

The Role of Constraints in Expert Memory

Fernand Gobet
University of Nottingham

Andrew J. Waters
Georgetown University

A great deal of research has been devoted to developing process models of expert memory. However, K. J. Vicente and J. H. Wang (1998) proposed (a) that process theories do not provide an adequate account of expert recall in domains in which memory recall is a contrived task and (b) that a product theory, the constraint attunement hypothesis (CAH), has received a significant amount of empirical support. We compared 1 process theory (the template theory; TT; F. Gobet & H. A. Simon, 1996c) with the CAH in chess. Chess players ($N = 36$) differing widely in skill levels were required to recall briefly presented chess positions that were randomized in various ways. Consistent with TT, but inconsistent with the CAH, there was a significant skill effect in a condition in which both the location and distribution of the pieces were randomized. These and other results suggest that process models such as TT can provide a viable account of expert memory in chess.

Ever since the works of Piaget (1954), Brunswik (1956), and Simon (1969), environment has played an important role in psychological theories. However, although all psychologists agree that cognitive systems adapt to the structure of their environment, there are disagreements about the consequence of this on the optimal way to study cognition. On the one hand, it has been proposed that the best approach is to develop process theories, preferably in the form of computer programs, and compare their predictions with human data (e.g., Anderson & Bower, 1973; Newell, 1990; Newell & Simon, 1972). On the other hand, it has been argued that a careful study of the properties of the environment should come before any attempt to theorize about cognitive processes (Anderson, 1990; Brunswik, 1956; J. J. Gibson, 1966; Vicente & Wang, 1998). Once these properties are fully understood, developing process theories becomes an easier, more constrained task.

Recently, this debate has focused on the *expertise effect* in memory recall. Experts are vastly superior to nonexperts in memorizing briefly presented material taken from their domains of expertise. This effect, first reported about 50 years ago in chess (de Groot, 1946/1978), has been observed in many additional studies, in domains including games, programming, and medical expertise (see Ericsson, 1996; Vicente & Wang, 1998, for recent reviews). A number of process theories have been developed by authors who purport to explain the cognitive mechanisms underlying this effect,

including the chunking theory (Simon & Chase, 1973), the skilled memory theory (Chase & Ericsson, 1982; Ericsson & Staszewski, 1989), the long-term working-memory theory (Ericsson & Kintsch, 1995; Ericsson, Patel, & Kintsch, 2000), and the elementary perceiver and memorizer (EPAM-IV) and template theories (Gobet & Simon, 1996c, 1998, 2000; Richman, Staszewski, & Simon, 1995; Simon & Gobet, 2000). In contrast, Vicente and Wang (1998) proposed an ecological theory of expert memory—the constraint attunement hypothesis (CAH)—that focuses more on understanding the goal-relevant constraints in a domain. The CAH is also a *product theory* of expert memory in that the functional relation between input and output is described, but—like all product theories—it is silent about specific psychological mechanisms.

In their review, Vicente and Wang (1998) made a distinction between domains in which memorizing stimuli is an *intrinsic task* (a task that is a definitive feature of that domain of expertise) and domains in which memorizing stimuli is a *contrived task* (a task that is not part of that domain of expertise). They argued that the process theories they discussed “cannot provide an adequate theoretical explanation for expertise effects in memory recall for the considerable number of domains in which memory recall is a contrived task” (pp. 34–35). In contrast, they noted that “all but one of the results reviewed in this section were consistent with the predictions generated by the constraint attunement hypothesis, thereby providing a significant amount of empirical support for the theory” and that “[many of the results that were reviewed] still could not be accounted for, a posteriori, by other competing theories” (p. 46). Vicente and Wang’s analysis is further discussed by Ericsson et al. (2000), Simon and Gobet (2000), and Vicente (2000).

Here, we address this challenge in part by making a detailed comparison between one process theory (the template theory; TT) and the CAH in one domain, chess. (See Simon & Gobet, 2000, for a comparison across a range of domains.) We chose TT because it is a well-articulated process theory of expertise and because it is discussed in detail by Vicente and Wang (1998). We focused on

Fernand Gobet, School of Psychology, University of Nottingham, Nottingham, England; Andrew J. Waters, Lombardi Cancer Center, Georgetown University.

Andrew J. Waters is now at the Department of Behavioral Science, MD Anderson Cancer Center, University of Texas.

This article was supported by an Economic and Social Research Council grant awarded to Fernand Gobet. We thank Peter Lane, Julian Pine, and Herbert Simon for valuable comments on drafts of this article, as well as Ken Liu for statistical advice and Susan Marx for editorial assistance.

Correspondence concerning this article should be addressed to Fernand Gobet, who is now at the Department of Human Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, England. E-mail: Fernand.Gobet@brunel.ac.uk

chess because Vicente and Wang's review contained 6 chess experiments (of 10) and because—as we discuss later in this article—they explicitly proposed an experiment in which the predictions of the CAH and TT differ. Thus, a major goal of the article was to compare the predictions of TT and the CAH under experimental conditions in which they make different predictions.

We start by describing the CAH and TT. We then argue that the theories can be disentangled by examining recall for different types of randomized chess positions. Predictions of TT are derived from a simulation study. Predictions of the CAH are derived from Vicente and Wang's (1998) article. Finally, we report a human study that tested these predictions.

The CAH

The CAH is derived from E. J. Gibson's (1969) specificity theory and Rasmussen's (1985) notion of abstraction hierarchy. It represents an outgrowth of ecological theories of cognition, which stress the importance of the environment over internal processes. Vicente and Wang (1998) stated the CAH as follows:

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 they do not pick up on this information, then no expertise advantage is expected. (p. 36)

To define the goal-relevant constraints, Vicente and Wang (1998) proposed that for each particular domain, an abstraction hierarchy should be constructed (p. 36). In the abstraction hierarchy, there are constraints on relationships both within and between levels of the hierarchy. For example, one important feature of an abstraction hierarchy is that levels of the hierarchy are connected by means-end relationships (p. 37). Vicente and Wang claimed that these types of constraint are the goal-relevant constraints that experts can exploit when recalling material within their domain.

Vicente and Wang (1998, pp. 55–57) sketched an abstraction hierarchy for chess. This hierarchy has five levels: board, paths, tactics, strategies, and purpose. The lowest level, board, consists of the number and type of pieces as well as the physical properties of the board. For example, the number of pieces imposes constraints on the possible physical configurations (e.g., not more than 16 pieces of the same color). The second level, paths, consists of the constraints resulting from the rules governing the movement of pieces. The next level, tactics, “is constrained by the meaningful and effective ways in which moves can implement goal-relevant strategies” (p. 56). About the next level, strategies, Vicente and Wang noted that “masters do not constrain their moves by the rules of the game or tactics alone. Rather, their actions are also highly constrained by the higher order strategic plans that they have adopted to achieve a win (cf. Holding, 1985)” (p. 56). Finally, the highest level, purpose, constrains the strategies that can be meaningfully chosen.

The CAH explains expertise effects for standard chess positions in the following manner. When an expert looks at a position, he or she may attune to a strategic configuration (e.g., white attacking

black with a wing attack), thereby ruling out other strategic factors. The strategic configuration constrains tactical features of the position (e.g., white may have stockpiled his or her pieces on a given file), which in turn will constrain the path level (certain moves will be expected). The expert, but not the novice, has the capacity to attune to these multiple constraints, allowing better recall. Vicente and Wang (1998, p. 57) put it as follows: “The links between levels define what is goal relevant for that domain, thereby allowing people who are aware of these relations to reduce the number of meaningful alternatives that need to be considered in reconstructive recall” (p. 57).

