chess players, and involved a test of *recall ability*. Each participant was shown a position on a chess board containing approximately 23 pieces for a few seconds, and was then asked to reconstruct the position from memory. Candidates for the world title managed to reconstruct the position with few, if any, errors; average players managed much worse. As there was no difference in the participants' intelligence level, other visual skills or general memory, clearly the difference was related to their level of expertise.

An explanation for the difference had to await the development of the chunking theory [2, 3] and cognitive models of human learning such as EPAM (Elementary Perceiver And Memoriser) [4]; the most detailed model of chess expertise is now CHREST (Chunk Hierarchy REtrieval STructures) [5–8], which is a version of EPAM with the addition of templates and more sophisticated perceptual mechanisms. The computational modelling and experiments have shown that, essentially, the difference in perceptual skills can be explained by two factors: the development of a large (approximately 300,000 chunks) set of knowledge about

the domain, and the use of templates by the visual system to actively seek out higher-order clusters of information in a stimulus.

We can summarise the role of attention within CHREST as follows. First, an image is perceived, and, as the eye has a limited field of view, a portion of the image has its features extracted. These features are then used to sort through long-term memory, seeking a familiar pattern. Any retrieved pattern is placed into short-term memory. The contents of short-term memory, some high-level domain-specific knowledge, and any items on the periphery of the field-of-view will all combine to guide the model's eye to locate a new point of the image to focus on. This process continues, and the model will attempt to build up, in short-term memory, a set of pointers to familiar patterns in long-term memory which 'cover' the image. Here we see the importance of prior experience, as the capacity of short-term memory is limited, but experts will have larger familiar patterns (chunks), and so can store much more relevant information in their short-term memory. However, attention is also important, as the larger patterns must be confirmed to be present in the image if they are to be maintained in short-term memory. Later, we explain in detail how a set of heuristics manages this process in a dynamic and flexible manner.

This paper continues with an overview of the CHREST architecture, details of the attention mechanism, and summaries of some results from experiments using CHREST to explore visual abilities. Finally, we discuss how the findings with CHREST relate to current issues in the study of attention.

2 Overview of CHREST

CHREST is an example of a *cognitive architecture*, that is, an implementation of a theory of human cognition for developing detailed models of human behaviour in a range of domains. The strength of cognitive architectures as scientific theories is that their implementation ensures a high degree of precision in the theory's formulation, providing a sufficiency proof that the proposed mechanisms can carry out the tasks in the domain of interest. Analysis of how the theory's predictions match actual behaviour, using measures such as eye movements, reaction times, and error patterns, also establish the quality of the cognitive architecture, by confirming its behaviour against actual human data.

Together with EPAM, from which it is derived, the CHREST architecture has been a source of successful models of human learning and perception over a 50-year period [3]. Beyond the results discussed in this paper, these successes have included results in: verbal learning, language acquisition [9], categorisation and problem solving. In this section, we provide an overview of CHREST, explain its key processes, and highlight some important empirical results. The next section will focus in more detail on the important attention mechanisms.

Fig. 1 shows the principal components of CHREST. The architecture follows a classic subdivision of the cognitive processes into three components: one of input/output and two of memory. The input/output mechanisms are used to interact with the environment. Input stimuli are separated into features, and out-

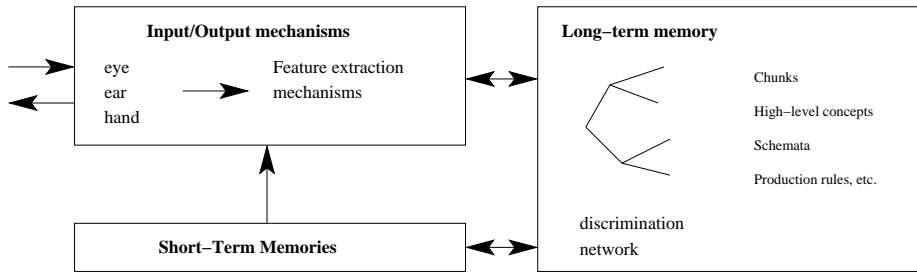


Fig. 1. The CHREST cognitive architecture

put actions are converted into motor controls. The Long-Term Memory (LTM) is used to store information in a more-or-less permanent state; all information learnt from the environment is held in LTM. The LTM uses a discrimination network, which is constructed incrementally during the system's lifetime, to index a pool of familiar patterns (chunks), higher-level concepts, schemata and productions. The final component of the architecture is the Short-Term Memory (STM), which provides a temporary store for information retrieved from LTM whilst considering a particular stimulus or problem. Different parts of STM relate to different modalities of information, such as visual or verbal. The STM maintains an *hypothesis node*, which is the most informative node retrieved so far, and is used to focus CHREST on retrieving or constructing large chunks.

