Expertise Effects in Memory Recall:
a Reply to Vicente and Wang

Simon, H. A. & Gobet, F. (2000). Expertise effects in memory recall: A reply to Vicente and Wang. *Psychological Review, 107,* 593-600.

Herbert A. Simon
Department of Psychology
Carnegie Mellon University

Fernand Gobet
ESRC Centre for Research in Development
Instruction and Training
Department of Psychology
University of Nottingham

Correspondence to:
Prof. Herbert A. Simon
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

has@cs.cmu.edu
412-268-2787

Running head: Expert memory

Abstract

In the January 1998 *Psychological Review,* Vicente and Wang propose a "constraint attunement hypothesis" to explain the large effects of domain expertise upon memory recall observed in a number of task domains. They claim to find serious defects in alternative explanations of these effects which their theory overcomes. Re-examination of the evidence shows that their theory is not novel, but has been anticipated by those they criticize, and that other current published theories of the phenomena do not have the defects Vicente and Wang attribute to them. Vicente and Wang's views reflect underlying differences (a) about emphasis upon performance versus process in psychology, and (b) about how theories and empirical knowledge interact and progress with the development of a science.

Expertise Effects in Memory Recall: a Reply to Vicente and Wang

Herbert A. Simon
Carnegie Mellon University

Fernand Gobet
University of Nottingham

In the January 1998 *Psychological Review*, Vicente and Wang (henceforth, V-W) employed an ecological approach to explain the "significant correlation between domain expertise and memory recall performance after a very brief exposure time" (p. 33). They claim that "this *constraint attunement hypothesis* [henceforth, CAH] . . . predicts . . . a memory expertise advantage in cases in which experts are attuned to the goal-relevant constraints in the material to be recalled and that the more constraint available, the greater the expertise advantage can be" (p. 33), and explains "several findings in the literature [that] have no satisfactory theoretical explanation." We will claim that, on the contrary, these findings have already been explained by process theories, and the memory advantage has long been understood in terms of its goal relevance.

After summarizing V-W's theoretical proposal, we address several errors in their account of past and current research on expertise. We then show why the ten experiments V-W review do not support CAH unless one adds to it a number of auxiliary assumptions, and why they provide better support for the contemporary process theories criticized by V-W. Process theories need not wait, and have not waited for product theories such as CAH to clear the way for progress.

0. Introduction to CAH

The authors define CAH very briefly:

> There can be expertise effects [in memory recall] when there are goal-relevant constraints (i.e., relationships pertinent to the domain) that experts can exploit to structure the stimuli. The more

> constraint available, the greater the expertise advantage can be. Fully random stimuli have no constraints, so no expertise advantage would be expected. To realize these potential advantages, experts must be attuned (i.e., they must attend) to the goal-relevant constraints in question. [If not,] then no expertise advantage is expected. (V-W, p. 36)

Thus, CAH asserts that goal-relevant patterns in the task situation, and only such patterns, must be attended to. Nowhere does CAH specify when, within the limitations of learning rates and memory capacities, the potential advantage of experts will or won't actually be realized.

The authors disclaim novelty for CAH, which they describe as deriving from E. J. Gibson's (1969) specificity theory and Rasmussen's (1985) notion of abstraction hierarchy; but they claim that its *application* to explain expertise effects in memory recall is novel.

In terms of goal-relevant constraints, V-W distinguish between "domains in which memorizing stimuli is an *intrinsic task* (i.e., . . . a definitive feature of that domain . . .) . . . [and domains] in which [it] is a *contrived* task (i.e., . . . not part of that domain of expertise)" (p. 34). V-W classify reproduction of briefly viewed chess boards as a contrived task, unlike most other tasks where superiority of expert memory has been demonstrated, quoting Gobet's statement (1994, p. 45) that "Chess players' main occupation is not to perform recall experiments but to play chess." They then claim (pp. 34-35) that "skilled-memory theory, LTWM theory, and EPAM IV theory cannot provide an adequate theoretical explanation for expertise effects in memory recall for . . . domains in which memory recall is a contrived task," and generalize this to claim that "an adequate process theory of expertise effects in memory recall has remained elusive." We will refute this claim for chess, the principal "contrived task" they examine, and show the problems that CAH faces in dealing with the other tasks they discuss.

V-W seek to use well-known experiments to show the superiority of their informal and non-quantitative CAH over other proposed explanations of the phenomena: the 1946-1973 "chunking" theories of de Groot, Jongman, Barenfeld, Chase, Gilmartin and Simon (CHNK); the skilled-memory theory (Chase

& Ericsson, 1981, 1982; Ericsson & Staszewski, 1989); the long-term working memory (LTWM) theory of Ericsson and Kintsch (1995); and the EPAM-based template theory (EPAM-TEMP) of Gobet, Richman, Simon and Staszewski. Skilled-memory theory and LTWM, like CAH, are (mainly) non-quantitative; EPAM-TEMP, and to a lesser extent CHNK, are embodied in computer simulation models that make quantitative predictions. We will deal first with V-W's specific claims and criticisms, and then with broader methodological issues.

