THEORETICAL NOTES

How Experts' Adaptations to Representative Task Demands Account for the Expertise Effect in Memory Recall: Comment on Vicente and Wang (1998)

K. Anders Ericsson
Florida State University

Walter Kintsch
University of Colorado at Boulder

K. A. Ericsson and W. Kintsch's (1995) theoretical framework of long-term working memory (LTWM) accounts for how experts acquire encoding and retrieval mechanisms to adapt to real-time demands of working memory during representative interactions with their natural environments. The transfer of the same LTWM mechanisms is shown to account for the expertise effect in unrepresentative "contrived" memory tests. Therefore, K. J. Vicente and J. H. Wang's (1998) critique of the generalizability of the LTWM framework is rejected. Their proposed refutation of LTWM accounts is found to be based on misrepresented facts. The process-based framework of LTWM is shown to be superior to their product theory because it can explain interactions of the expertise effect in "contrived" recall under several testing conditions differing in presentation rate, instructions, and memory procedures.

A few years ago, two of us (Ericsson & Kintsch, 1995) proposed in this journal a theoretical framework for how individuals could acquire skills to maintain access to relevant information in long-term working memory (LTWM) during comprehension of text and a wide range of different types of expert performance. We considered our research to be consistent with pioneering efforts in ecological psychology, as Vicente and Wang's (1998) quote eloquently put it:

Skill acquisition consists of changing what one attends to, the goal being to identify diagnostic high-order information that can be used to satisfy task goals. Training of attention is accomplished by abstraction, filtering, and optimization of perceptual search (see E. J. Gibson, 1969, 1991, for more details). (p. 36)

We attributed the experts' ability to consistently select more appropriate actions than novices in similar situations to their superior ability to identify and maintain access to relevant information. Consequently, we were very disappointed that Vicente and Wang's recent article on "expertise effects in memory recall" argued that our framework was insufficient to explain experts' superior memory performance for "natural" stimuli from their domains. In this comment, we show that our original framework can readily explain all of the valid empirical evidence on expertise cited by Vicente and Wang in support for their new framework and against accounts based on the LTWM. Given that their argument against the generalizability of LTWM accounts is rather complex and does not agree with our understanding of the proposed mechanisms of LTWM, we first describe their conception of superior expert memory and summarize their criticisms of accounts based on LTWM before turning to our rebuttal of those claims in the remaining three sections of this article.

Vicente and Wang's (1998) Main Proposal and the Claimed Inadequacy of Long-Term Working Memory

Vicente and Wang (1998) distinguished two different forms of experts' superior memory. The first type of memory superiority is very distinctive; it could not possibly be exhibited by nonexperts because of their memory limitations for maintaining information. Consequently, we were very disappointed that Vicente and Wang's recent article on "expertise effects in memory recall" argued that our framework was insufficient to explain experts' superior memory performance for "natural" stimuli from their domains. In this comment, we show that our original framework can readily explain all of the valid empirical evidence on expertise cited by Vicente and Wang in support for their new framework and against accounts based on the LTWM. Given that their argument against the generalizability of LTWM accounts is rather complex and does not agree with our understanding of the proposed mechanisms of LTWM, we first describe their conception of superior expert memory and summarize their criticisms of accounts based on LTWM before turning to our rebuttal of those claims in the remaining three sections of this article.

1 Vicente and Wang (1998, p. 35) raised the problem of finding a suitable set of verbal labels. To minimize further problems of communication, we use quotation marks in this article when we use concepts and labels in the manner proposed by Vicente and Wang, such as "experts," "contrived" memory task, and "natural" materials.
For example, chess players could not play chess blindfolded unless they could keep the current chess position accessible in working memory (WM), and mental calculators could not multiply 32,759 and 83,793 "in their heads" without an exceptional ability to keep the intermediate numerical products in WM. As long as the demands for superior memory are representative and "intrinsic" and reflect "a definite feature of that domain of expertise," Vicente and Wang (1998, p. 34) accept Ericsson and Kintsch’s (1995) account of memory phenomena in terms of LTWM. However, Vicente and Wang do not think that LTWM can account for a second type of experts' superior memory that seems to be independent of any explicit need for superior memory to perform at a superior level. The best evidence for this second type of memory superiority comes from laboratory studies in which experimenters have tested experts' and novices' memory for stimuli that are representative of natural objects and situations in their domain of expertise, that is, "natural" stimuli. For example, chess masters are able to recall presented chess positions far better than less skilled players (Chase & Simon, 1973; de Groot, 1946/1978). Given that chess players are evaluated only by their ability to play chess games—not their memory for chess positions—the experts' superiority cannot be explained by their intentional efforts to improve that aspect of their performance. In fact, most of the experts would never be given these unexpected and "contrived" tests of their memory. Vicente and Wang (1998) argued that no existing theoretical account, including Ericsson and Kintsch’s (1995) LTWM, can adequately explain these expertise effects in memory recall when the memory tests are "contrived," that is, not occurring naturally in the domain, and the presented material includes "natural" stimuli as well as various degrees of randomly rearranged versions of such stimuli.