CHREST is an example of a symbolic cognitive architecture, which means that all information retrieved by and stored within a CHREST model is at a high level. We shall use the term 'patterns' and 'chunks' throughout this paper. A pattern is what CHREST perceives in the external world, and a chunk is a familiar pattern, one retrieved from its internal memory. Both patterns and chunks are represented as lists of items with their positions. For example, in chess, the items are actual pieces, and the positions are squares on the chess board, so (Kg1 Rf1 Ph2 Pg2 Pf2) would describe a typical castled position for the White player.

3 CHREST's LTM: Chunking networks

CHREST's long-term memory is a *chunking network*: a discrimination-network representation with cognitively plausible learning mechanisms, based on the chunking theory. The chunking network includes mechanisms for forming and using *lateral links*, as well as clustering techniques, to form *templates*.

3.1 Learning and retrieving perceptual chunks

A *chunk* is a familiar pattern: CHREST acquires and retrieves chunks from its long-term memory through a set of learning and storing operations acting on a discrimination network. The network is formed from *nodes* holding the chunks;

the chunk stored at a node is known as its *image*. There is a *root node* from which all sorting operations begin, and the discrimination network is built up from the *test links* between nodes. Patterns are sorted through the network by following those test links whose tests match the given pattern.

The network is constructed using two learning mechanisms. *Discrimination* is used when a pattern is sorted to a node which has no succeeding test links matching the pattern, and the pattern mismatches the chunk stored at the node. The part of the pattern which mismatches the chunk is used as the test for a new test link. *Familiarisation* occurs when a pattern is sorted to a node which has no succeeding test links matching the pattern, and the pattern matches the chunk stored at the node. If the pattern contains more information than is held at the node, then extra information is added to the stored chunk. Thus, node images are specialised as patterns are re-encountered.

In describing the discrimination and familiarisation operations, part of the sorted pattern is used, either as a new test, or to augment an existing chunk. The amount of the sorted pattern used is a function of the rest of the network in the following sense. The part of the pattern to be used is *re-sorted* in the network, and the chunk that is retrieved is used in the above processes. This chunking process ensures that learning is slow in the initial stages of getting to know a domain, but as the network grows, and some of the patterns become known, the rate of learning will speed up.

Efficiency of accessing a large store of data is often a concern, as in the utility problem [10]. However, the usable capacity of a data structure need not be problematic if it is sufficient in the given domain [11]. Human experts typically pick out familiar patterns in a time of around 250ms, which is easily matched by CHREST, using a network of around 300,000 chunks [8].

3.2 Constructing templates

The *template* is a critical element of CHREST's explanation of the recall abilities of human experts [8]. A template collects together the information stored in a number of separate nodes so as to highlight the constant *core* information and the variable information, held in *slots*.

Template creation is assumed to occur whenever a critical condition is met by a specific node within the network. Similarity between nodes is made explicit by providing *similarity links* between them (such links are discussed in the next subsection). When a node has a sufficient number of such links, and the contents of these nodes satisfy an overlap criterion, the information in these nodes is aggregated to form a template. Work in chess [8] has assumed a threshold for the number of similarity links of 4, with at least 5 features in common.

3.3 Lateral links

Lateral links [12] are created when the model has retrieved two chunks (retrieved chunks are stored in STM); two classes of lateral link may be distinguished. The first is where the two nodes match a similarity function of some kind. The

1. Clears the record of which heuristic was last used.
2. Selects a point in the centre of the field of view to start from.
3. Performs following cycle:
 - (a) Store the currently fixated item in the list ‘fixated-items’
 - (b) Gets the next fixation point using `get-next-fixation` (see text).
4. If timing is used and model has used its presentation time, or if the maximum number of allowed fixations is reached, then finish.
5. Otherwise, repeat from step 3.