1. Data and Theories of Chess Memory

Theories and data in science generally undergo progressive growth and refinement, unless the theories are discarded in favor of better theories, or the data rejected because of evidence of inaccuracy. Treating each theory as a separate entity, V-W misread the cumulative development of knowledge of chess memory as a series of failures, reversed only by the appearance of CAH. A brief review of the actual history reveals the errors in their critique of the theories proposed prior to CAH.[1]

1.1. The Work of de Groot and its Ramifications

De Groot, in his landmark work (in Dutch, 1946), reported that chessmasters could recall a chessboard almost perfectly after seeing a game position for a few seconds, while weaker players could replace only about half the pieces correctly. Two decades later it was shown[2] that most of the expert's advantage vanished if the players viewed the same pieces arranged randomly[3] (Jongman, 1968, p. 57, fn. 1; Chase & Simon, 1973a, 1973b).

---

[1]We will trace only one strand of the history, and that incompletely. Many important contributors to our knowledge of chess memory, for example, Holding and Saariluoma, are almost absent from our account, as is research on chess play, as distinct from memory.

[2]Unpublished study by Lemmens and Jongman, footnoted in Jongman's thesis.

[3]V-W note that these positions are not "truly" random, for the sets of pieces are taken from game positions, and they propose an alternative definition. In the chess literature, "random" explicitly and correctly means "pieces from game positions rearranged randomly." We follow this standard usage and discuss the issue of

These findings have been replicated frequently (and analogous findings reported for many other tasks): that (1) experts remember much more about chess positions, on brief exposure, than novices; but (2) their advantage decreases as the positions deviate from positions typically encountered in games, and (nearly) disappears when the pieces from game positions are re-arranged randomly. The "nearly" was only established reliably about 1996 (Gobet & Simon, 1996b) by a new experiment and meta-analysis. Until then, only two studies (Gold & Opwis, 1992; Saariluoma, 1994), showed statistically significant differences between experts and novices in the random condition.

These findings concern not just what expert chessplayers are *attuned to*, which is CAH's sole concern as a product theory, but what they *learn, know* and can *recall* after exposure that is too brief for fixation in LTM. Therefore they have important implications for the structure of LTM and STM, and for the processing times required to store new information. CAH has no basis for making differential predictions for long and short stimulus presentations; therefore it cannot predict masters' difficulty in remembering random boards after short exposure without process assumptions (not present in CAH) about acquisition, storage processes and size limits on storage and retrieval systems. De Groot explained the findings by the familiarity of chessmasters with patterns that appear frequently in games. Unlike weaker players, masters did not have to retain individual pieces in STM, but only *labels* for about a half dozen meaningful patterns. Jongman (1968, p. 48) compared these patterns to the "perceptual chunks" of Miller (1956). Chunking was also used by Feigenbaum (1961) in the EPAM theory of LTM and STM.

The chunking theory was extended by Simon and Barenfeld (1969), by simulating, with PERCEIVER, how knowledge of chess-relevant relations among pieces can guide chessplayers' eye movements to recognize meaningful LTM chunks. The chess-relevance of the relations guiding the eye movements and chunking demonstrates that the chunking theory, contrary to V-W's assertion (p. 35), took account of both the meaning of materials and their familiarity (Feigenbaum & Simon, 1984; Simon &

---

residual pattern, quite comparable to the "non-randomness" of cards dealt in a particular hand. Lemmens and Jongman had already examined four degrees of randomization, V-W's is a fifth.

Feigenbaum, 1964). Meaningfulness influenced attention, hence learning; familiarity permitted subsequent recognition (Simon & Barenfeld, 1969, p. 481).

But de Groot, in 1946, had already recognized the relevance of chunks (his "complexes") for understanding chess positions: "The position is perceived in *large complexes*, each of which hangs together. . . . [and] . . . is to be considered as a unit of perception *and* significance. (de Groot, [1946] 1978, pp. 329-30). Here we see, a half century prior to V-W's paper, a key aspect of the CAH theory clearly stated and used, explaining chunks as acquired for one purpose, but used for another. CAH only predicts a *possible* connection between task-relevant knowledge and task performance, while de Groot provides also learning and recognition *processes* to account for performance in two distinct tasks. The chunks were learned and stored as part of the "intrinsic" task of playing skillfully, but, once stored, prove useful in the "contrived" task of recalling positions.

Today, CHREST (De Groot and Gobet, 1996; Gobet, 1993; Gobet & Simon, in press), a template theory used to simulate chess memory experiments, acquires chunks by scanning a database of games. The CHUMP model (Gobet & Jansen, 1994), using the same input, plays chess by growing a discrimination net for moves and learning to associate these moves with the patterns acquired by its CHREST component.[4] Thus, the model embodies the "theory of expertise in domain-relevant problems, not just expertise effects in memory recall," that V-W (p. 48) call for.

1.2. The C-S Experiments and Theory

In the 1970s, Chase and Simon (1973a, 1973b) began to explore the CHNK hypothesis more directly:

---

[4]CHUMP plays weak chess because no search mechanisms are implemented, although they are part of the chunking and template theories (Chase & Simon, 1973b, pp. 269-270, Gobet, 1997, Gobet & Simon, 1998b).