The Chunking and Template Theories

Simon and Chase's (1973) chunking theory (CT), which is closely related to the EPAM theory of memory and perception (Feigenbaum & Simon, 1984; Richman & Simon, 1989; Richman et al., 1995), was the first theory to specify precisely the cognitive mechanisms involved in chess memory. CT proposes that at the core of expertise lies the ability to rapidly recognize important features in the problem at hand, such as the location of groups of pieces on the board. These features are stored internally as chunks—symbols in long-term memory (LTM) having an arbitrary number of subparts and properties—that can be used as knowledge units. Chunks act as access points to semantic LTM and as the conditions of productions, whose actions may be carried out internally or externally. In the case of chess expertise, perceptual mechanisms in CT allow recognition of patterns of pieces on the board. These patterns suggest moves, which are used to update the internal representation of the board in the “mind's eye.” The mind's eye acts as a relational system that stores perceptual structures. These structures, which can originate both from external inputs and from memory stores, can be manipulated by visuospatial mental operations. Finally, recognition mechanisms apply both when perceiving the external position and when examining the positions that are generated in the mind's eye during search. A subset of CT was implemented as a computer program by Simon and Gilmartin (1973).

Two weaknesses were later uncovered with CT: (a) In an interference paradigm, the theory underestimates speed of encoding into LTM (Charness, 1976), and (b) the theory does not relate mechanisms at the chunk level with the higher level representations that players use (Cooke, Atlas, Lane, & Berger, 1993; Holding, 1985). TT (Gobet & Simon, 1996c, 1998; 2000) was developed to eliminate these two weaknesses while keeping the strength of the original chunking theory. TT is able to account both for perceptual phenomena, such as players' eye movements during short presentations of unknown positions, and memory phenomena, such as immediate recall of positions. The theory is implemented as a computer program, Chunk Hierarchy and REtrieval STRuctures (CHREST), that offers detailed simulations of the empirical data (de Groot & Gobet, 1996; Gobet, 1993, 1998; Gobet & Simon, 1996a, 2000). Finally, an extension of the program (Gobet & Jansen, 1994) plays (weak) chess by pattern recognition—recognized chunks elicit potential moves or sequences of moves.

Like CT, TT proposes that chunks are accessed by traversing a discrimination net. A discrimination net consists of a set of nodes (chunks) connected by links, which together form a treelike structure. The nodes have tests, which can be applied to check features

of the external stimuli. The outcome of each test determines which link will be taken below a node. This net is grown by two learning mechanisms, *familiarization* and *discrimination*, which together produce a self-organizing, dynamical system. When a new object is presented to the model, it is sorted through the discrimination net, starting from the root node, until no further test applies. When a node is reached at the end of this process, the object is compared with the *image* of the node, which is the internal representation of the object. If the image underrepresents the object, new features are added to the image (familiarization). If the information in the image and the object differ on some feature or some subelement, a new link and a new node are created below the current node (discrimination).

Chunks are linked to other information stored in LTM, such as moves, plans, and tactical motives. (In its current implementation, CHREST acquires only information about patterns of pieces and sequences of moves.) Moreover, chunks that are often used in a player's practice evolve into more complex data structures (templates), which have *slots* allowing variables to be instantiated rapidly. In particular, information about piece location, piece type, or chunks can be (recursively) encoded into template slots. Slots are created at chunks where there is substantial variation in squares, pieces, or groups of pieces in the test links below. In addition to slots, templates contain a *core*, basically similar to the information stored in chunks. Altogether, templates typically store about 10 pieces, but the number of pieces can be much larger—more than 20 in some cases.

Templates may be linked to other templates, allowing search at a higher level than moves (i.e., planning). TT also proposes that the information stored in the mind's eye decays rapidly and that it needs to be updated either by inputs from the external world or by inputs from memory structures. Finally, TT, like CT, proposes that search is carried out by recursively applying pattern recognition processes to the internal representation (Gobet, 1997, 1998).

Memory for Random Chess Positions

Given that the amount of constraint in the chess stimulus is central to the CAH, one way to examine the CAH is to manipulate this variable and to observe the effects on recall. The amount of constraint can be manipulated by various randomization procedures. Indeed, historically, random material has been one of the most powerful tools for studying experts' memory in contrived tasks (see Gobet, 1998). It is important to note that Gobet and Simon (1996a, 1996b) have shown that chess experts keep some superiority with random positions, even if the skill advantage is

much less than with game positions. As simulated by CHREST (Gobet & Simon, 1996a, 2000), this result fits the predictions of the chunking and the template theories. Both of these theories assume that experts acquire a large number of perceptual chunks and that they are therefore more likely than nonexperts to pick up patterns that occur by chance in random positions.

At first sight, it might be thought that the presence of expertise effects in recall of random positions should disentangle the CAH and TT. If a position is random, there are presumably no goal-relevant constraints that can be exploited by experts, and therefore the CAH should not predict an expertise effect. However, Vicente and Wang (1998) noted that published studies have used a randomization process in which the location of the pieces on the board, but not their distribution, is randomized. In most experiments, the random positions were generated by randomizing the pieces of positions from *master-level games* after about 20 moves. Their insight is that these master-level games may have statistical properties at the board level that will be inherited by the random positions. For example, suppose that a master-level position after 20 moves has an 80% chance of containing one white queen and one black queen. The experts—who know the board level constraints of master-level games—may be able to exploit this constraint to aid recall. In contrast, the weaker players cannot exploit these constraints. Therefore, both the CAH and TT predict a small skill effect on classic random positions and cannot be distinguished with this material.

However, Vicente and Wang (1998) identified a type of random position for which the CAH makes a different prediction than does TT:

Our observation leads to a new experiment that contrasts our theory with that of Gobet and Simon (1994, 1996b). The two theories make different predictions for a truly random chess position, in which both the position and the selection of the pieces are randomly determined (as far as we know, this procedure has not been adopted before). Gobet and Simon (1996b) predicted a small but significant expertise effect for completely random positions, whereas our theory would predict no expertise effect. (p. 45)

We tested Vicente and Wang's (1998) prediction in this study. Appendix A shows the algorithm we used to generate truly random positions, and Figure 1 shows an example of a position (truly random) generated by this algorithm. As can be seen, this position violates the most basic board level constraint of chess—that a position has one white king and one black king. In addition, the position contains other unusual features such as three black knights (two or fewer would be expected) and three white bishops (two or

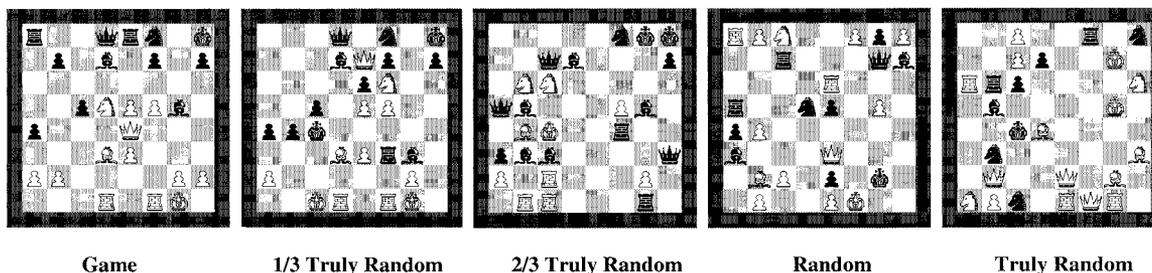


Figure 1. Examples of stimuli used in the experiment.

fewer would be expected). Thus, unlike the classic random stimuli, these positions do not inherit the statistical properties of master-level positions. There are no constraints that can be exploited by the masters but not by the weaker players. Thus, the CAH predicts no skill effect in recall of these positions, whereas TT predicts a small skill effect (we later quantify this effect), because experts are still more likely to recognize more chunks in these positions than nonexperts.

Overview of Experiments

We first address the effect of true randomization on recall with a computer simulation. This allows us to formalize TT's predictions about the relationship between skill and recall on truly random positions. Additionally, we included positions where only one third or two thirds of the pieces are truly randomized; this allows more fine-grained testing of TT's quantitative predictions (as well as of the CAH's ordinal predictions). We selected three levels of true randomization (i.e., one third, two thirds, and fully randomized), as we wished to examine the relationship between degree of randomization and template access by TT's large nets (and by the strong human players) while being careful not to overburden the human participants with too many experimental conditions (and therefore trials). This was particularly relevant, as we also included game and random positions that are typically used in expert memory studies in chess. In the human experiment that followed, chess players had to recall the same types of positions as were presented in the simulations. In this article, we show that in the case of memory for chess positions, both TT and the CAH make a number of correct predictions but that in the truly random condition proposed by Vicente and Wang (1998) TT's prediction—but not CAH's—is supported.