Fig. 2. CHREST’s perception cycle; learning occurs during `get-next-fixation`.

simplest of these is the direct *similarity link*, as used above in the creation of templates. More complex *generative links* [9] connect those nodes with similar descendant test links. The second class of links are those where the chunks are of different type. For instance, *production links* are used to associate a perceptual pattern with its corresponding action or conclusion, and so the production link forms a basis for the model’s problem-solving behaviour. Gobet et al. [3] provides a summary of these links and their use in different domains; Section 5.2 provides more details on how such links are formed.

4 CHREST’s Attention Mechanism

CHREST uses a simulated eye to retrieve information from its target stimulus. The eye is directed to a focus of attention, the *fixation point*, and has a limited field of view. Its movement is governed by a set of heuristics, which combine low-level and high-level information. A *perception-learning-perception* cycle guides CHREST’s eye movements around the current stimulus for the presentation time; see Fig. 2. In the following, we describe the heuristics and perception processes to be introduced in the latest release of CHREST, version 3.0.

The list of ‘fixated-items’ is part of CHREST’s short-term memory. The fixated-items record the items and their positions fixated upon during the current presentation cycle. Each item observed will be added to this list until a termination condition is met. This termination condition is when an empty location is fixated, or CHREST fixates something already in the list, or the heuristic used is a random or weak heuristic. Marking the list as fixated will then lead to the fixated-items list being learnt as a pattern. After learning, the fixated-items list is cleared, and CHREST will begin building a new list, as it continues its perception cycle. Worth noting is the implicit connection between perception and learning; the list of fixated items is constructed by sequences of eye fixations, which means that information learnt by CHREST has a locality bias.

Heuristics for selecting eye fixations can come from two sources. First, there are the generic, domain-neutral heuristics which are part of the general architecture. These heuristics include: using LTM, fixating a part of the scene not observed yet, locating a random object on the periphery, and locating a random position on the periphery. Second, there are domain-specific heuristics which are

part of the model for that domain alone. For example, in the chess models there is an heuristic to guide the eye to typical positions of the kings; attacking or defending the king is the ultimate key to winning at chess, and so the king position is an important initial factor in assessing a position. Again in chess, there is an heuristic to guide the eye to a square attacked by the currently fixated piece.

The process `get-next-fixation` is responsible for applying the various heuristics to retrieve a new fixation point. The process uses a hierarchy of heuristics: first it tries to use LTM, second it tries fixating a new object or a position suggested by the domain-specific heuristics, finally it defaults to choosing a new random position. There is a stochastic element, which means CHREST will fail to use an applicable heuristic in a percentage of cases (currently set at 10%).

As an example we describe in more detail the most interesting heuristic, which selects a new square based on information in LTM. This heuristic uses pointers into long-term memory as a basis for selecting the next fixation point. There are two main sources. The first is the *hypothesis node*, held in short-term memory, which acts as an anchor, guiding CHREST to retrieve or construct the largest chunk possible for the given stimulus. The test links of the hypothesis node (its descendants) are considered in turn, as described below. The second is the node retrieved by learning something from the currently perceived scene. The process is as follows: (1) `current-node` is the result of learning from the current scene; (2) if there are any remaining descendants of the hypothesis to consider, let `current-child` be the first descendant, and remove it from the remaining list; (this step attempts to retrieve the largest chunk) (3) otherwise, let `current-child` be the first child-link of the `current-node`; (4) return the first potential square of `current-child` as the next fixation point. The learning that can occur in step (1) attempts to extend the hypothesis with information currently perceived, or else using the list of fixated-items, if that list has been marked as complete; hence, CHREST is biased towards learning the largest chunk possible.

One consequence of linking low-level perceptual processes directly with the higher-level information stored within the long-term memory is a blurring of the boundary between primitive and complex visual objects. We assume that CHREST initially contains single-element chunks for each of the primitive elements which may occur in the given domain, e.g. if CHREST is scanning text documents, all the single letters will be provided as primitive features. As CHREST learns about the domain, it begins to form larger chunks, consisting of groups of these primitive features. In recognition, CHREST will typically retrieve complete chunks from the domain, bypassing the more elementary features. Thus, in scanning text documents, CHREST soon begins to work with words as ‘primitive’ elements, instead of single letters. Such creation and use of higher-level features occurs in any domain, and is typical of the perceptual knowledge acquired and used by experts. It is also a typically hard problem in unsupervised learning tasks, especially perceptual domains [13, 14].