> Previous studies . . . make highly plausible . . . that the chess master encodes information . . . in chunks but *provide no direct methods for delimiting the chunk boundaries or detecting the relations . . . among the components of a chunk* [italics added]. Evidence is needed . . . to discover how many pieces typically constitute a chunk, . . . the relative sizes . . . of the chunks of masters and weaker players, and how many chunks players retain after a brief view of a position. . . . The central objective of this study, then, is to isolate and define the chunks . . . in chess perception tasks (1973a, pp. 56-57).

C-S then use this definition to segment the chunks in the board recall task. Thus, the main goal of the C-S experiment was not, contrary to V-W, to "test" the chunking hypothesis, but to provide new observational means for characterizing chunks. While adding new evidence for the large master-novice difference in the recall of game positions, the experiment did not detect a difference in random positions. It did detect (C-S, 1973a, p. 70 and Tables 3 and 7; 1973b, p. 232) that the chunks in random positions, as in game positions, had more than an average number of chess relations among their pieces, which should, according to theory, produce increase in recall with skill — but didn't in this experiment.[5]

C-S designed two converging definitions of chunks: latencies between placement of successive pieces, and number of chess-relevant relations between such pieces. (See C-S, 1973a, Figure 5 and associated text).[6] With these measures, they could now ask whether experts' superiority derives from recalling larger chunks or more chunks. In fact, their chessmaster recalled "both more chunks and larger chunks" than the other players, while the theory predicted no differences in number. In 1973a and 1973b, C-S suggested possible explanations for the observed variation in number of chunks, but left the question for further study.

---

[5]In another experiment, C-S detected large differences between stronger and weaker players in recalling random sequences of legal moves, a finding we will discuss later.

[6]The convergence of the two methods has recently been replicated and extended by Gobet and Simon (1998a).

C-S mention another failure of the original model to explain *all* the empirical findings: recall again showed great superiority of stronger players on game positions, but no superiority on random positions. However, C-S also found that, "in the randomized boards, players are noticing the same kinds of structures as those they perceive in the coherent positions, even though these structures occur rarely in the randomized boards" (1973b, p. 232).

Hence, contrary to V-W's claim, the *theory* predicted that the strongest player should have shown some small superiority even on randomized boards, but the *data* did not support the prediction, nor, of course, the identical prediction that CAH makes.

At this same time, the processes of the CHNK theories were simulated by MAPP, a simplified version of EPAM (Simon & Gilmartin, 1973). The chunks learned by MAPP produced a level of recall consistent with theoretical estimates of the required number of chess patterns (and also consistent with Gobet & Simon, 1996a). The simulation found "substantial correlations between chess master and MAPP as to which pieces . . . are remembered, . . . the organization of these pieces in chunks, and . . . the sequence in which pieces are replaced on the board" (p. 46).

None of these findings are predicted by CAH. Both experimental findings and the results of simulation supported the CHNK theory, as implemented in PERCEIVER and MAPP,

> to account for the *main* [italics added] features of the human performances . . . by bringing together . . . basic mechanisms, for chunking in short-term and long-term memory and for managing attention,[7] whose significance had already been demonstrated . . . in experimental paradigms and tasks quite remote from the present one (Simon and Gilmartin, 1973, p. 46).

Although explicitly recognizing that the theory did not explain the observed lack of expert superiority on random boards or the increase in number of chunks with expertise, C-S took the usual and

---

[7]Notice the role of the attention mechanism in EPAM, again anticipating CAH.

justified position that failure to account for *all* known phenomena is not grounds for rejecting a theory's core, but an invitation to explain the deviation (Lakatos, 1970).

1.3. Charness (1976)

Charness (1976) found an even more serious gap in CHNK as a *complete* theory. Intervening tasks that use STM should interfere with ability to reproduce positions after the interruption (as in the Peterson and Peterson, 1959, paradigm), but they did not. (CAH would not, of course, account for the effects of intervening tasks.)

Where were the chunks or the pointers held while the intervening task was being performed? Simon (1976), citing independent evidence from verbal learning, proposed that a chunk could be inserted in an existing LTM node in 1 s or 2 s. This new hypothesis would account for Charness's results.

Charness's anomaly led to experiments with two mnemonicists who learned (after two years of almost daily training) to memorize, in one trial, lists of 80 or more digits presented at one digit per second — far too rapidly, according to the original EPAM theory, to transfer them to LTM. Chase and Ericsson (1981) and Staszewski (1988) showed that the participants gradually learned new LTM schemas that they called "retrieval structures": deliberately acquired discrimination nets containing "slots" for variables that could be filled within 1 s or less each. The evidence for their existence and use was extensive, permitting quantitative prediction of learning rates over the entire learning period (Richman, Staszewski & Simon, 1995).

1.4. Recent Developments of Chunking Theory

Subsequently, CHNK has evolved on two complementary lines. First, the skilled-memory theory (Chase & Ericsson, 1982; Ericsson & Staszewski, 1989) and LTWM (Ericsson & Kintsch, 1995) use retrieval structures to explain expert recall in various domains. Second, the direct descendant of CHNK, EPAM-TEMP ("EPAM with templates"), comprises both EPAM-IV (Richman, Staszewski & Simon, 1995) and CHREST, the latter with schemas specialized for chess.[8] These simulations predict both outcomes, and processes. In its account of the sources of the expert's knowledge, and its use of the attention mechanism, EPAM-TEMP again anticipates the much weaker CAH product theory.