Computer Simulation Study

As noted above, Vicente and Wang (1998, p. 45) proposed an experiment that makes it possible to discriminate between the CAH and TT. The idea was to use truly random positions where not only the location but also the distribution of the pieces is randomized. The CAH predicts no skill difference for these positions, whereas TT predicts a small skill difference. As indicated above, we extended their proposed experiment by using positions where one third of the pieces are truly randomized and positions where two thirds of the pieces are truly randomized. These additional position types are useful for generating further predictions from TT and the CAH. As noted above, we also included game and random positions that are typically used in chess memory studies.

Method

Materials

Five types of positions were used. We refer to these as *game*, *random*, *one-third truly random* (¹), *two-thirds truly random* (²), and *truly random* positions. Five hundred stimuli were selected by randomly sampling (without replacement) from a database of 3,100 positions taken from master-level games after about 20 moves. In the game condition, the stimuli were kept unchanged. The four types of random positions were generated using the algorithms described in Appendix A, and examples of all position types are shown in Figure 1.

Procedure

Because TT is implemented as a computer program (CHREST), one may easily run it on the material proposed above and obtain quantitative predictions. We used the same version of CHREST that was used by Gobet and Simon (2000)—no parameter was altered to fit the data. We created 16 nets that ranged in size from 100 to 300,000 nodes, by letting the program scan a database of about 50,000 positions taken from master-level games (the positions were middle-game positions, taken from games played in the last 50 years). During the learning phase, the program moved its simulated eye around the board and attempted to learn the patterns within its visual field (± 2 squares from the fixation point) using the mechanisms of familiarization and discrimination. Template slots were created when two conditions were satisfied: (a) The number of nodes below a given node that share identical information (either a square, a type of piece, or a chunk) is greater than 3, and (b) the node to which a slot could be added contains at least five elements. Slots could encode only information referring to their types: that is, either to kind of piece, location, or chunks having a common subchunk (see chapter 8 of de Groot & Gobet, 1996, and Gobet & Simon, 2000, for details on the model).

During the testing phase, each of the 16 nets was used to simulate the recall of 500 positions of each type. Each position was presented for a simulated time of 5 s. During the presentation of the position, the model moved its simulated eyes around the board and attempted to recognize chunks (or templates). In the case of successful recognition, a pointer to the chunk was placed in visual short-term memory (STM), the capacity of which was set to 3. During the 5 s of presentation, the model could add information to LTM in two ways: by familiarization (augmenting the internal representation of a pattern), which took 2 s, and by filling information into a template slot, which took 250 ms. The third learning mechanism, discrimination, could not be used by the model, because its execution takes longer than the 5-s presentation time. All time parameters were the same as they were in the simulations reported by Gobet and Simon (2000). One should note that TT's mechanisms are the same in all position types (from game to truly random). Thus, any difference in recall performance reflects the probability that external patterns will elicit chunks or templates in LTM.

Data Reduction and Analysis

We selected the 4 of the 16 nets (with 1,010, 3,008, 15,003, and 300,009 chunks) that most closely matched the mean recall of four groups of human subjects on game positions (reported in detail later in the human study¹). To facilitate comparisons with the human data, we considered the 300,009-chunk model to have a rating of 235.6 in *human* units (the mean skill rating of our top group of humans). That is, we treated the 300,009-chunk model as a computer participant with a rating of 235.6. The 15,003-chunk model was considered to have a rating of 201.2 in human units (the mean skill rating of our second group of humans), the 3,008-chunk model a rating of 150.9, and the 1,010-chunk model a rating of 112.3.

We derived estimates of the skill slopes in each position type in the following manner. We used linear regression to predict recall performance (the dependent variable) from the ratings (the independent variable). We did this for each position type (i.e., game, $\frac{1}{3}$, $\frac{2}{3}$, random, and truly random) separately. We used the unstandardized parameter estimates from

¹ For now, the reader should note that we collected recall data on four groups of human participants with mean British Chess Federation (BCF) ratings of 235.6 (Grandmasters, the strongest group), 201.2 (Masters/Experts), 150.9 (Class A/B players), and 112.3 (Class C/D players, the weakest group). As described in more detail later, the mean recall on the game positions for the four groups was 83.7%, 61.3%, 55.1%, and 40.1%, respectively. As indicated in the text, these percentages were used to select four nets (displaying similar recall); the recall performance of these nets on the random positions was further examined.

these models as an index of the skill slope in each position type, and we used the *p* value of the parameter estimate to determine whether the slope was significantly different from zero. To determine whether the skill slope on one position type was steeper than that on other position types, we used regression analysis (*proc glm* in SAS) to test for Skill × Position Type interactions, where position type was entered as a repeated measures variable. To compare the skill slope on the 1/3 positions with the mean slopes on the other random positions, we used the helmert comparison option in SAS. Given the limited sample size, one-tailed tests were used.

Results

Figure 2A shows the recall performance of the four selected nets on the different position types. In general, recall improves as a

function of the number of chunks (because more and bigger chunks can be accessed with larger nets) and decreases as a function of the amount of randomization (because, as randomization increases, the likelihood of matching a chunk decreases). The estimates of the slopes in the different position types were as follows: game = 0.331 (*SE* = 0.032), 1/3 = 0.216 (*SE* = 0.058), 2/3 = 0.077 (*SE* = 0.021), random = 0.071 (*SE* = 0.016), and truly random = 0.061 (*SE* = 0.011); the skill slopes for all position types were significantly different from zero (all *ps* < .05). For the truly random positions, the parameter estimate (0.061) indicates that an increase in skill level of 100 human units (about the difference between an average club player and a Grandmaster) yields an increased recall performance of 6.1%, or roughly 1.5 pieces.

Regression analysis indicated that the skill slopes on the four randomized position types (1/3, 2/3, random, and truly random) were significantly different from one another: Skill × Position Type interaction, *F*(3, 6) = 11.1, *p* = .004 (data from the game position types were not included in this analysis). Figure 2 indicates that the skill slope for the 1/3 positions is steeper than it is for the other randomized positions; statistical analysis confirmed that the skill slope on the 1/3 position types was greater than the mean slope on the other three random position types, *F*(1, 2) = 11.6, *p* = .04. The slopes on the random and truly random position types did not differ significantly from each other, *F*(1, 3) = 1.36, *p* = .18; the same was true for the 2/3–truly random comparison, *F*(1, 3) = 2.01, *p* = .15.

Discussion

The purpose of the simulation study was to formalize TT's predictions. As we expected, TT predicts a skill effect in all position types, including the truly random positions. TT also predicts that skill effects become smaller with randomization. This Skill × Position Type interaction is mainly explained by the fact that (a) large nets make it possible for CHREST to access more and larger chunks, and thus to obtain better recall performance, and (b) chunks get harder to find as positions get more random. The abrupt change in slope with game and 1/3 position is explained by the fact that CHREST can find templates and use their slots to fill in either the piece location or its type rapidly in game positions and, to some extent, in 1/3 positions, but not in the other types of positions. With smaller nets, CHREST can rarely find templates, even in game positions, and has to rely mainly on chunks.

There are only small differences between the skill effects on the 2/3, random, and truly random position types, and they do not reach statistical significance. Thus, taken at face value, TT predicts no difference in skill slopes on these three position types. However, we note that these analyses had limited power to detect significant Position Type × Skill interactions. In any case, TT predicts that the difference in slopes between the 1/3 and truly random conditions should be greater than the difference in slopes between the 2/3 and truly random conditions and greater than the difference in slopes between the random and truly random conditions.