5 Experiments with CHREST on Attention

We describe two sets of experiments from previous work to illustrate the impact of CHREST’s perceptual mechanisms: data from chess, on the details of the human attention mechanism; and data from a word-recognition task, explaining how expectations assist in disambiguating input data.

5.1 Chess

In the introduction of this paper, we mentioned de Groot’s [1] classic experiments on chess memory. It is of interest that these experiments were motivated by a need to understand chess masters’ perception and attention: how can a world-class master understand more of a position after five seconds than a candidate master after fifteen minutes? How can search be so selective with strong players, who often consider no more than one hundred positions during their thinking, and ignore the billions of positions that could, in principle (but not in practice), be of interest? CHREST provides mechanisms explaining how strong players attend to the relevant while ignoring the irrelevant. These mechanisms have been used to explain both how players direct their attention when considering a position, and how they can limit the number of moves they anticipate during look-ahead search. The organisation of these two types of behaviour is hierarchical, in the sense that the pattern of eye movements is part of the explanation of selectivity at the level of move choice. In both cases, computational simulations have reproduced key aspects of human behaviour. Given space constraints, we limit our attention to three classes of phenomena simulated by CHREST.

Eye movements as indicators of attention In their analysis of players’ eye movements during the brief presentation of a chess position, de Groot and Gobet [5] identified a few striking differences between novices and masters. Masters’ eye movements were shorter on average than novices’ (260 ms vs. 310 ms), and also showed less variability (sd = 100 ms vs. sd = 140 ms). Masters’ eyes covered more squares on the board, and also covered more of the squares that were important in the position. CHREST does a very good job of simulating these data. For example, it replicates the skill effect with the average duration of eye movements (272 ms for the simulated masters vs. 315 ms for the simulated novices) and the difference in variability (97 ms vs. 154 ms). Fig. 3 shows an example set of eye fixations; what is important is how CHREST covers approximately the same amount of the important part of the board as the expert, whereas the novice player and model (not shown) cover a much smaller area. CHREST also captures the skill differences in the percentage of the board covered, and in the percentage of critical squares covered. The speeding-up of the eye movements and the increased number of important squares fixated are due to the fact that many more fixations are directed by the structure of the discrimination network with the master version of the model than with the novice version, as the discrimination network is larger with the former than with the latter.

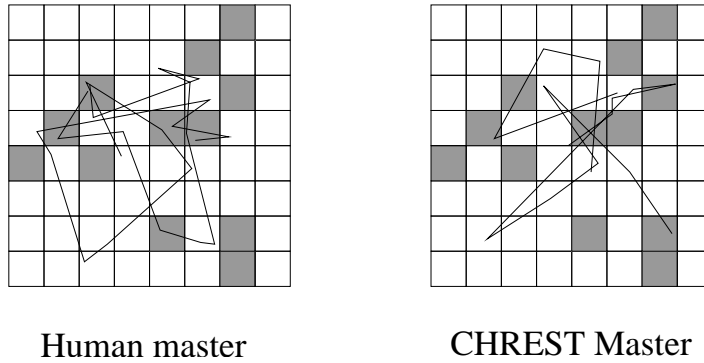


Fig. 3. Example of a Master’s eye movements for a specific position (left) and its simulation by CHREST (right). Grey squares indicate important squares in the position. The field of view will cover several squares around each fixation point. (After de Groot and Gobet [5].)

More qualitatively, CHREST captures masters’ tendency to fixate perceptually salient pieces (e.g. a white knight on the black side of the board) early on, and their tendency not to fixate parts of the board that are “normal” in a given situation (e.g. a standard castling position).

Although we have emphasised the way perceptual knowledge helps develop more efficient attention, it should be pointed out that masters’ attention is not infallible. In several of the positions used by de Groot and Gobet, immediate threats (e.g. checkmate in one move) were missed by some of the masters. The longer the distance between the two pieces, the more likely it was for the threat to be missed. Assuming that chunks play a key role in rapidly identifying threats, as we have done in this paper, this result is in line with the way CHREST learns chunks, giving precedence to relations of proximity, as described next.