In chess, most templates and other patterns in LTM are acquired while studying games and books in order that common patterns will be recognized, giving access to information relevant for choosing strategies and moves. For skilled chess play, a store of recognizable patterns and associated chess information is indispensable. It is also invaluable, if inadvertently, for performing the recall task. The EPAM-TEMP theory (Gobet & Simon, 1996a, 1996b, 1996c, 1998a, in press), incorporating this feature, experiences none of the difficulties that V-W attribute to process theories of recall when performing "contrived" tasks. Information, learned for whatever purpose, can be evoked and used for a new task.

1.5. EPAM-TEMP

EPAM learns both to discriminate and recognize familiar stimuli. The information at the various terminal nodes (semantic memory) is connected by associations (see Richman, Simon & Staszewski, 1995). EPAM-TEMP retains these same mechanisms, adding retrieval structures and templates, learned in the ordinary way, whose form may depend on the task to be addressed.

For example, in recalling digit strings, the retrieval structures were abstract hierarchies (Richman, Simon & Staszewski, 1995), explicitly memorized and associated with pre-existing semantic schemas through the standard EPAM learning mechanisms, with unchanged learning parameters. Their novelty was ability to insert digits in an already-learned hierarchy about as rapidly as it could insert them in STM.

---

[8]In both skilled-memory theory and EPAM-TEMP, "retrieval structure" refers to an LTM structure deliberately acquired. The other memory structures, which also have variable slots, are called "templates." Templates need not be deliberately acquired and fully available to consciousness, but are *available* to recall when the task calls attention to them. Occasionally, when mode of acquisition and awareness are not at issue, we use “retrieval structure” in a generic sense, including both “template” and “retrieval structure” in the narrower meaning defined here.

CHREST incorporates templates with slots for squares, which can accept information about the piece on each square or a familiar chunk of pieces located around it. The largest templates (often containing a dozen or more pieces) correspond mainly to typical opening positions. These templates represent no essential novelty. The rapid LTM encoding they allow explains why interference has little effect in Charness' (1976) and Frey and Adesman's (1976) experiments, but do not otherwise much alter the earlier chunking theories. Templates are created automatically during learning, reflecting the patterns in the environment (Gobet & Simon, 1998a, in press), thus achieving one of V-W's goals for future research.

CHREST, using Web-stored data bases of chess games, has grown a discrimination net of 300,000 nodes. Performance in the recall task improves steadily with size of net, and approaches master level performance: 85% versus 90% (Gobet, 1998; Gobet & Simon, in press). This removes V-W's objection (p. 34) that CHREST did not simulate master-level recall performance. (Simulating any performance is beyond the power of CAH.)

Hence, far from refuting the CHNK theory or the EPAM theory, the experimental results of Charness and Chase-Ericsson-Staszewski showed how to amend the theory by incorporating variable slots in LTM schemas. The contemporary versions of the chunking hypothesis embodied in EPAM and CHREST account for all of the phenomena that caused trouble for C-S and their successors with the earlier CHNK simulations. Wherever CAH makes predictions at all, the new chunking models make good quantitative predictions, whereas CAH predicts only differences of sign.

2. The evidence reviewed by V-W

We will next comment on the specific experiments that V-W cite as supporting CAH or refuting competing theories. Because these experiments are described by V-W, and the original papers are available in the literature, we can be very brief. They encompass both chess and other tasks.

2.1. Chase and Simon (1973b)

Participants reconstructed move sequences from master games: actual moves, or random but legal moves, dictated at 5 s a move. All participants did much better with actual than with random move sequences; and the stronger players much better than the weaker with actual moves. But contrary to the experimenters' reported expectations, performance with random legal moves also varied strongly with the strength of the player.

V-W regard these findings as refuting the chunking theory, but they only refute the incautious prediction of the experimenters, who did not use MAPP to check their guess (illustrating the frequent superiority in reasoning from precise models rather than words). The positions were set up by the participants and studied for 30 s before the moves were presented. Then, the players had 5 s to study each move on the slowly changing board: long enough to fixate at least half the moves, at the usual 8 s per chunk. Even in the random-move condition, numerous chess-relevant patterns were visible so that all chunking theories would predict expert superiority. CAH, which does not deal with processes, much less their durations, cannot predict the outcome at all without making additional ad hoc assumptions about the "constraints" that distinguish this from the ordinary board recall task.

2.2. Frey and Adesman (1976)

In this experiment an interposed task (a second chess position) interfered only minimally with recall of the board. This is predicted by the EPAM-TEMP version of the chunking theory.

Frey and Adesman also compared game positions, random placement, and move-by-move presentation of the preceding moves in the usual board replacement paradigm. For the first two conditions, as usual, stronger players showed a clear advantage in game but not random positions. The three groups, but especially the strongest players, did still better with move-by-move presentation.