Human Study

The main purpose of the study was to compare TT's and the CAH's predictions with human data. As indicated above, TT

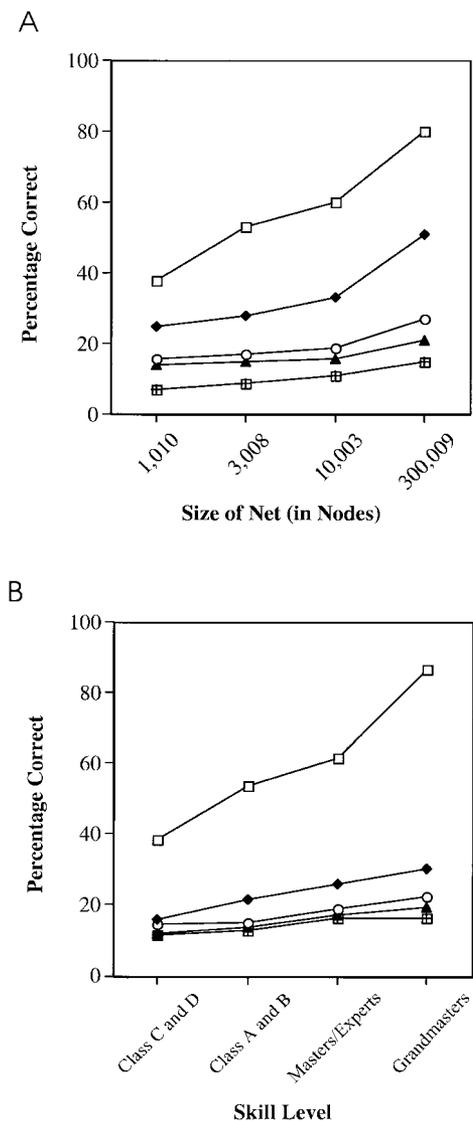


Figure 2. Percentage of pieces correctly recalled as a function of type of position and number of chunks in the discrimination network for the model (A) and of skill level for the human data (B). Open squares indicate game positions; filled diamonds indicate 1/3 truly random positions; open circles indicate 2/3 truly random positions; filled triangles indicate random positions; checkered squares indicate truly random positions.

Table 1
Age and Chess-Related Data for the Four Groups

Measure	Grandmasters (<i>n</i> = 7)		Masters/Experts (<i>n</i> = 12)		Class A/B (<i>n</i> = 10)		Class C/D (<i>n</i> = 7)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	34.7	3.8	28.7	8.2	25.5	9.3	25.6	5.8
BCF rating, August 1998	235.6	5.6	201.2	14.8	150.9	11.9	112.3	9.1
FIDE rating, July 1998	2,493	44	2,245 ^a	98	2,045 ^b			
No. slowplay games in the past year	67.6	49.8	46.4	31.4	20.7	13.5	28.3	14.3
No. rapidplay games in the past year	20.9	27.6	25.2	27.3	19.3	17.1	7.6	5.3
Days since last slowplay game	22.4	19.9	38.6	24.0	63.6	39.2	44.1	31.3
Studying time (hours per week)	16.4	13.7	6.4	9.1	3.7	3.2	1.6	1.7

Note. Grandmasters: 7 men; Masters/Experts: 10 men, 2 women; Class A/B: 9 men, 1 woman; Class C/D: 7 men. *Slowplay* games normally involve each player having more than 75 min thinking time for a certain number of moves. *Rapidplay* games are those in which the participant has about 30–40 min thinking time for the game. BCF = British Chess Federation; FIDE = Fédération Internationale des Echecs.

^a An average of 11 players; one expert did not have a FIDE rating in the July 1998 list.

^b The rating of one Class A player who had a FIDE rating in the July 1998 list.

predicts a skill effect on all position types, including the truly random positions. In contrast, the CAH does not predict a skill effect on truly random positions. It does predict skill effects on the other position types.

On the randomized positions, TT predicts that skill effects become smaller in the following manner: $\frac{1}{3}$ (predicted slope = 0.212) > $\frac{2}{3}$ (predicted slope = 0.077) > random (predicted slope = 0.071) > truly random (predicted slope = 0.061). TT also predicts that the slope on the $\frac{1}{3}$ positions is greater than the mean slope on other randomized position types. The CAH does not give quantitative estimates of the skill effects but predicts that the ordering of the skill slopes should be as follows: game > random > truly random and game > $\frac{1}{3}$ > $\frac{2}{3}$ > truly random. As we indicated earlier, the CAH predicts that skill effects should be greater in random positions than in the truly random positions, because the random positions inherit constraints that benefit the Masters more than the other players. The CAH does not make any straightforward predictions as to whether the skill slopes for the random positions should be smaller or greater than those in the $\frac{2}{3}$ (and $\frac{1}{3}$) position types.

In sum, the main goal was to examine whether skill effects were present in truly random positions, as predicted by TT but not by the CAH. We also evaluated the other predictions of TT and the CAH about the size and ordering of skill effects in different position types.

Method

Participants

Thirty-six participants completed the study. All participants were active chess players who had BCF ratings in both the August 1997 and August 1998 ratings list. Nineteen participants also had international ratings in the World Chess Federation (*Fédération Internationale des Echecs*; FIDE) July 1998 ratings list.² Table 1 shows the characteristics of the sample by skill level. We split the participants into four groups. The participants in the top group (Grandmasters, *n* = 7) were players with ratings of above BCF 225. The participants in the second group (Masters/Experts, *n* = 12) had BCF ratings between 175 and 224. This group contained three male International Masters, one male FIDE Master, one female Grandmaster,

one female International Master, and six male Experts. The third group (Class A and B players, *n* = 10) consisted of players with ratings between 125 and 174; these players can be thought of as moderate to strong club players. The final group (Class C and D players, *n* = 7) contained players with ratings lower than BCF 125; these players are considered weak club players. The Grandmasters were paid £40 (U.S. \$64) for participating, Masters/Experts were paid £20 (U.S. \$32), and all other participants were paid £10 (U.S. \$16). Because previous studies (e.g., Gobet & Simon, 2000) have shown that Masters are not very keen to recall random positions, a monetary reward was used to try to keep players' motivation level high; for each skill level, the prize for the best player was the equivalent to the participation fee for this skill level. The names of the winners were kept confidential. The experiment lasted on average about 1.5 hr.

Materials

Chess stimuli. The same five types of position were used as were used in the simulation study (see Figure 1). Twenty-five positions (with an average of 25 pieces) were taken from master games after about 20 moves and were randomly assigned to one of the five types of positions. There were thus five positions in each condition.

Presentation software and hardware. Chess stimuli were presented on a portable Apple Macintosh computer using specialized software for presenting chess stimuli and recording responses (see Gobet & Simon, 1998, for a detailed description of the software used). Participants were required to use a mouse to select pieces, move pieces onto squares, and delete pieces. To go on to the next trial, the participant pressed an *OK* button on the top left corner of the computer screen.

² FIDE publishes rating lists of its members every 6 months and awards the following titles, in descending order of merit: Grandmaster, International Master, and FIDE Master. The BCF rating is an interval scale ranking competitive chess players in the United Kingdom. Skill levels have standard names, which are used in this article (the approximate corresponding ranges in BCF points are in parentheses): Grandmaster (normally above 240), International Master (225–240), Master (200–225), Expert (175–200), Class A players (150–175), Class B players (125–150), and so on. Chess ratings tend to be stable between years in adult players; in our sample, the correlation between the BCF August 1997 and August 1998 grades was .99.

Chess questionnaire. Participants completed a questionnaire assessing basic information on chess play in the past year, such as the number of tournament games played, the time since the last game played, and the average amount of time per week each player spent studying chess. Because some players report that they find it much easier to look at board positions (i.e., a position with pieces on a chess board) than chess diagrams (i.e., the representations used in this study), the questionnaire also contained the following question: "How easy do you find it to read chess positions presented in two dimensions (on a page or computer screen) compared with the three dimensional board positions?"

Computer questionnaire. Participants filled out a brief questionnaire assessing computer use in which they were asked to rate their abilities at using a computer mouse on a 7-point scale from 1 (*not skilled at all*) to 7 (*very skilled*).

Motivation questions. Participants were asked to rate their motivation to perform the task, their interest in the task, and their irritation with the task on 7-point Likert-type scales both after completing the game positions and after completing the random positions.