The structure of chunks: An archaeology of attention A crucial assumption in CHREST is that the information stored as chunks is a reflection of the attention mechanisms used during learning. This offers an indirect way of testing how attention is directed when playing chess. Concurrent and retrospective protocols [1, 5] suggest that a fair amount of attention is directed to consideration of moves and counter moves. However, they also suggest that attention is directed to patterns of pieces. Chase and Simon [15] analysed the structure of the groups of pieces recalled together in a memory task, and found that a surprisingly small number of these groups contained relations of attack. The vast majority of these groups contained relations of proximity, same colour and defence. This result turns out to be robust, and has been replicated by [16] with a larger sample. In a simulation of the memory experiment used by Chase and Simon, Gobet [17] showed that CHREST closely simulates the detail of the pattern of relations found in chessplayers’ chunks. This outcome supports the importance given by CHREST to proximity during the acquisition of chunks.

Selectivity in choosing a move A characteristic of experts is the rapidity with which they can propose solutions. For example, chess grandmasters literally ‘see’ the good moves straight away [1, 18]. With very short decision times, less than 10 seconds, their choice will not always be the best possible move, however, it will almost always be a very plausible move in the position. CHUMP (CHUnks and Moves Patterns) [19] is a variant of CHREST that implements the idea that recognising patterns of chess pieces on the board makes it possible to access information about moves in long-term memory, and thus to rapidly identify fairly good moves. CHUMP uses two different but linked discrimination networks to store two types of knowledge: first, patterns of pieces (the kind of chunks acquired by CHREST), and, second, moves and sequences of moves (see the next section for more on multi-modal learning in CHREST). During learning, the program scans positions taken from master games, and patterns of pieces are associated with moves. When selecting a move in a position in the test phase, patterns of pieces act as conditions, and moves as actions. If different patterns suggest a move, and/or if the same patterns suggest different moves, the conflict is resolved by using a function combining the number of different chunks voting for a given move and the number of times the move has been associated with a given pattern.

CHUMP provides a demonstration that the idea of selectivity through recognition and association can be implemented in a computer program. However, it should be pointed out that the level of play of CHUMP is rather low. This is because it plays chess by pure pattern recognition, without being able to look ahead, and it is well established that look-ahead abilities are important in playing chess and other board games at a high level [20].

5.2 Expectations

Expectations are important in guiding what we look at and how effectively we can recognise what we see. Expected objects are recognised with greater accuracy than unexpected objects, particularly in noisy domains [21]. Expectations may also relate to complex collections of objects, or *schemata*. Perceptual classification of objects within a familiar schema can be quicker than when the objects are not in the schema. For instance, Biederman [22] describes an experiment in which participants took longer to identify a fire-hydrant when positioned above street level than when at its expected position. Finally, noisy or ambiguous scenes may be reconstructed, if the visible elements are constrained to fit a compatible schema; Lindsay and Norman [23] describe such an experiment with words composed of distorted or ambiguous letters.

The main lesson from the results described above is that perception is not a simple flow of information from scene to memory, but instead perception is an active process, with the attention forcing a shift of fixation point to different parts of the scene based partly on what is observed and partly on what is expected (or anticipated) to be present. CHREST has been used to construct a model of this process, and demonstrate the qualitative results described above [24]. The key factor in supporting this process was for CHREST to support links between visual information, representing the scene being analysed, and verbal

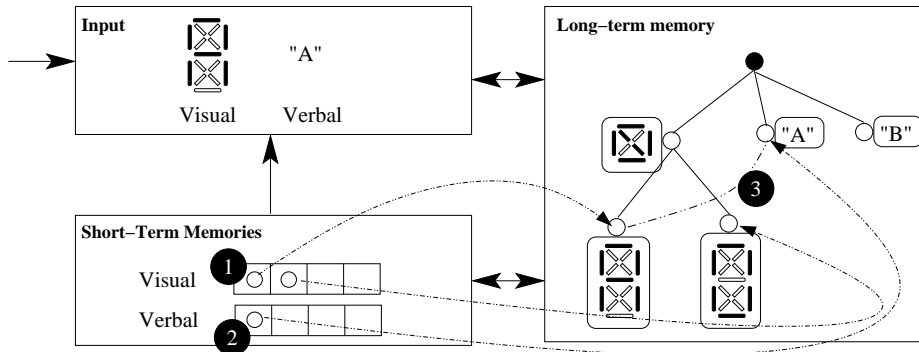


Fig. 4. Learning to associate information across two modalities: (1) visual pattern is sorted through LTM, and a pointer to the node retrieved is placed into visual STM; (2) verbal pattern is sorted through LTM, and a pointer to the node retrieved is placed into verbal STM; and (3) a naming link is formed between the two nodes at the top of the STMs. (Taken from [24].)