According to Vicente and Wang’s (1998) proposal, experts must extract the structure of "natural" stimuli in their domain. With increased experience, individuals discover or attune to more and more constraints of "natural" stimuli during representative activities in the domain, and experts become able to rely on these constraints during the presentation and subsequent recall of the "natural" material. Vicente and Wang’s constraint attunement hypothesis also explains why experts’ recall is reduced to the novice level when randomly rearranged versions of the "natural" stimuli are presented, because the random stimuli are not limited by the constraints of the "natural" stimuli when they occur in a domain-representative context. However, Vicente and Wang claimed not to be ready to fully explicate the processes and detailed mechanisms involved in the encoding and recall of "natural" stimuli or how the specific constraints are learned through experience. Instead, they proposed a product theory of expertise effects in memory recall that predicts when the expertise effects will be observed and the relative magnitude of the effects across experimental conditions in "contrived" memory tests. By explicitly stating that their hypothesis is not a process theory, they insulated themselves from criticism regarding any issues concerning mechanisms that mediate this type of superior expert memory performance. However, they still criticized the mechanisms specified by other competing theories.

Vicente and Wang (1998) not only argued that Ericsson and Kintsch (1995) did not give an account of "expertise effects in domains in which memory recall is a contrived, rather than an intrinsic, task" (pp. 47–48) but also claimed that the framework of LTWM could not do so because of inherent theoretical limitations, such as its "need for deliberate memory enhancement" (p. 34) and its inability to account for "the influence of the environment" (p. 48). They also made additional, highly specific criticisms, assuming that our LTWM framework can provide plausible accounts for the expertise effect in most domains. For example, Vicente and Wang proposed that "LTWM cannot account for the lack of expertise effects in several memory recall studies in the domain of medicine" (p. 48) and that the result of one study (Coughlin & Patel, 1987) "explicitly contradicts the recent LTWM theory" (p. 44). The latter result takes on particular significance because it allegedly provides evidence—the only explicitly stated instance that we could uncover in their article—for their central claim that "a product theory . . . can make a unique contribution by guiding the search for a viable process theory of the same phenomenon" (p. 48).

Vicente and Wang (1998) explicitly based their rejection of accounts based on LTWM on their difficulties in understanding how the mechanisms of LTWM could possibly explain the emergence of superiority of experts’ memory for "contrived" memory tests. Their difficulties appear to be linked to a conception of LTWM defined merely as the deliberate acquisition of mnemonic encoding methods and retrieval structures in the manner proposed by Chase and Ericsson (1981, 1982). In support of that interpretation, Vicente and Wang limited their discussion of LTWM to expertise where memorizing is "a distinctive feature of that domain" (p. 34). For example, they made explicit reference to a waiter’s memory skill for dinner orders (Ericsson & Polson, 1988a, 1988b). In that particular case, the resulting memory skill is an independent addition to the natural skill of waiting and relies on deliberately acquired mnemonic encoding methods. Furthermore, the waiter, in that particular case, is only trying to remember dinner orders and is only required to reproduce the literal information from memory. Consequently, the internal mnemonic encoding does not need to represent the intrinsic semantic structure of the "natural" material in the domain and thus does not reflect the structure in the environment or "active attunement to that structure," to quote Vicente and Wang (1998, p. 48).