V-W claim the improvement in the latter condition is predicted by CAH, because the presentation provides "redundant information about the tactics and strategies that led to the position to be recalled" (p. 40), hence an extra "constraint." EPAM-TEMP would obviously also show the expert advantage, for many positions are reached by common sequences of opening moves, and this information would be retrieved and used in recall. Moreover, in the move-by-move condition, S's had considerable time for fixation, as the position changed only one piece at a time. Hence, the chunking hypothesis makes the same prediction in this case as CAH, but can be quantified, as the latter cannot.[9]

2.3. Goldin (1978)

As V-W admit (p. 45), the results from Goldin's 1978 experiment "contradict CAH." The results do not conflict with chunking theories in either CHNK or EPAM-TEMP form, although assumptions would have to be made about costs of errors of omission and commission to predict outcome. (See section on Myles-Worsley et al., 1988, experiment, below.)

2.4. Lane and Robertson (1979)

In this study of recall of chessboards, in the "semantic task" condition, participants decided which player had the advantage and chose the best move; in the "formal" condition, participants counted the numbers of pieces on the black and the white squares. All participants also memorized another position ("intentional" learning condition).

Recall was better in the "semantic" than in the "formal" condition, and equally good in the "semantic" and "intentional learning" conditions. CAH predicts this, as does the chunking theory on the basis of difference in focus of attention.

2.5. Reynolds (1982)

This is a study of recall of positions constructed randomly, but under certain constraints ("order of grouping"). V-W note that "the only statistically significant effect of expertise occurs in the condition with

---

[9]In the EPAM-TEMP theory, redundancy plays two facilitating roles: it reduces the rate of forgetting in LTM, and it provides alternative retrieval paths for stored information. See Richman, Staszewski and Simon (1995).

the highest order grouping" (p. 41).[10] The same result would be predicted by any form of CHNK or EPAM.

2.6. Barfield (1986)

In this experiment, computer codes instead of chessboards are recalled: intact code, the same code with sub-blocs arranged randomly, or the same code with lines arranged randomly. The chunking theory in any of its forms predicts better recall at all levels of expertise, and increasing advantage of more expert players with more coherence. CAH makes the same prediction.

2.7. Coughlin and Patel (1987)

In this experiment, medical students and physicians recall sets of symptoms. The order of appearance of symptoms is relevant, or irrelevant. The symptoms were presented in that order, or randomly. In the random condition, physicians performed significantly better than students when order of symptoms was irrelevant but not when it was relevant. As relevance is a condition for attention, hence for learning in the chunking theory, all chunking theories predict this result. There is no need to heed or encode irrelevant information, hence its removal from the set of symptoms will not affect recall.

2.8. Myles-Worsley et al. (1988)

This experiment compared the ability of participants at various levels of expertise to distinguish abnormal (presumably pathological) X-rays from normal X-rays. With greater expertise, more abnormal X-rays were recognized as such, but more normal X-rays were also judged abnormal.

The chunking model would predict that experts would detect subtler conditions of possible abnormality than would less expert judges. Without specifying participants' beliefs about the costs of errors

[10] Figure 5 of V-W is incorrectly labeled. The skill levels used by Reynolds (1982) were Class C, Class A and Masters, not Class C, Class B and Class A.

of omission and commission, chunking cannot predict (nor can CAH) how experts' capabilities for recognizing symptoms would bias their decisions in borderline cases.

2.9. Vicente (1992)

Experts and novices were shown the behavior of a simulated thermal-hydraulic process plant and asked to recall the final state of the process variables and to diagnose whether the trial was "normal," "faulty" (one fault introduced into the operation), or "random" (the process variables driven in a physically meaningless fashion). CAH makes the correct prediction**[,]** of expert superiority in all three conditions, and of descending accuracy from "normal" through "faulty" to "random." Again, V-W need an ad hoc assumption about relevance to support the predictions.

The chunking model would learn patterns (chunks) of process variables in each of the conditions and expand its EPAM net to discriminate among the three conditions. If we equate randomness in chess boards with "abnormality" in a machine, then a chunking theory would predict that experts (a) would recognize degrees of departure from normality, and (b) would remember more of the properties (familiar chunks) of "normal" configurations.

2.10. Gobet and Simon (1994, 1996b, 1998a)

The absence of expert-novice differences in recalling random positions had long been questioned by the CHNK theorists, because their theories predicted at least small differences. A meta-analysis of 13 studies (Gobet & Simon, 1996b) showed a small but statistically significant superiority of experts over novices on random boards, confirming the suspicions roused by the C-S prediction and revived by V-W.

Gobet and Simon (1998a) also examined the question, first raised by C-S (1973a, 1973b), of whether the difference in <u>number</u> of chunks recalled by experts and novices was a reality or an artifact. In particular, "as the subjects used actual chess boards and pieces, the maximum number of pieces they could grasp in one hand could have limited apparent chunk sizes" (Gobet & Simon, 1998a, p. 226).

New experiments, using the mouse to replace pieces on a computer screen, show that chunk sizes had been substantially underestimated, so that size, rather than number, is the principal basis for the superior recall of stronger players. The new data replicated the original findings of negative correlation between placement times and numbers of chess relations between successively placed pieces, and the superiority of strong over weak players in recall.[11] They also gave a good fit to the visual STM capacity parameter originally estimated from printed characters (Zhang & Simon, 1985).