Visual memory test. All participants completed a test of visual memory (the Shape Memory Test from the Educational Testing Service [ETS] kit of factor-referenced cognitive tests; Ekstrom, French, Harman, & Derman, 1976). Ekstrom et al. reported that the cognitive tests within the ETS kit exhibited good to excellent internal reliability, generally in excess of 0.70 (when computed by calculating the split-half correlation and adjusting using the Spearman-Brown correction). Factor analyses indicated that the shape memory test loaded strongly onto a visual memory factor (Ekstrom et al., 1979).

Procedure

Participants first completed the chess and computer questionnaires, followed by the visual memory test.³ Participants then completed a brief measure of competence at using the computer: The starting position was shown for 5 s on the computer screen, and participants were subsequently required to reconstruct it as quickly and as accurately as possible on an empty chess board. This was followed by the recall task. On each trial, a position was presented for 5 s. The screen was then blank for 2 s, and then an empty chess board appeared. The participants were instructed to try to recall the positions as completely and as accurately as possible. They had unlimited time to make their responses. The 20 random positions (four position types \times five stimuli) were first presented in a different random order for each participant. After participants completed the motivation questions, the five game positions were presented, also in a different random order for each participant, and participants completed the motivation questions once again. Participants had two practice trials on random positions and one practice trial on game positions.⁴

Data Reduction and Analysis

We used participants' ratings in the BCF August 1998 rating list as an index of chess skill, as this rating reflects performance in games most proximal to the completion of the experiment. We used Pearson's r to test the relationship between skill and recall on the position types. To obtain estimates of the skill slopes, we performed regressions in which recall performance (the dependent variable) was predicted from skill, age, and visual memory entered together (the independent variables). Age was entered into the models because research has shown age to be an important variable in chess memory (Charness, 1981a, 1981b) and because age could plausibly be related to recall on this task (Schulz & Salthouse, 1999), even under the relatively restricted range of the study sample (age range 18–47 years). In addition, controlling for age is particularly relevant in this study as age and chess skill were correlated in our sample ($r = .38, p = .02$), with the better players being older. Thus, age effects could potentially reduce the size of the association between skill and recall performance; by includ-

ing age in the regression models we were able to obtain estimates of skill slopes when statistically controlling for any effects of age. Visual memory performance was included in the regression models because general visual memory ability could plausibly be associated with both chess skill (e.g., Frydman & Lynn, 1992) and recall performance on this task (Waters et al., 2002), and so it is important to demonstrate that any relationship between chess skill and memory recall is independent of general visual memory ability. Indeed, controlling for general visual memory ability is particularly important in this study because TT predicts skill effects (of different magnitude) in all position types, thereby leaving no position type in the study to serve as a control for general visuospatial memory ability. Separate regressions were carried out for each position type. The unstandardized parameter estimate of the skill variable (BCF rating) provided our estimate of the skill slope, and the p value of this statistic determined whether it was significantly different from zero. To determine whether the skill slope on one position type was steeper than that on another position type, we tested Skill \times Position Type interactions in which position type was entered as a repeated measures variable with two levels. To compare the skill slope on the $\frac{1}{3}$ positions with the mean on the other random positions, we used the helmert comparison option in SAS. Because the model makes directional and unambiguous predictions about recall performance on all conditions, we used one-tailed tests for all analyses of skill effects on recall; two-tailed tests were used for effects of age and visual memory on recall and for all other analyses.

Results

First, we examined the data on game and random positions; the recall percentages were 83.7 ($SD = 14.4$), 61.3 ($SD = 19.4$), 55.1 ($SD = 17.2$), and 40.1 ($SD = 10.1$) on the game positions and 18.3 ($SD = 6.95$), 17.0 ($SD = 6.25$), 14.1 ($SD = 4.71$), and 12.3 ($SD = 5.55$) on the random positions for the Grandmasters, Masters/Experts, Class A/B players, and Class C/D players, respectively. These values are typical of those that have been observed in the literature (e.g., Gobet & Simon, 1996b). As we expected, there was a strong correlation between skill and recall of game positions ($r = .68, p < .001$, one-tailed). There was also a smaller, but reliable, correlation between skill and recall on the random positions ($r = .36, p < .025$, one-tailed). Thus, consistent with the prediction of the TT, the prediction of the CAH, and previous data (Gobet & Simon, 1996b), better players have better recall on the random positions even with presentation time as short as 5 s.

Are There Skill Effects in Recall of Truly Random Stimuli?

Next, we focused on whether there were skill differences even on the truly random positions. Figure 2B shows recall as a function of group and position. Recall on the truly random stimuli was generally poor; the mean percentages were 14.8 ($SD = 5.14$), 16.3 ($SD = 4.47$), 13.7 ($SD = 4.95$), and 12.0 ($SD = 2.68$) from

³ Details of administration and analysis of this test are presented in Waters, Gobet, and Leyden (2002).

⁴ This experiment is part of a study in which both randomization and the location of pieces on squares (pieces on center of squares vs. pieces on the intersection of squares) were manipulated. Presentation of standard and intersection positions was blocked, and the order of presentation counter-balanced over participants. Here we report data from the standard positions; the results from intersection positions will be reported in a separate article.

stronger to weaker players. There was a small but reliable correlation between skill and recall on the truly random positions ($r = .34, p < .025$, one-tailed).⁵

Table 2 shows the results of the multiple regression analyses for all five conditions. On the truly random positions, skill was a significant predictor of recall performance. The skill slope, indexed by the unstandardized parameter estimate, was 0.052 ($SE = 0.016$), which compares to TT's prediction of 0.061. Thus, an increase of 100 BCF grading points (e.g., the difference between an average club player and a Grandmaster) yielded an increase of about 5.2% in recall, which corresponds to about 1.25 pieces. TT predicts that there is a linear relationship between skill and recall; consistent with this, a quadratic term for skill in a second block in the regression did not significantly improve prediction of recall ($p = .19$). Table 2 also indicates that age is a predictor of recall on the truly random positions, with recall declining about 2.6% for each decade of life (age was not, however, univariately associated with recall performance on these positions: $r = -.24, p = .15$). Visual memory was not a predictor. None of the possible two-way interaction terms approached significance (all $ps > .37$).

We note that the relationship between skill and recall is a correlation and that the effect size is small. It is possible that other factors unrelated to knowledge of chess structure underlie this association. We considered four possible confounds.

Better players have better visual memories in general. There was no evidence for this explanation because there was no correlation between skill and performance on the visual memory task (Waters et al., 2002) and because performance on the visual memory task did not predict recall performance on these positions (see Table 2). In our sample, the visual memory test exhibited a split-half correlation of .76 when the Spearman-Brown correction is used, providing further evidence for the reliability of the measure.

Better players were better at using the computer interface. There was no evidence for this explanation because the better players reported being (nonsignificantly) less skilled at using the mouse ($r = -.18, p = .28$) and were (nonsignificantly) slower to manipulate the mouse in reconstructing the opening position ($r = -.19, p = .29$). There was no correlation between chess skill and preference for viewing 3-D board positions versus 2-D positions ($r = .01, p = .94$).

Better players were more motivated to perform well. The data showed that the better players reported trying (nonsignificantly) less hard ($r = -.16, p = .34$), being marginally less interested ($r = -.32, p = .054$), and being (nonsignificantly) more irritated ($r = -.22, p = .19$) by the random positions. Although these data refer to responses on all random conditions rather than just the truly random condition, it seems unlikely that different data would be obtained if effort, interest, and irritation responses had been elicited from just the truly random trials. Thus, differences in motivation are unlikely to underlie the skill effect on the truly random condition.

Better players guess more, and their guesses are of higher quality. Better players may be more willing to guess "blindly" when performing the task and gain a few extra pieces through pure guesswork. If participants make a lot of guesses, then we would expect them to show large numbers of errors of commission (i.e., incorrectly placed pieces). Table 3 shows that the better players (i.e., the Grandmasters and Masters/Experts) did not make many

more errors of commission than did the weaker players.⁶ The correlation between skill and number of errors of commission on the truly random condition was .03 ($p = .87$).⁷

Skill Effects on 2/3, Random, and Truly Random Positions

Table 2 indicates that the skill slopes for the 2/3 condition (0.063), the random condition (0.059), and the truly random condition (0.052) follow the ordering predicted by TT. Statistical analyses indicated that the Position Type \times Skill interactions for the 2/3–truly random comparison, $F(1, 32) = 0.25, p = .31$, one-tailed, and the random–truly random comparison, $F(1, 32) = 0.11, p = .37$, one-tailed, did not approach significance.