information, representing the interpretation of the scene. The links between these two input modalities are formed when the model is presented with visual and verbal information simultaneously, as illustrated in Fig. 4.

In the situation depicted, the model is presented with a visual stimulus and a verbal label. Both the stimuli are sorted through the long-term memory, and pointers to the retrieved nodes are placed in their respective short-term memory. A *naming link* is formed between the two nodes at the top of the two short-term memories. This naming link can be used in future to retrieve the verbal label from a visual stimuli. By a similar process, sequences of letters and words can be acquired and stored in the network.

The attention mechanism can use this stored information in various ways. First, verbal priming (e.g. being told to find particular words) can be used to highlight specific visual chunks, which the attention mechanism will then attempt to locate. Second, when part of a chunk is located, the visual or verbal information can be used to identify schemata which can guide the attention mechanism. For instance, verbally learning a sequence of words can be used to trigger recognition of the visual representations of those same words.

6 Discussion

In this paper, we have described CHREST and shown how it provides an account of the links between perception, attention, learning, memory and action. Thus, CHREST addresses several key issues in the study of attention, and in particular how it links to other aspects of human cognition. One of the differences of CHREST with respect to other theories of attention is the level at which it works; CHREST is a *symbolic* system, meaning that information is stored at a

‘meaningful’ level. The basic information used by CHREST in the chess experiments is the *piece-on-square*, which encodes that a piece is located at a particular position of the board. This perspective distinguishes our work from that by other authors on perception, such as Tsotsos [25] or Wolfe [26]. CHREST’s perspective is psychological, but related to high-level conceptual processing, and is not concerned with neurological or other low-level processes.

The key contribution of this paper is to show how a symbolic treatment of attention can affect our understanding of attention in high-level areas of human cognition, and we now review some of these areas.

Perception is cognition. This was the central focus of the studies reported in de Groot and Gobet’s book [5]. This role has sometimes been overlooked in recent attempts to provide general architectures of cognition. For example, in both Soar [27] and ACT-R [28], perception is considered, but its role is not as central as in CHREST. As we have seen, in CHREST the central theoretical construct is that of a perceptual chunk, and the learning of chunks directly depends on perceptual mechanisms.

Attention directs learning, and that which has been learned directs attention. This relates to the previous point. Expertise in a domain develops because the use of fairly weak heuristics leads to the acquisition of more perceptual knowledge. These heuristics can be either general (“look at a part of the scene you don’t know anything about”) or domain-specific (“verify whether your Queen is attacked”). With sufficient experience, this perceptual knowledge enables the fluid and rapid behaviour that is characteristic of experts’ intuitive decision making.

Action is closely linked to perception. The entire CHREST architecture rests on the assumption that many actions are elicited by perceptual patterns. This is particularly the case with expert behaviour, where the link between action and perception is made automatic.

Attention is directed by a combination of top-down and bottom-up processes. The CHREST simulations show how novices rely more on weak heuristics to direct their attention, while the masters rely more on the structure of the discrimination network to suggest the next fixation position. Note that some of the (top-down) heuristics used by the masters may also be useful as information-seeking devices. Top-down heuristics are also important to make sure that global goals are heeded.

Context matters in focusing attention. In CHREST, this is readily captured by the fact that the fixations directed by the discrimination network are sensitive to even fine details in the context. Indeed, different positions lead to different eye movement patterns. Just like human masters, CHREST does not use stereotyped sequences of eye movements, but adapts these as a function of the environment.