Whereas Vicente and Wang (1998) seemed to view the structure of the waiter’s memory as representing the full scope of LTWM, we view that account as a relatively simple extension of skilled memory theory in the manner originally proposed by Ericsson and Polson (1988a, 1988b). Ericsson and Kintsch’s (1995) proposal for LTWM was designed to go beyond skilled memory theory, to explain WM in skilled and expert performance (see in particular pp. 217–222). It is notable that Vicente and Wang never explicitly considered and discussed the new fundamental ideas that define LTWM as a unique contribution. In fact, one of the new central ideas of LTWM is that the same basic mechanisms governing accessibility of encoded information in long-term memory seen in exceptional memory skills can be generalized to skilled and expert performance in which the generated internal encodings preserve the semantic structure and the representation of the environmental constraints in WM. The primary challenge for WM in skilled and expert performance is to maintain access to information so that it can be efficiently retrieved whenever relevant.

It is necessary to keep large amounts of information accessible in WM during text comprehension and expert performance in domains such as chess and medicine. For example, to successfully comprehend a particular sentence within a text, the reader must
maintain access to all relevant aspects of the information presented earlier in the text. Similarly, a medical doctor needs to maintain access to a lot of information about a patient to generate the correct diagnosis. Finally, during tournament games, chess masters almost invariably plan and mentally explore consequences of long sequences of chess moves as they search for the best move for a position. More generally, experts in all sorts of domains mentally reason and evaluate their plans of action as part of their expert performance, thus requiring access to a lot of information in WM.

According to the theoretical framework of LTWM, an essential aspect of expert performance is the ability to anticipate which information might be needed in the future and to maintain access to that relevant information. To maintain access, expert performers need to develop skills that allow them to encode information in long-term memory during representative activities such that the same information can be easily retrieved whenever it is relevant. The ability to anticipate future retrieval demands for information is closely linked to the development of knowledge of the domain and to acquisition of the methods and skills mediating superior performance in the domain. Consequently, as Ericsson and Kintsch (1995) proposed, the encoding skills that support efficient storage and reliable retrieval in LTWM are tightly integrated aspects of the performance of experts, and these skills must be acquired in close interaction with other aspects of their performance. The tight interaction between memory and performance makes it difficult to experimentally study the memory mechanisms.

Ericsson and Kintsch (1995) argued that the same type of memory mechanisms mediate the comprehension of texts, chess, and medical diagnosis. An extended part of their article (pp. 222–232) reviews how the comprehension of text requires encoding of previously read text in such a way that relevant information can be effortlessly retrieved when needed later. They proposed that the construction of new semantically integrated structures in the form of situation models explicates an LTWM mechanism for coordinating encoding and retrieval of information from the text (see Kintsch, 1998, in press; Kintsch, Patel, & Ericsson, 1999, for more recent and comprehensive accounts). In a well-written text, the author facilitates the integration of new sentences with earlier presented information through the organization of the text and the use of explicit and implicit references. For example, when a skilled reader encounters the words the tall policeman in a sentence, then these words provide a semantically based retrieval cue to relevant information about the associated character in the text that the reader generated during the prior reading of the text. Similarly, semantically based access occurs in medical diagnosis when a new symptom is reported to a medical expert by a patient. The expert’s encoding of the symptom provides retrieval cues for accessing other relevant encodings of the patient’s condition that the medical expert had generated earlier in the interview. Unlike the case of the well-written text in which the author guides the reader’s encoding and integration, the medical expert must actively anticipate the future potential relevance of the information to ensure its reliable access whenever relevant.

It is almost inconceivable that memory could be deliberately improved independently of domain-related performance, because memory and task-related processing are so tightly integrated during text comprehension and expert performance. Any useful improvements in an expert’s memory would be an indirect consequence of deliberate changes to mechanisms mediating representative performance in these cases. For example, chess players often study published chess games between chess masters by trying to predict the “correct” move (the actual move selected by the chess master making the next move). Whenever players select a different move, they mentally plan out the consequences of alternative moves to discover which aspect or relation they should have considered to select the “correct” move (Ericsson, Krampe, & Tesch-Römer, 1993). Aspiring chess players can spend as much as 4 hr per day engaged in this type of study (Charness, Krampe, & Mayr, 1996); consequently, their ability to mentally represent chess positions becomes so good that it is even possible for them to play a single game of blindfolded chess without prior specific practice (Ericsson & Oliver, as cited in Ericsson & Staszewski, 1989; Saariluoma, 1989). More generally, Ericsson and Kintsch (1995) argued that “superior working memory capacity reflects a domain-specific memory skill acquired to meet specific demands of working memory” (p. 233). Experts deliberately try to enhance their representative performance, such as their ability to select superior chess moves. The improvement of memory is an indirect consequence.