Skill Effects on 1/3 Positions

Table 2 and Figure 2 indicate that the skill gradient was greater on the 1/3 condition than it was on other randomized positions. Statistical analysis confirmed that the skill slope on the 1/3 condition was greater than the mean slope on the other three conditions, $F(1, 32) = 4.15, p = .025$, one-tailed.

We noted that the skill slopes on the randomized conditions tended to be less steep than those predicted by TT, particularly in the 1/3 condition (humans, slope = 0.110, $SE = 0.033$ vs. TT, slope = 0.216, $SE = 0.058$). As indicated in Figure 2, the Grandmasters did not perform as well on these positions as TT had predicted (Grandmasters, $M = 28.1, SE = 4.1$ vs. TT, $M = 53.0$). It is possible that the stronger human players adopted strategies on the random positions that hindered recall performance on the 1/3

⁵ The correlation between skill and recall for the 2/3 condition was $r = .33, p < .05$, one-tailed. For the 1/3 condition, the correlation was $r = .46, p < .005$, one-tailed.

⁶ The large means and standard deviations for the errors of commission in the truly random condition in Table 3 were caused by a few participants who appeared to guess prolifically; 7 participants placed at least five more incorrect pieces than correct pieces (on average) on this condition. These participants ($n = 7, M = 13.6\%, SD = 6.4$) did not perform better on the truly random position types than did the nonguessers ($n = 29, M = 14.6\%, SD = 4.1\%$), despite having (nonsignificantly) higher ratings (guessers, $M = 186.9, SD = 36.4$ vs. nonguessers, $M = 174.1, SD = 47.3$).

⁷ Pure guessing is unlikely to greatly improve recall on the truly random positions. We simulated the effects of guessing by determining the amount of pieces that would be "recalled" correctly through two guessing methods: guessing from the legal distribution of chess pieces at the beginning of a game (i.e., one white king, one white queen, eight white pawns, etc.) and guessing from a rectangular distribution. The estimates were based on 1,000 positions with an average of 24.6 pieces. Both methods showed that putting down 32 pieces on each truly random position would yield one piece correct and that putting down 10 pieces gives an expected yield of 1/3 of a piece correct. Of course, the kind of guessing that participants will use in the recall task is likely to be more sophisticated than pure guesswork. As an example of this, one Grandmaster verbalized after one random trial that he knew there was not a knight in the position. He was not sure what there was, but he was sure there was not a knight. If he were then to guess, he could use his knowledge of the absence of a knight to constrain his response, and so this would not represent pure guesswork. Thus, better players' "guesses" may be of higher quality than weaker players', which may be partially explained by the larger chunks they can access (see Gobet & Simon, 2000). However, an investigation of this issue was beyond the scope of the present study.

Table 2
Results of Multiple Regression Analyses in Which Skill, Age, and Visual Memory Are Entered Together in One Block

Position type	R^2 (%)	BCF rating		Age		Visual memory	
		<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>
Game	50	0.35††	0.063	-0.50	0.37	-0.45	0.40
1/3	31	0.11††	0.033	-0.22	0.19	0.33	0.21
2/3	26	0.063†	0.022	-0.29*	0.13	0.09	0.14
Random	17	0.059†	0.023	-0.16	0.14	-0.09	0.15
Truly random	28	0.052††	0.016	-0.26**	0.09	-0.06	0.10

Note. For each position type, a regression was performed by entering British Chess Federation (BCF) rating, age, and visual memory into the same block. On the 2/3 positions, an increase in 100 BCF points yields about a 5.2% increase in recall. 1/3 = one-third truly random; 2/3 = two-thirds truly random.

* $p < .05$, two-tailed. ** $p < .01$, two-tailed. † $p < .01$, one-tailed. †† $p < .005$, one-tailed.

positions. For example, some players reported that they just concentrated on a small part of the board (e.g., the first row or a quadrant of the board) on the nongame positions and just tried to remember the pieces in these areas. We tested the effect of this strategy on TT's performance by running simulations on which TT fixated only two rows on the board. As we expected, this strategy had the effect of decreasing recall performance on the 1/3, 2/3, random, and truly random positions, and the resulting skill slopes more closely matched the human skill slopes. Because we did not formally assess the strategies that participants adopted, we were unable to carry out more detailed simulations on this issue.

Overall Fit Between Model and Data

Figure 3 shows the fit between the predicted scores and the human observed scores for the 20 data points (4 skill levels \times 5 position types). The model captures the human data well ($r^2 = .889$, $p < .001$, RMSD = 6.71) with respect to both the overall trend and the absolute values. The main deviations occur on the 1/3 positions, where—as discussed above—the model overestimates the human data (on average, by 12.1%), and on the truly random positions, where the model underestimates the human data (on average, by 3.7%). One should note also that the variability of the human data, and in particular of the model, is low. The fit remains good when the data from the game positions, which were used to match the model with the human data, are excluded from analyses ($r^2 = .885$, $p < .001$, RMSD = 7.21).

Discussion

The main finding from the experiment was that there was a small, but reliable, skill effect on the truly random positions. There was no evidence that this association was driven by any confounding variable. Thus, the prediction of TT was upheld, and the prediction of the CAH refuted.

TT's prediction that the skill slope on the 1/3 condition should be greater than the slope on the other random positions was also upheld. We also note that the skill slopes on the 2/3, random, and truly random conditions did not reliably differ, as was observed in the simulations. Although the experiment may have had insufficient power to detect differences in skill slopes on these conditions, the parameter estimates presented in Table 2 tend to argue against a real interaction in the population. In general, the overall fit between TT and the human data was good. As noted earlier, we used the same version of CHREST as that used by Gobet and Simon (2000) to simulate the role of presentation time on memory recall, with no free parameter adjusted to fit the human data. Hence, the correspondence between model and data supports the chunking mechanisms embodied in CHREST.

The CAH's predictions concerning the relative sizes of skill effects (game > random > truly random, and game > 1/3 > 2/3 > truly random) were all upheld. Skill effects were greater in stimuli with more constraint. The CAH would have received further support if the skill slopes on the random stimuli had been significantly greater than those on the truly random stimuli, and if the

Table 3
Mean Numbers of Errors of Commission by Position Type

Position type	Grandmasters ($n = 7$)		Masters/Experts ($n = 12$)		Class A/B ($n = 10$)		Class C/D ($n = 7$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Game	2.43	2.37	4.48	3.22	4.34	2.78	3.94	3.08
1/3	4.00	4.44	2.55	4.24	4.44	4.83	2.40	2.61
2/3	4.40	5.85	3.15	4.82	5.70	6.78	2.69	2.67
Random	4.31	5.07	3.15	4.09	5.98	6.71	3.17	1.93
Truly random	6.14	7.75	3.28	4.51	6.04	7.47	2.91	2.00

Note. Errors of commission = pieces placed incorrectly. 1/3 = one-third truly random; 2/3 = two-thirds truly random.

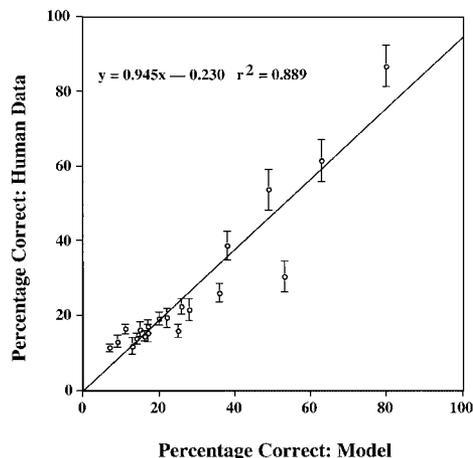


Figure 3. Predicted (x-axis) and observed (y-axis) percentage recall on the different conditions. The vertical error bars indicate the standard errors of the mean for the human data. The standard errors of the mean for the model (not shown in the graph) are all less than 1%.

skill slopes on the $\frac{2}{3}$ stimuli had been significantly greater than those on the truly random stimuli, because the CAH predicts that “the more constraint available, the greater the expertise advantage can be” (Vicente & Wang, 1998, p. 36). However, as noted above, the experiment may have had insufficient power to detect these differences.