Attention is crucial at many levels. While this paper has focused on the role of attention in the first seconds of presentation of a new complex stimulus, it should be pointed out that, in a full model, attention would be essential at several levels. For example, when chess players examine a position in order to choose the best move, attention would be selective, not only in selecting the next fixations, but also in deciding which moves should be selected for further consideration,

and also in deciding what part of the information currently gathered about the position is worth the cost of storing in long-term memory. In both cases, selectivity – and not just random guess – is made possible by knowledge acquired over years of dedicated practice [15].

Why is attention limited? And why is the capacity of short-term memory limited? From an evolutionary standpoint, limits in attention and short-term memory capacity make sense; when facing a danger, organisms that immediately identify the presence of the danger are more likely to survive (and reproduce) than organisms that have to process the many stimuli they focus on, and sift through masses of information in their short-term memory. Indeed, Simon [29] argued that selective attention is a key feature of the broader notion of bounded rationality – the assumption that humans make decisions in science, business and everyday life with only a small amount of search. Thus, selective attention enables rapid decisions that may not be optimal, but that are good enough.

While CHREST is a computational architecture of cognition that captures many of the aspects of attention as studied in cognitive psychology, it was not developed as a model of attention only. As we have seen earlier, it also captures many aspects of (high-level) perception, learning, memory and decision-making. CHREST may also have interesting things to say about machine learning and robot perception. For example, chunking turns out to be a robust statistical learning mechanism, and we have provided several examples in this paper of how chunking is linked to attention.

The way attention is used in CHREST can also shed light on questions central to the development of autonomous intelligent systems. The frame problem is such a question. How can a system notice the relevant changes in the environment, while ignoring the irrelevant changes? Due to the various limitations that characterise CHREST, it actually does not face the frame problem. Together, CHREST’s attentional restrictions implemented by a limited capacity memory enable it to pay attention just to a few features of the environment. With increasing expertise in a domain, the perception-learning-perception cycle we have discussed above leads to increasingly rapid and adequate decisions, whilst not necessarily increasing the attention span of the system.

Evolutionary considerations highlight two factors important in modulating attention, which we have started incorporating into CHREST: emotions and motivations. In a classic paper, Simon [30] argued that, with systems characterised by serial organisation and control hierarchy, motivation refers to what is controlling attention at a specific time; in particular, given that these systems have multiple goals, motivation controls how attention is focused on a specific goal. Furthermore, it is necessary to have a provision for interrupt mechanisms; Simon proposes that at least two sources inform these mechanisms: first, drives (for example, hunger), and, second, the information gained by EPAM’s process of noticing. In particular, emotional tags might be added to perceptual chunks during the learning process. Later on, the emotions associated with these chunks, in particular when they are negative, such as fear, may direct attention to specific aspects of the environment. For example, a chess master might have suffered

a painful loss in a given type of position. In future games with the same type of position, her attention is likely to be modulated by the knowledge of this previous game, and the emotions associated with it – for better or for worse.

7 Conclusion

In this paper, we have described how the CHREST cognitive architecture explains, simulates and employs attention mechanisms. Two aspects of CHREST's implementation are central to explaining the attention mechanism in humans: the first is the tight cycle of perception-learning-perception; the second is the use of discrete *chunks* of information, both in long-term and short-term memory. The perception-learning-perception cycle guides the attention mechanism through a set of heuristics, which select a new fixation point by combining bottom-up and top-down information. The discrete nature of chunks enables a limited short-term memory to refer to a far larger store of information, and also supports the use of cross-modal or image-action associations. As CHREST's problem-solving and action abilities are further extended, we anticipate greater insights will emerge about the link between an expert's selective attention and their ability to produce rapid and skillful responses.

Acknowledgements

We thank the reviewers for their comments on a previous version of this paper.