In the remainder of this article, we rebut Vicente and Wang’s (1998) criticisms, showing that Ericsson and Kintsch’s (1995) original theoretical framework for LTWM can explain Vicente and Wang’s second type of superior expert memory revealed by “contrived” memory tasks. We demonstrate that the LTWM framework offers a deeper understanding of experts’ memory than does Vicente and Wang’s product theory. The discussion is divided into three parts. First, we show how a rigorous definition of expertise based on reliably superior performance allows us to identify individuals who have acquired superior memory performance to meet the demands of representative tasks in their domain. This analysis provides an account of why some “experts” do not display superior recall for “natural” stimuli when compared with less experienced individuals—an issue that Vicente and Wang (1998, p. 47) conceded could not be accounted for by their proposed product theory. This section also explains the methodological problems associated with studying performance on unrepresentative and unfamiliar tasks, such as “contrived” memory tasks, and why Ericsson and Kintsch (1995) intentionally avoided a detailed discussion of those empirical phenomena, explicitly restricting their review to WM in “well-defined tasks that capture the expert performance in each of these domains (Ericsson & Smith, 1991)” (p. 233).

In the second section, we rebut Vicente and Wang’s (1998) criticism by demonstrating in considerable detail how the LTWM framework can be successfully extended to explain the mechanisms mediating performance on “contrived” memory tests. For each type of “contrived” memory task, such as unexpected recall of a representative task, and intentional memorization of “natural” stimuli, we identify the memory-related mechanisms that mediate superior “contrived” memory performance and specify their relation to mechanisms mediating experts’ superior representative performance, such as selecting chess moves. Only by examining the degree to which “contrived” memory tasks match real-time task demands for experts’ representative activities can accurate predictions be made for superior “contrived” memory performance. Although Ericsson and Kintsch (1995) did not describe these arguments in much detail, they cited related accounts (Ericsson & Pennington, 1993; Ericsson & Smith, 1991), and Ericsson
has addressed them elsewhere (Ericsson & Delaney, 1999; Ericsson & Lehmann, 1996).

In the final section of this article, we show how the expertise effect in memory recall can be accounted for by using the experts’ LTWM acquired for the purpose of superior representative performance in chess and medicine. In our discussion of the studies featured by Vicente and Wang (1998), we provide a rebuttal of their most specific criticisms of the LTWM framework in the domain of medicine. We also review related findings from chess and medicine that raise issues about the generalizability of their product theory.

The Empirical Phenomenon of the Expertise Effect for Memory Retrieval and Its Relation to Superior Representative Performance

Any time that scientists from different disciplines discuss evidence collected by investigators with different theoretical backgrounds, such as cognitive and ecological psychology, it is essential for them to reach consensus about the definition and general characteristics of the studied empirical phenomenon before a meaningful exchange of ideas can occur.

Toward a Definition of the Empirical Phenomenon

Vicente and Wang’s (1998, p. 35) product theory was explicitly designed to answer the question “Under what conditions will there be an expertise advantage?” and thus to specify when “experts” will show better recall than “novices” for “natural” and randomly rearranged materials from their domain of expertise. Unfortunately, their product theory is difficult to evaluate because Vicente and Wang did not clearly anchor their theoretical concepts empirically. They did not address the current controversies surrounding the problem of identifying “experts” and failed to offer unambiguous definitions of the terms expert, novice, beginner, and natural material. However, these problems are not just Vicente and Wang’s; they were shared by most research on expertise during the 1970s and 1980s.

In the past couple of decades, expertise researchers’ definitions of expert have ranged from world-class athletes and chess players with vastly superior performance to college students who have taken a couple of courses in a subject. Some investigators even used the term experts to refer to regular school children with a greater than average knowledge of various sports (Schneider, Körkel, & Weinert, 1989). In response to the daunting problems of developing theories and predictions for this heterogeneous group of “experts,” Ericsson and Smith (1991) proposed focusing on an empirically well-defined group of “experts” (expert performers), who were able to exhibit consistently superior performance in activities that defined expertise in their domain. Because the traditional domains of expertise, such as chess and medicine, meet this criterion of superior performance, Ericsson and Kintsch (1995) adopted this definition for their review of superior memory in expert performance.