2.11. A Proposed Experiment

V-W (p. 45) propose an experiment that would lead to different predictions by CAH from those of EPAM-TEMP theory. "Truly" random positions ("$random_2$") would be constructed by randomizing both the pieces selected and their location, not just the location, as in the standard ("$random_1$") procedure. V-W state that CAH predicts no difference in recall in $random_2$ across skill levels, while EPAM-TEMP (or any chunking theory) would predict a skill effect, for stronger players will more often than weaker players recognize the patterns produced adventitiously by the random process (just as a random bridge deal occasionally produces 13 spades in one hand).

Data recently collected by Gobet and Waters (in preparation) including 36 participants, from amateurs to grandmasters, show that stronger players recall $random_2$ positions better than weaker players. V-W might now argue that $random_2$ sets are really not random, because there are still constraints (e.g., a square can only be occupied by one piece). However, this defense underlines the ad hoc way in which CAH defines and can redefine the "constraints of a domain." In any event, the predictions of the chunking theories, but not of CAH, are supported by the observed data.

---

[11] There was a clear difference, by expertise, in the random positions, as both the chunking theory and, more recently, CAH predict, thus removing one experimental (not theoretical) anomaly that C-S had noted 25 years earlier.

2.12. Empirical Evidence: Conclusions

At the heart of CAH is the idea that in acquiring expertise, attention is paid to information relevant to expert performance, and only such information. This idea was announced by de Groot in 1946 and elaborated by the subsequent research. It remains embedded in the chunking models in their current LTWM and EPAM-TEMP forms. Process models do not ignore the goal-oriented and adaptive nature of behavior, for they fully incorporate adaptive mechanisms (e.g., goals, means-ends analysis). In addition, contrary to V-W's claim (p. 36), (itself contradicted by the citation of De Groot (1966) in their appendix [p. 56]), proponents of chunking processes have made detailed analyses of the constraints offered by the statistical properties of the chess environment (Chase & Simon, 1973a, b, De Groot, 1966; De Groot & Gobet, 1996; Jongman, 1968; Simon & Gilmartin, 1973), thereby adding to the quantitative strength of the theory.

The experiments cited by V-W show no superiority of CAH over CHNK and EPAM-TEMP. The chunking theories make many predictions about process as well as product, hence explain, as CAH cannot, how the outcomes come about. Moreover, CAH makes only qualitative predictions; and even for these, constraints must be constructed and rationalized ad hoc for each task.

In contrast, the chunking theory makes many quantitative predictions, using parameters that are constant over all the tasks. For example, in all EPAM simulations, 8 s are required to store a single chunk in LTM unless slotted templates are accessed, but these require about 1/4 s each. Converging evidence has produced information about the sizes and nature of chunks and templates. In CHREST, adapted specifically to chess, normal human study materials (e.g., tournament games) are used for learning the discrimination net with its chunks and templates.

3. Phenomena and Theories

V-W's approach, focused on testing theories by generating observations that may disconfirm them, takes an unproductive view of the relation between phenomena and theories, and fails to explain where theories come from. In the actual practice of science (Gooding, 1990; Hanson, 1961; Simon, 1977; Thaggard, 1988), a major motivation for experimentation is to discover and characterize new phenomena. Oersted discovers that an electric current creates a magnetic field, or Faraday that movement of a magnet induces an electric current. Experiments explore; they do not just test hypotheses; they often initiate theory rather than following it.

The whole sequence of experiments on expert recall of chess positions is most fruitfully viewed as a series of problem-solving explorations that became increasingly interesting as new phenomena were revealed and ties disclosed to other important phenomena (e.g., verbal learning, Feigenbaum & Simon, 1984; categorization, Gobet et al., 1997; architecture and capacities of STM and LTM, Gobet & Simon, 1996; attention, Richman & Simon, 1989; the nature of expertise, Ericsson & Staszewski, 1989).

Theories are more useful the stronger the constraints they place on the phenomena, yielding sharper and more quantitative predictions. They are more useful if more general, provided that generality is not purchased at the cost of introducing free parameters and task-specific ad hoc conditions. They are more useful and meaningful if they not only describe and predict products, but also explain them in terms of their processes. In contrast to V-W's account, the line of research we have reviewed was not a series of isolated experiments, each pronounced a failure by significance tests, but a cumulative problem-solving search, in which early errors and insufficiencies led to new experiments, the discovery of new phenomena and improved theory.

3.1. Product Theories and the Ecological Stance

Vicente and Wang had a very sound message, though one already well known to evolutionists and ecologists: "To predict the shape of jello, look at the mold in which it is jelling." An adaptive system is just that: it adapts to its environment. More explicitly:

> A thinking human being is an adaptive system; . . . . To the extent that [humans] are effectively adaptive, possessing the relevant knowledge and skills, their behavior will reflect characteristics largely of the outer environment (in the light of their goals) and will

> reveal only a few limiting properties of their inner environments — of the physiological machinery that enables them to think (Simon, 1969, pp. 25-26).[12]

If we ignore a system's internal limits of adaptation, psychology becomes an impoverished science. There are, first, the limits of knowledge, and second, the limits of memory capacity and speed of storing new information. Dealing with these limits requires a theory of the organism, not just of the environment — that is, a process theory.