References

1. de Groot, A.D.: Thought and Choice in Chess. The Hague: Mouton (1946/1978)
2. Simon, H.A., Gilmarin, K.J.: A simulation of memory for chess positions. *Cognitive Psychology* **5** (1973) 29–46
3. Gobet, F., Lane, P.C.R., Croker, S.J., Cheng, P.C.H., Jones, G., Oliver, I., Pine, J.M.: Chunking mechanisms in human learning. *Trends in Cognitive Sciences* **5** (2001) 236–243
4. Feigenbaum, E.A., Simon, H.A.: EPAM-like models of recognition and learning. *Cognitive Science* **8** (1984) 305–336
5. de Groot, A.D., Gobet, F.: Perception and Memory in Chess: Heuristics of the Professional Eye. Assen: Van Gorcum (1996)
6. Gobet, F., Simon, H.A.: Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology* **31** (1996) 1–40
7. Gobet, F.: Expert memory: A comparison of four theories. *Cognition* **66** (1998) 115–52
8. Gobet, F., Simon, H.A.: Five seconds or sixty? Presentation time in expert memory. *Cognitive Science* **24** (2000) 651–82
9. Freudenthal, D., Pine, J.M., Gobet, F.: Modelling the development of children's use of optional infinitives in English and Dutch using MOSAIC. *Cognitive Science* **30** (2006) 277–310

10. Minton, S.: Quantitative results concerning the utility of explanation-based learning. *Artificial Intelligence* **42** (1990) 363–91
11. Lane, P.C.R., Cheng, P.C.H., Gobet, F.: Learning perceptual schemas to avoid the utility problem. In Bramer, M., Macintosh, A., Coenen, F., eds.: *Research and Development in Intelligent Systems XVI: Proceedings of ES99, the Nineteenth SGES International Conference on Knowledge-Based Systems and Applied Artificial Intelligence*, Cambridge, UK: Springer-Verlag (1999) 72–82
12. Gobet, F.: Discrimination nets, production systems and semantic networks: Elements of a unified framework. In: *Proceedings of the Second International Conference of the Learning Sciences*, Evanston, IL: Northwestern University (1996) 398–403
13. Burl, M.C., Asker, L., Smyth, P., Fayyad, U., Perona, P., Crumpler, L., Aubele, J.: Learning to recognize volcanoes on Venus. *Machine Learning* **30** (1998) 165–94
14. Talavera, L.: Feature selection as a preprocessing step for hierarchical clustering. In: *Proceedings of the International Conference on Machine Learning*. (1999)
15. Chase, W.G., Simon, H.A.: Perception in chess. *Cognitive Psychology* **4** (1973) 55–81
16. Gobet, F., Clarkson, G.: Chunks in expert memory: Evidence for the magical number four... or is it two? *Memory* **12** (2004) 732–47
17. Gobet, F.: Is experts' knowledge modular? In: *Proceedings of the Twenty-Third Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum (2001) 336–341
18. Campitelli, G., Gobet, F.: Adaptive expert decision making: Skilled chessplayers search more and deeper. *Journal of the International Computer Games Association* (2004)
19. Gobet, F., Jansen, P.: Towards a chess program based on a model of human memory. In van den Herik, H.J., Herschberg, I.S., Uiterwijk, J.W., eds.: *Advances in Computer Chess 7*. Maastricht: University of Limbourg Press (1994) 35–60
20. Gobet, F., de Voogt, A., Retschitzki, J.: *Moves in Mind: The Psychology of Board Games*. Hove and New York: Psychology Press (2004)
21. Neisser, U.: *Cognitive Psychology*. New York: Appleton-Century-Crofts (1966)
22. Biederman, I.: On the semantics of a glance at a scene. In Kubovy, M., Pomerantz, J.R., eds.: *Perceptual Organization*. Hillsdale, NJ: Lawrence Erlbaum (1981) 213–254
23. Lindsay, P., Norman, D.: *Human Information Processing*. New York: Academic Press (1972)
24. Lane, P.C.R., Sykes, A.K., Gobet, F.: Combining low-level perception with expectations in CHREST. In Schmalhofer, F., Young, R.M., Katz, G., eds.: *Proceedings of EuroCogsci*, Mahwah, NJ: Lawrence Erlbaum Associates (2003) 205–210
25. Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y., Davis, N., Nuflo, F.: Modeling visual attention via selective tuning. *Artificial Intelligence* **78** (1995) 507–545
26. Wolfe, J.M.: Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review* **1** (1994) 202–238
27. Newell, A.: *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press (1990)
28. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebière, C., Qin, Y.L.: An integrated theory of the mind. *Psychological Review* **111**(4) (2004) 1036–1060
29. Simon, H.A.: A behavioral model of rational choice. *The Quarterly Journal of Economics* **69** (1955) 99–118
30. Simon, H.A.: Motivational and emotional controls of cognition. *Psychological Review* **74** (1967) 29–39