In contrast, Vicente and Wang’s (1998) review did not specify any verifiable criteria for “experts.” The “experts” in their prominently featured study on process control (Vicente, 1992) were engineering graduate students whose only direct experience of the complex system was restricted to an hour-long general introduction given to all participants before the start of the experiment. The “novices” were other graduate students who “had never enrolled in a science or engineering major” (Vicente, 1992, p. 363). Vicente (1992) reported that “the novices did not perform significantly differently from the experts” (p. 365) on the main task of diagnosing the state of the system. Furthermore, both “experts” and “novices” had major difficulties even distinguishing “random” states of the system from the “natural” states. Consequently, Ericsson and Kintsch (1995) did not cite Vicente’s (1992) findings in their review of WM of expert performers who exhibited consistently superior performance on representative tasks and were very accurate in distinguishing between “random” and “natural” stimuli from their domain of expertise.

Within Ericsson and Kintsch’s (1995) LTWM framework, the superior recall of Vicente’s (1992) “experts” over his “novices” for the presented states of the system would be attributed to individual differences in domain-specific knowledge. However, Vicente’s (1992) procedure for testing recall by “experts” and “novices” was innovative and different from traditional recall of text in which accurate recall is defined by recall of specific ideas and particular propositions. Vicente’s participants were required to give a numerical estimate for each of 34 different process variables by marking a value on a number line for each variable after presentation of a complex state lasting only 25 to 30 s. Not surprisingly, memory was far from perfect for both groups. As we argue in the concluding section of this article, the LTWM framework would

---

2 The average proportion of correct decisions can be extracted from Figure 5 in Vicente (1992). When participants were asked to judge the state as “random” or “semantic” (“natural”), the accuracy for both “experts” and “novices” was less than 60% and less than 80% for the two display conditions, respectively, with 50% expected from pure guessing in both cases. In a couple of post hoc analyses using nonparametric tests, Vicente coded for which of the two display conditions each participant generated more accurate diagnoses or, alternatively, whether the diagnostic performance was equally good (tie). The accuracy of the “experts” was higher for the advanced display than the other display type, when making more detailed diagnoses about “natural” states, whereas the accuracy of the “novices” remained closer to the same level for both types of displays. These interesting analyses, however, did not provide evidence for a reliable accuracy difference between the “experts” and “novices” in regard to their ability to distinguish “random” and “natural” states.

3 Vicente (1992) calculated an error score “by subtracting the subjects’ recall from the actual state of the variable, normalizing this difference with respect to the maximum scale value for each variable, and then taking the absolute value” (p. 364). The least accurate reproductions were provided by the “novices” for the “random” states. Under the assumption that the recall deviations have a rectangular distribution, the reproduction of the “novices” would be within ~20% and 20% of the correct value half of the time, and the other half of the time, deviations would be greater (values extracted from Vicente, 1992, Figure 6). Thus, the recall of the “novices” would cover 80% of the total range and corresponded to pure guesses, virtually completely. The most accurate condition corresponded to the recall of the experts for the “semantic” or “natural” states, in which the mean deviation was only around half of the memory performance of the “novices” for the “random” states. This suggests that the “experts” could accurately identify which half or third of the variables’ total range corresponded to the presented value, on average. Vicente and Wang’s (1998) Figure 2 suggests that performance was even more accurate in another condition, but these differences were not found to be statistically reliable in the original article (Vicente, 1992, p. 367).
account for the demonstrated superiority of the “experts’” memory estimates for “natural” states by better guessing constrained by their superior knowledge of the domain and their partial memory for some aspects of the state of the system.

More generally, Vicente and Wang (1998) claimed, in their introduction, that “memory recall performance on meaningful stimuli” and “domain expertise” have “almost always been found to be correlated” (p. 33), and they cited evidence from 51 studies from 19 different domains. However, many of those studies did not measure “domain expertise”; they simply inferred it from the length of experience in the domain. Recent reviews (Ericsson, 1996; Ericsson et al., 1993; Ericsson & Lehmann, 1996) have demonstrated that the length of professional experience after completed training has often been a weak predictor of performance in representative professional activities, such as medical diagnosis (Norman, Coblentz, Brooks, & Babcock, 1992; Schmidt, Norman, & Bosshuizen, 1990), auditing (Bédard & Chi, 1993; Bonner & Pennington, 1991), text editing (Rosson, 1985), judgment and decision making (Camerer & Johnson, 1991; Shanteau & Stewart, 1992), and therapy in clinical psychology (Dawes, 1994).