To make even the predictions it does about chess positions, CAH must assume: (1) that experts acquire information that facilitates description of a game position greatly, but description of a random position only slightly (a process assumption about learning); (2) that, because of STM limits, "raw" perception alone cannot reproduce the board without pre-stored information (a process assumption); (3) that intervening tasks do not always destroy the availability of this information (an assumption about memory architecture). CAH claims that experts cope better than novices — because they have inherited and/or acquired what it takes to do that. This claim doesn't tell us *what* it takes, nor under what conditions or how the "what" was acquired. These have to be supplied from outside the theory.

EPAM-TEMP theory replaces the required assumptions with integral components of the theory itself: that is, with mechanisms that satisfy these assumptions. EPAM-TEMP, exposed to standard chess data bases, learns chunks and templates, thereby acquiring the predicted (and testable) knowledge structures that produce recall very close to that of human chess players at corresponding levels of skill.

3.2. In Sum

The difference between CAH and theories like EPAM-TEMP is that the latter predict many phenomena that the former does not, and with fewer ad hoc assumptions, hence are stronger theories.[13]

---

[12]Revised to remove gender references.

[13]The reader should not be misled by V-W's incomplete quotation of Richman et al. (1995, p. 327): "the range of tasks studied in detail is very limited to date." What they said was, "But though the range of tasks

Further, recognizing that experimentation is exploration, the research sequence from de Groot in 1946 to today's experiments and models, illustrates how psychological knowledge advances by continuing mutual interaction of experiment with computer modeling and other theorizing.

This interaction between experiment and theory replaces the strategy, advocated by Popper (1959), of testing and refuting hypotheses with the much more productive strategy of gradually modifying and extending successful theories to improve their scope and accuracy. In the present case, the advance involves areas of cognitive psychology, extending far beyond expertise in recall of chess boards into the whole range of cognitive psychology, anticipating V-W's vision (p. 49) of "cumulative theory of skill acquisition that can account for learning in a wide variety of tasks" (Anderson, 1983; Feigenbaum & Simon, 1984; Newell, 1990).

V-W view CAH "as a stepping stone to developing viable process theories of expertise effects in memory recall. . . This role is valuable because it can make the search for a viable process theory more efficient than it has been" (p. 47). This is a hollow claim when a process theory that incorporated the characteristics of the task domain as integral components long preceded CAH. Product theories, incorporated in process theories to the benefit of both, are no new thing in this domain.

References

Anderson, J. R. (1983) *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Barfield, W. (1986). Expert-novice differences for software: Implications for problem solving and knowledge acquisition. *Behaviour and Information Technology, 5,* 15-29.

[i.e., memory for digit sequences] studied in detail [i.e., with this version of EPAM IV, including retrieval structures] is very limited to date, the fact that most of the mechanisms used (with the principal exception of the retrieval structures) have already served within EPAM to explain many perceptual, learning, and memory phenomena in other domains gives reasons for being sanguine about generalization." See also their following paragraphs. The past three years has produced substantial additional evidence for the retrieval structures in the quite different chessboard recall task, as the references cited in this paper show.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 641-653.

Chase, W. G., & Simon, H. A. (1973a). Perception in Chess. *Cognitive Psychology*, *4*, 55-81. Reprinted in H. A. Simon, *Models of Thought*. New Haven: Yale University Press, 1979.

Chase, W. G., & Simon, H. A. (1973b). The mind's eye in chess. In W. G. Chase (Ed.) *Visual information processing*. New York: Academic Press. Reprinted in H. A. Simon, *Models of Thought*. New Haven: Yale University Press, 1979.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Erlbaum.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16). New York: Academic Press.

Coughlin, L. D. & Patel, V. L. (1987). Processing of critical information by physicians and medical students. *Journal of Medical Education, 62,* 818-828.

de Groot, A. D. (1946). *Het denken van den schaker*. Amsterdam, Noord Hollandsche.

de Groot, A. D. (1966). Perception and Memory Versus Thought: Some Old Ideas and Recent Findings. In B. Kleinmuntz (Ed.), *Problem Solving, Research, Method and Theory*. New York: Krieger, 1966.

de Groot, A. D. (1978). *Thought and choice in chess*. The Hague: Mouton Publishers (English translation of de Groot, 1946).

de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess. Heuristics of the professional eye*. Assen: Van Gorcum.

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr. & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon.* Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.

Feigenbaum, E. A. (1961). The simulation of verbal learning behavior. *Proceedings of the 1961 Western Joint Computer Conference, 19,* 121-32.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8*, 305-336.

Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory & Cognition*, *4*, 541-547.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York, NY: Appleton-Century.

Gobet , F. (1993). A computer model of chess memory. In *Proceedings of 15th Annual Meeting of the Cognitive Science Society* (pp. 463-468). Hillsdale, NJ: Erlbaum.