A secondary finding of the study was that age was associated with recall performance on the truly random positions (when controlling for the effects of skill), with younger participants recalling more pieces than their older counterparts. The significance of this finding for the theoretical mechanisms described in TT or the CAH is unclear, but it may suggest that recall performance on these taxing stimuli is facilitated by strategic flexibility or some other domain-general ability that may favor the young. Whatever the explanation, given that domain-specific knowledge and domain-general abilities may jointly underlie recall performance on the randomized stimuli, the finding underscores the importance of examining the effects of subject variables such as age in this paradigm.

The experiment also brought to light the role of strategic processes on recall performance. As noted earlier, some players reported using a limited fixation strategy when responding to the randomized positions, presumably as a cognitive coping strategy. It seems possible that chess players can use multiple processes to encode the randomized stimuli, with the task conditions influencing which process is recruited at any given time. Thus, a Master may process a truly random position differently when it is embedded in a sequence of game positions compared with when it is presented with other truly random positions.

The study had a number of limitations that should be addressed in future research. It would have benefited from having a larger sample size in order to derive point estimates for the skill slopes with smaller standard errors. As noted above, strategic factors may have influenced scanning and recall of the randomized positions; these strategies could be more fully assessed by eye-movement recordings as well as by postexperiment interviews or questionnaires so that their influence can be further evaluated. Finally,

these strategic factors may have had a different impact if the presentation of the randomized positions had been blocked by position type; this could be evaluated in future studies.

General Discussion

Our goal in this study was to test a number of predictions of two theories of expert memory—the CAH and TT—on the recall of randomized chess positions. In particular, our simulation study showed that TT predicts an expertise effect on the truly random positions, whereas Vicente and Wang (1998) noted that the CAH predicts no expertise effect on these stimuli. Our human study showed that stronger human players did recall truly random positions reliably better than weaker players, thereby supporting the prediction of TT. Both the CAH and TT make a number of correct predictions about the ordering of expertise effects over the different conditions, although TT overpredicted the expertise effect on the $\frac{1}{3}$ positions. In sum, there was support for both theories of expert memory, although TT’s prediction prevailed in the truly random condition.

Following Vicente and Wang, we defined the types of random positions by the algorithm that created the material. As can be seen in Figure 1 and Appendix B, this algorithm produced stimuli that violate the most basic board level constraints. Put another way, there is nothing about the statistical properties of the truly random stimuli that should give Masters an advantage over weaker players. Thus, Vicente and Wang explicitly stated that the CAH predicts no skill effect on truly random positions constructed like those in our study.

Now that a positive association between skill and recall on the truly random stimuli has been demonstrated, it may be tempting for proponents of the CAH to reconsider how the CAH might be able to accommodate this new finding. For example, as noted earlier, the lowest level for the abstraction hierarchy for chess includes the physical appearance of the board and pieces (Vicente & Wang, 1998, p. 56, Table A2). Because we used conventional representations for boards and pieces, it could be argued that there are appearance-related constraints in the truly random stimuli that can be exploited by experts to aid recall (Vicente, 2000). However, we would assume that these appearance-related constraints are of equal value to the weaker players, who—like the Masters—have presumably only experienced chess environments containing conventional boards and pieces. In lay terms, both Masters and weaker players play chess using chess pieces and chess boards that look the same. Given that the chess environments of stronger and weaker players do not differ at the level of physical appearance, the better players should not gain any differential benefit from appearance-related constraint.

TT predicts a skill effect on both the random and the truly random stimuli because low-level patterns arise by chance in these stimuli, and these patterns facilitate recall in the larger nets. Thus, it could be argued that constraints can also appear by chance and that the truly random stimuli can contain constraints that are exploitable by the experts. But there is no reason to suppose that the algorithm that generates the truly random stimuli will produce stimuli that contain constraints of more benefit to the Masters (compared with the weaker players). In general, the probability of producing a stimulus with even the most board-basic chess constraint (two kings of opposite colors) is small: We have computed

that only 1 of 34 truly random positions will have this property. But it is not clear that this board-level constraint would differentially benefit the Masters because this constraint is always present in the chess environments of both stronger and weaker players. In general, for the CAH to predict an expertise effect on randomized stimuli in a straightforward manner, there need to be constraints in the test stimuli that reflect the statistical properties of the Master-level chess environments. This is the case for the random positions, because these positions contain the distribution of pieces from Master-level games. It is also true for the $\frac{2}{3}$ positions, because some of the pieces and their location derive from Master-level games. But there is no reason to suppose that there are constraints in the truly random stimuli that reflect the statistical properties of the chess environments of Master-level players, and thus the CAH—as indicated by Vicente and Wang (1998)—does not predict an expertise effect in this condition.

In sum, we have new evidence that in chess—the domain Vicente and Wang use most often to support the CAH (in 6 of the 10 experiments they review)—the CAH does not do better in explaining skill effects in recall than does one process theory, TT. We acknowledge that this study examined expert memory in only a relatively narrow range of conditions and in only one domain. We also accept that TT did not predict expertise effects perfectly and that many of the CAH's ordinal predictions were supported. Nonetheless, our findings seem of interest given Vicente and Wang's (1998) claim that process theories "cannot provide an adequate theoretical explanation for expertise effects in memory recall for the considerable number of domains in which memory recall is a contrived task" (pp. 34–35).

The theoretical explanation for recall on this contrived task invokes several mechanisms. First, by learning chunks and templates, CHREST captures regularities in the environment. Second, templates are necessary to explain the huge skill superiority on the game positions, and to some extent on the $\frac{1}{3}$ positions, over the other types of positions. Third, attentional mechanisms are important, as was shown by the simulations of the $\frac{1}{3}$ positions. Finally, a limited visual STM is necessary to account for most of the simulations; with an unlimited capacity, little skill effect would be found as the information about individual pieces could simply be stored in STM. Thus, through a variety of cognitive mechanisms, CHREST provides an adequate theoretical explanation for expert recall on a contrived task.

Nonetheless, we recognize that cognitive mechanisms distinct from those proposed in this article may mediate the skill effects observed in this study. For example, Masters may be faster at encoding the features necessary to identify the symbols for pieces. Thus, a faster piece detector or a piece detector that operates in parallel over a larger visual field could conceivably contribute to skill effects on the randomized stimuli. (Simulation studies in TT in which encoding time is manipulated across skill levels could help determine the impact of this low-level mechanism.) One should note also that chess skill itself explained only a modest amount of variance in the recall of truly random positions (around 10%); presumably there are (as-yet-undiscovered) domain-general factors that are associated with recall on these positions. For example, participants may use an assortment of strategies during the performance of the task (e.g., Siegler, 1999), particularly under the more taxing random conditions, and individual differences in strategy use may plausibly explain some of the variance in recall.

Further work is required to more fully understand the multiplicity of mechanisms—both domain specific and domain general—that underlie recall performance on the randomized chess positions.

It is important to recognize that the ultimate goal of TT, within the chess domain, is to model human chess play. Our underlying assumption is that pattern recognition is the key process underlying chess skill. The critical idea is that TT learns patterns (chunks) not with the explicit goal of excelling in memory experiments but to use them as conditions of productions that may be useful later in playing chess (e.g., Gobet & Jansen, 1994). As with the chunking theory, expert performance in recall tasks is a side effect of acquiring goal-oriented knowledge for performing a task at a high level. That is, the chunks are an invaluable, if inadvertent, aid to performance in the recall task (Simon & Gobet, 2000).