Even more directly relevant to Vicente and Wang’s (1998) claimed generalization about the expertise effect in recall, the relation between level of expertise and performance on “contrived” memory tests has been found to be similarly inconsistent. Problems with replicating the expertise effect in recall have not been limited to medicine—the only problematic domain acknowledged by Vicente and Wang—which we discuss in considerable detail later in the section on Medicine. In fact, there are many instances in which experienced individuals’ memory performance for representative stimuli is not superior to that of novices and, in some studies, to that of control participants who lack additional experience in the corresponding domain. Consider the examples of actors who are memorizing text (Intons-Peterson & Smyth, 1987; H. Noice, 1991), computer programmers’ memory of program statements (Adelson, 1984), experienced map users’ memory of regular planimetric maps (Gilhooly, Wood, Kinmear, & Green, 1988; Thordyke & Stasz, 1980), and musicians’ reproductions of presented melodies (Sloboda & Parker, 1985).

**Expertise as Reliably Superior Performance on Representative Tasks**

In response to the aforementioned and other related problems with the concepts of expert and expertise, Ericsson and Smith (1991) resurrected the original approach advocated by de Groot (1946/1978), which focused on the identification and study of captured superior expert performance. By examining the ecological conditions in which the essential aspects of expert performance naturally occur, researchers are able to design similar task conditions to reproduce experts’ superior performance in the laboratory. This type of expert performance represents an integrated, stable adaptation to all relevant constraints of the central task demands attained after years and decades of practice (Ericsson & Lehmann, 1996). Consequently, Ericsson and Kintsch (1995) restricted their review of expertise to studies of expert performers who exhibited a superior level of performance on representative tasks. They restricted their review to the “intrinsic” demands of WM mediating the observed superior performance that assures a close connection between LTWM and performance. However, will the relation between level of expertise and performance on “contrived” memory tasks be similarly tight?

Performance on “contrived” memory tests does not relate closely to expert chess performance. The correlations are quite modest for representative samples of chess players—in the .04 to .50 range, which is considerably less than those of other tests of knowledge and skills related to move selection (Charness, 1981a; Pfau & Murphy, 1988). In a study on the effects of aging, Charness (1981a) found that even participants’ age was a more accurate predictor of contrived memory performance than was expertise (chess rating). De Groot and Gobet (1996) also reported “substantial” (p. 60) individual differences in chess masters’ recall on “contrived” memory tests, including one chess master whose performance could not be distinguished from that of weaker players. Furthermore, there are large practice effects on contrived memory tests. Doll and Mayr (1987) found the recall performance of chess masters averaged less than 57% without prior practice. In contrast, after task familiarization and adoption of appropriate techniques, such as taking a 5- to 10-s break to consolidate the perceived position prior to the initiation of recall, chess masters are expected to be able to recall well above 80% of the presented pieces (de Groot & Gobet, 1996). Similarly, Pfau and Murphy estimated the average recall of chess masters and experts to be greater than 70%.

In sum, Ericsson and Kintsch (1995) deliberately avoided the problems associated with “experts’” contrived memory tests for natural material and confined their review to LTWM mediation of reliably superior performance on representative tasks that capture the essence of the expertise. The captured representative performance is closely related to external objective measures of performance, such as chess ratings, whereas the relation to contrived memory performance is much more modest (Ericsson & Lehmann, 1996). Given that representative tasks have been designed to preserve the relevant ecological constraints, the experts’ performance should reflect the stable normal operation of WM and is unlikely to change during a few hours of additional testing—in contrast to the reported changes in contrived memory tests discussed above.

The LTWM mechanisms derived from an analysis of stable processing during representative tasks can be generalized to account for the same experts’ performance on “contrived” memory tasks. In the next section, we show how performance on different types of “contrived” memory tasks can capitalize on mechanisms acquired in support of performance on representative tasks.

**How Can Accounts of Superior Performance on Representative Tasks Explain the Expertise Effect for “Contrived” Memory Tests?**

Vicente and Wang’s (1998) central dichotomous distinction between “contrived” and “intrinsic” memory tasks is replaced within our LTWM framework by degree of transfer of the specific memory mechanisms acquired for performance on representative
tasks to performance on "contrived" tasks. The general mechanisms mediating LTWM are illustrated in Figure 1, in which the solid arrow marks how experts encode representative stimuli in LTWM to allow cue-based access to relevant information whenever needed during task-relevant processing. Even when participants are given unexpected and delayed ("contrived") recall tests, much of the semantically encoded information in long-term memory remains accessible and can be recalled to reconstruct the presented information even literally (see the dashed lines in Figure 1). We next discuss the circumstances in which the LTWM mechanisms mediating experts’ superior representative performance can be generalized to "contrived" memory tasks created to assess specific types of memory processing.