Gobet, F. (1994). *Memory in chess players: Chunks, schemata, or both?* Complex Information Paper #517, Carnegie Mellon University, Pittsburgh, PA 15213.

Gobet, F. (1997). A pattern-recognition theory of search in expert problem solving. *Thinking and Reasoning, 3,* 291-313.

Gobet, F. (1998). Memory for the Meaningless: How Chunks Help. *Proceedings of the 20th Meeting of the Cognitive Science Society.* (pp. 398-403). Mahwah, NJ: Erlbaum.

Gobet, F. & Jansen, P. (1994). Towards a chess program based on a model of human memory. In H. J. van den Herik, I. S. Herschberg, & J. W. Uiterwijk (Eds.), *Advances in Computer Chess 7*. Maastricht: University of Limburg Press.

Gobet, F., Richman, H., Staszewski, J., & Simon, H. A. (1997). Goals, representations, and strategies in a concept attainment task: The EPAM model. *The Psychology of Learning and Motivation, 37,* 265-290.

Gobet, F. & Simon, H. A. (1994). *Role of presentation time in recall of game and random chess positions.* Complex Information Paper #524 Carnegie Mellon University, Pittsburgh, PA 15213.

Gobet, F., & Simon, H. A. (1996a). Recall of random and distorted positions: Implications for the theory of expertise. *Memory & Cognition, 24,* 493-503.

Gobet, F., & Simon, H. A. (1996b). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review, 3,* 159-163.

Gobet, F., & Simon, H. A. (1996c). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31,* 1-40.

Gobet, F., & Simon, H. A. (1998a). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6,* 225-255.

Gobet, F., & Simon, H. A. (1998b). Pattern recognition makes search possible: Comments on Holding (1992). *Psychological Research, 61,* 204-208.

Gobet, F., & Simon, H. A. (in press). Presentation time in expert memory. *Cognitive Science.*

Gobet, F. & Waters, A. (in preparation). *Constraints in chess memory.*

Gold, A., & Opwis, K. (1992). Methoden zur empirischen Analyse von Chunks beim Reproduzieren von Schachstellungen. *Sprache & Kognition, 11*, 1-13.

Goldin, S. E. (1978). Memory for the ordinary: Typicality effects in chess memory, *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 605-616.

Gooding, D. (1990) *Experiment and the making of meaning.* Dordrecht, NETH: Nijhoff/Kluwer.

Hanson, N. R. (1961). *Patterns of discovery*. Cambridge: The University Press.

Jongman, R. W. (1968). *Het oog van de meester*. Amsterdam: Van Gorcum.

Lakatos, I. (1970). Falsification and the methodology of scientific research program. In I. Lakatos, & A. Musgrave, (Eds.), *Criticism and the Growth of Knowledge*. Cambridge, England: Cambridge University Press.

Lane, D. M., & Robertson, L. (1979). The generality of the levels of processing hypothesis: An application to memory for chess positions. *Memory and Cognition*, *7*, 253-256.

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

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, Cognition, 14,* 553-557.

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

Peterson, L. & Peterson M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58* , 193-198.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-15,* 234-243.

Reynolds, R. I. (1982). Search heuristics of chess players of different calibers. *American Journal of Psychology*, *95*, 383-392.

Richman, H. B., & Simon, H.A. (1989) Context effects in letter perception: Comparison of two theories. *Psychological Review*, 1989, *96*, 417-432.

Richman, H. B., Staszewski, J. J., & Simon, H.A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review, 102*, 305-330.

Saariluoma, P. (1994). Location coding in chess. *The Quarterly Journal of Experimental Psychology, 47A,* 607-630.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Simon, H. A. (1976). The information-storage system called 'human memory'. In M. R. Rosenzweig, & E. L. Bennet (Eds.), *Neural Mechanisms of Learning and Memory*.

Cambridge, Mass: MIT Press. Reprinted in H. A. Simon, *Models of Thought*. New Haven: Yale University Press, 1979.

Simon, H. A. (1977). *Models of discovery.* Dordrecht, Holland: Reidel.

Simon, H. A., & Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. *Psychological Review*, *76*, 473-483. Reprinted in H. A. Simon, *Models of Thought*. New Haven: Yale University Press, 1979.

Simon, H. A. & Feigenbaum, E. A. (1964). An information processing theory of some effects of similarity, familiarity, and meaningfulness in verbal learning. *Journal of Verbal Learning and Verbal Behavior, 3*, 385-396. Reprinted in H. A. Simon, *Models of Thought,* New Haven: Yale University Press, 1979.

Simon, H. A., & Gilmartin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, 5, 29-46. Reprinted in H. A. Simon, *Models of Thought*. New Haven: Yale University Press, 1979.

Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The Nature of Expertise*. Hillsdale, NJ: Erlbaum.

Thagard, P. (1988) *Computational philosophy of science.* Princeton, NJ: Princeton University Press.

Vicente, K. J. (1992). Memory recall in a process control system: A measure of expertise and display effectiveness. *Memory & Cognition, 20,* 356-373.

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105*, 33-57.

Zhang, G. & H. A. Simon (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition*, *13*, 193-201.

Authors note

Correspondence concerning this article should be addressed to Herbert A. Simon, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213.