When comparing the CAH and TT, it is important for one to distinguish between the process-product and the formal-informal dimensions, which are orthogonal. The CAH is clearly a product theory, and the TT is clearly a process theory. TT, implemented as a computer program, is a formal theory. Its implementation, CHREST, makes quantitative predictions, sometimes about small differences, which is certainly preferable to making predictions about ordering—the most informal models can do (cf. Grant, 1962; Meehl, 1967). The CAH is mostly formulated informally, except for the DUal REservoir System Simulation program, which specifies an abstraction hierarchy for thermal hydraulic process (Vicente & Wang, 1998, p. 53–55). In this case, the hierarchy allows more powerful predictions than do the informal hierarchies proposed in other domains such as chess. In particular, it is possible to manipulate several constraints in parallel and to make quantitative predictions. Thus, the desirability for formal theories applies both for process and product approaches.

In the same way in which the CAH can be used both formally and informally, TT can be applied to tasks even when no simulations are available. When used as a verbal theory, the EPAM family of models, of which TT is a member, can be shown to account for a large amount of data in other domains of expertise (cf. Simon & Gobet, 2000), including domains in which memory recall is a contrived task. That said, given that TT is a process theory, it is ultimately optimal to construct computational models for all of the other domains in order to derive quantitative predictions in these other domains (Vicente, 2000). Indeed, the scope of CHREST is being extended to language acquisition, memory for computer programs, and physics expertise (Gobet et al., 2001).

In sum, despite the wealth of recent theoretical articles on the nature of expert memory (e.g., Ericsson et al., 2000; Simon & Gobet, 2000; Vicente, 2000; Vicente & Wang, 1998), to the best of our knowledge this is the first study to undertake a detailed empirical analysis of the relative merits of the CAH and TT. Our data suggest that a process theory makes a number of correct predictions for expert recall in chess. Thus, although much further testing is required, process theories may be able to provide an adequate theoretical explanation for expert memory in contrived tasks.

References

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.

- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.
- Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 641–653.
- Charness, N. (1981a). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 467–476.
- Charness, N. (1981b). Visual short-term memory and aging in chess players. *Journal of Gerontology*, 36, 615–619.
- Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 1–58). New York: Academic Press.
- Cooke, N. J., Atlas, R. S., Lane, D. M., & Berger, R. C. (1993). The role of high-level knowledge in memory for chess positions. *American Journal of Psychology*, 106, 321–351.
- de Groot, A. D. (1978). *Thought and choice in chess* (G. W. Baylor, Trans.). The Hague, the Netherlands: Mouton Publishers. (Original work published 1946)
- de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess. Studies in the heuristics of the professional eye*. Assen, the Netherlands: Van Gorcum.
- Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, 79(2).
- Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.
- Ericsson, K. A. (Ed.). (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*. Mahwah, NJ: Erlbaum.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.
- Ericsson, K. A., Patel, V. L., & Kintsch, W. (2000). How experts' adaptations to representative task demands account for the expertise effect in memory recall: Comment on Vicente and Wang (1998). *Psychological Review*, 107, 578–592.
- 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* (pp. 253–267). Hillsdale, NJ: Erlbaum.
- Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science*, 8, 305–336.
- Frydman, M., & Lynn, R. (1992). The general intelligence and spatial abilities of gifted young Belgian chess players. *British Journal of Psychology*, 83, 233–235.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gobet, F. (1993). *Les mémoires d'un joueur d'échecs* [Chess players' memories]. Fribourg, Switzerland: Editions Universitaires.
- Gobet, F. (1997). Roles of pattern recognition and search in expert problem solving. *Thinking and Reasoning*, 3, 291–313.
- Gobet, F. (1998). Expert memory: Comparison of four theories. *Cognition*, 66, 115–152.
- 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. H. M. Uiterwijk (Eds.), *Advances in Computer Chess 7* (pp. 398–403). Maastricht, the Netherlands: University of Limburg Press.
- Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5, 236–243.
- 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. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, 6, 225–255.
- Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, 24, 651–682.
- Grant, D. A. (1962). Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychology Review*, 69, 54–61.
- Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Erlbaum.
- Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34, 103–115.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- 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.
- Richman, H. B., & Simon, H. A. (1989). Context effects in letter perception: Comparison of two theories. *Psychological Review*, 96, 417–432.
- Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory using EPAM IV. *Psychological Review*, 102, 305–330.
- Schulz, R., & Salthouse, T. A. (1999). *Adult development and aging: Myths and emerging realities* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Sciences*, 3, 430–435.
- Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, 61, 394–403.
- Simon, H. A., & Gilmarin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, 5, 29–46.
- Simon, H. A., & Gobet, F. (2000). Expertise effects in memory recall: Comment on Vicente and Wang (1998). *Psychological Review*, 107, 593–600.
- Vicente, K. J. (2000). Revisiting the Constraint Attunement Hypothesis: A reply to Ericsson, Patel, and Kintsch (2000), and Simon and Gobet (2000). *Psychological Review*, 107, 601–608.
- Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, 105, 33–57.
- Waters, A. J., Gobet, F., & Leyden, G. (2002). Visuo-spatial abilities in chess players. *British Journal of Psychology*, 93, 557–565.

(Appendix follows)

Appendix A

Methods Used to Generate the Various Types of Random Positions

1. Random positions: The pieces of a game position are randomly assigned to an empty square in the new position.
2. Truly random positions: Not only the location but also the type of piece are randomized. The following algorithm is used.
 For the desired number of pieces:
 1. pick up an empty square randomly with equal probability in the new position;
 2. pick up a type of piece, with equal probability, from the set of 12 black or white types of pieces; and
 3. assign that piece to that square.
3. Truly random 1/3 : With equal probability, randomly keep two thirds of the pieces on their location. Using the algorithm in #2, truly randomize the other one third of the pieces.
4. Truly random 2/3 : With equal probability, randomly keep one third of the pieces on their location. Using the algorithm in #2, truly randomize the other two thirds of the pieces.

Appendix B

Positions Used in the Truly Random Condition

The positions are given in Forsyth notation. In this notation, the board is scanned rank by rank from a8 to h8, then from a7 to h7, and so on to h1. Letters indicate pieces (uppercase for white, lowercase for black). Digits refer to the number of empty squares between pieces or the board's boundaries. Ranks are separated by slashes:

Position 1: Pb2bBqP/2B1k3/1pRK2pB/1Q2K2b/R1p3q1/k5r1/P2Q2b1/5p1p/

Position 2: 4bqP1/8/r2kkKrK/1r4b1/1qNrn2N/2b1Bb2/Q5Bn/1r2n1kP/

Position 3: 1n1Q1RQ1/4P1q1/1n1Kk2r/qN4P1/Q7/B5Q1/4bNpK/N1R2p1Q/

Position 4: b1q5/1k1q4/2BQ1k1P/5K2/3p3b/2nN1qKK/k3B2N/1B2pPrB/

Position 5: qQ1QR1nn/K1pr4/1n1R1k2/6q1/7n/1BN5/K3PpBq/b3qBr1/

The composition of the truly random stimuli is given in Table A1. One should note that the frequency of each piece type is reasonably close to the expected frequency, which is 2.08 (25 pieces/12 types). The table also indicates the expected frequency for each piece in master game positions, computed from a sample of 10,000 positions after 20 moves.

Table A1
Frequency Distribution of White and Black Pieces for the Five Truly Random Stimuli Used and for a Sample of 10,000 Master Games After 20 Moves

Position	White							Black							Total
	K	Q	R	B	N	P	Sum	k	q	r	b	n	p	Sum	
Truly random stimuli															
1	2	2	2	3	0	3	12	2	2	1	4	0	5	14	26
2	2	1	0	2	2	2	9	3	2	5	4	3	0	17	26
3	2	5	2	1	3	2	15	1	2	1	1	2	2	9	24
4	3	1	0	4	2	2	12	3	3	1	2	1	2	12	24
5	2	2	2	3	1	1	11	1	4	2	1	4	2	14	25
<i>M</i>	2.2	2.2	1.2	2.6	1.6	2.0	11.8	2.0	2.6	2.0	2.4	2.0	2.2	13.2	25.0
Master-level games															
<i>M</i>	1.0	0.8	1.9	1.3	1.1	6.1	12.3	1.0	0.8	1.9	1.3	1.2	6.2	12.4	24.7

Note. White pieces are indicated by uppercase letters; black pieces are indicated by lowercase letters.

Received August 30, 2002
 Revision received June 5, 2003
 Accepted June 8, 2003 ■