The closest relation between the cognitive processes mediating superior representative performance and memory performance on "contrived" tasks is found when the experts are unexpectedly instructed to recall presented information immediately after completion of a representative task (cf. Lane & Robertson, 1979; and Charness, 1981a, for examples of recall after selection of the best move for a chess position). In these instances, the participant is instructed about the recall task after the representative task is completed, thus eliminating the possibility that anticipation of having to report could change the encoding of information in memory. The encoding of information and storage in memory could not be termed contrived because the internal encoding of information was spontaneously generated and was thus intrinsic to the representative task. However, the participant has to be able to retrieve the memory of the internal encoding (see Figure 1). This process of recall may qualify as contrived because participants have to search for and extract information requested by the experimenter. As long as the requested information is the same as the participants’ spontaneous thoughts during performance on representative tasks, no changes in the structure of thinking would be expected. This finding would be consistent with the nonreactive nature of thinking aloud and retrospective verbal reports of one’s thoughts (Ericsson & Simon, 1993).

When the information requested in the recall task is spontaneously generated by the expert during the representative task (see Figure 1), one would not expect any effect of advance knowledge of the recall test. Lane and Robertson (1979) found that advance information did not reliably influence memory when chess players were engaged in a move-selection task that involved encoding relevant aspects of an entire chess position (de Groot, 1946/1978). Similar results have been found for medical diagnosis. Think-aloud protocols show that medical doctors normally encode relevant information about patients (Boshuizen & Schmidt, 1992). This evidence leads us to predict that advance knowledge of recall would not influence subsequent recall in studies, such as Coughlin and Patel's (1987), provided that recall is restricted to relevant information and the doctors are allowed to verbalize their higher

---

**Figure 1.** Schematic drawing of representative task and stimuli that allow the reproduction of representative expert performance in the laboratory, along with the associated mediating encoding and retrieval processes (solid lines). Additional processes required for performance on "contrived" memory tasks are illustrated as dashed lines. When the contrived memory task requires mere reproduction, it is possible for participants, including experts, to use a proportion of mnemonic encodings to maximize their memory performance (illustrated as a path bypassing normal semantic encoding).
level encodings of presented information (Schmidt & Boshuizen, 1993). More generally, Ericsson and Lehmann (1996) found that when expert performers were confronted with challenging problems in their domains of expertise, their think-aloud protocols consistently revealed extensive planning and reasoning using complex and highly refined mental representations.

When experts are presented with "natural material" that captures the essence of a typical situation in their domain of expertise, such as a game position for a chess expert or an athlete, it is possible that the experts could encode the presented information in the typical manner even when the experimental task is just a memory task. Experts have revealed superior memory for representative game situations but not for recall of random, unstructured stimuli in sports, such as basketball (Allard, Graham, & Paarsalu, 1980; Starkes, Allard, Lindley, & O'Reilly, 1994), field hockey (Starkes, 1987; Starkes & Deakin, 1984), and volleyball (Bourgeaud & Abernethy, 1987), and in games, such as chess (see Charness, 1991; and Saariluoma, 1995, for reviews), bridge (Charness, 1979; Engle & Bukstel, 1978), Othello (Wolff, Mitchell, & Frey, 1984), and GO (Reitman, 1976). In these cases, one would expect the "contrived" memory performance to be at least as good as that observed when the experimenter instructs the players to perform the representative task and test their incidental memory (see Figure 1). However, proactive interference is a major problem in domains such as sports, where the external environment is rapidly and continuously changing (Ericsson, 1996; Ericsson & Kintsch, 1995).

There are many types of "contrived" memory tasks involving "natural" materials from the domain of expertise in which the required processing does not consistently map onto encoding and retrieval in representative activities. A good example of this type of memory task is memorization of the specific code for a simple computer program by experts and novices (Barfield, 1986). Although a computer expert such as a computer science professor might have to consider even literal aspects of students' simple computer programs to check their accuracy, it is probably rare that such an expert has to memorize and reproduce the exact programming code. Expert performers acquire higher level representations that allow them to plan and reason about "natural" objects, such as computer programs, which may or may not be effective for literal memorization and reproduction of presented information. When we discuss medical diagnosis, we consider this issue in more detail.

"Contrived" memory tasks typically require only literal reproduction of the presented information, which makes them vulnerable to artifacts and practice effects. Hence, it is possible that even skilled participants rely on shortcuts, such as mnemonics, to encode the presented information (see the separate dashed pathway in Figure 1) when the memory tasks require only a literal reproduction. Furthermore, experts are not familiar with "contrived" memory tasks and may therefore change their encoding methods, resulting in significant practice effects similar to those noted in our previous review of memory for chess positions. However, these artifacts can be virtually eliminated by studying experts' memory during and after representative performance, as Ericsson and Kintsch (1995) recommended. Only when experts confront demanding representative tasks can research be assured that they are forced to attend to relevant information in the highly constrained manner that is necessary to generate superior performance.

A Long-Term Working Memory Account of Vicente and Wang's (1998) Featured Studies of "Contrived" Memory Performance

In this section, we provide accounts within the LTWM framework for the results from the "contrived" memory studies featured by Vicente and Wang (1998). We first discuss the results from the domain of chess and then address the findings from medical diagnosis, which Vicente and Wang found problematic for the LTWM framework.

Chess

The first step of our approach involves identifying measures of consistently superior performance in the domain. Since de Groot's (1946/1978) pioneering research, Elo (1986) developed a method for calculating chess-playing ability based on outcomes of games at chess tournaments that can accurately predict the outcome of future matches. In all of the studies on chess featured by Vicente and Wang (1998), level of expertise was assessed by the participants' ratings of chess skill, Ericsson and Kintsch (1995) focused on "the best laboratory task for capturing chess skill," namely, "selection of the next move for an unfamiliar chess position (de Groot, 1946/1978; Ericsson & Smith, 1991)" (p. 233). To select the best move at a critical point in a game, chess players take between 6 min (Charness, 1981b) and 15 min (de Groot, 1946/1978), on average, which is at the upper bound for the time chess players normally require to select a move for a similar position during a tournament. Following the original account of move selection given by de Groot, it is useful to distinguish the initial brief perception of the position from the extended subsequent planning and systematic evaluation of alternatives. The processes mediating the initial perception have been extensively studied with methods ranging from introspection to eye fixations, and they have revealed the remarkable perceptual abilities of chess masters (see de Groot & Gobet, 1996, for a review). The primary contribution by Ericsson and Kintsch consists of a review of the evidence from a long line of studies (starting with Charness, 1976; and Frey & Adesman, 1976) to show that the generated representation was stored in long-term memory in accessible form—LTWM.

If it can be demonstrated that chess players spontaneously generate a representation of a chess position while selecting a move, as shown by the analyses of immediate (de Groot, 1946/1978; de Groot & Gobet, 1996) and delayed unexpected recall (see Charness, 1981a; Goldin, 1978; and Lane & Robertson, 1979, for recognition memory), then we simply posit that the same mechanisms underlie performance during "contrived" memory tests for chess positions. We do not have anything significant to add to current models that explain individual differences in recall of brief...
presentations of regular positions based on differential availability of previously stored patterns, knowledge, and organized structures, such as templates, in long-term memory (Gobet & Simon, 1996b).

The parallels between the integrated structures generated during the comprehension of a chess position and those generated during the comprehension of texts are striking (Ericsson & Kintsch, 1995; Kintsch, 1998). The central question is why would highly skilled chess players encode chess positions spontaneously in long-term memory rather than merely keep the relevant information transiently activated in short-term memory, as Chase and Simon (1973) originally proposed and as appears to be true for many habitual everyday activities, such as typing and driving a car. Encoding of a chess position in LTWM appears to be a necessary prerequisite for planning and evaluation activities following the initial perception that allows players to manipulate positions and evaluate alternative move sequences. During the move-selection task, detailed move planning is found to increase as a function of chess skill at least until the players reach the expert level (Charness, 1981b, 1989; de Groot, 1946/1978). Better planning enables world-class players to uncover better chess moves than those that they retrieved from memory on the basis of their initial perception (de Groot, 1946/1978). Planning also minimizes the frequency of mistakes and failure to consider important aspects of the chess position (Saariuluoma, 1992). As mentioned in our introductory section, it is the "intrinsic" demands for extensive planning that appear to drive the acquisition of LTWM for chess rather than any supplementary memory training. Consequently, the research reviewed by Ericsson and Kintsch on LTWM in chess thus meets the definition of "contrived" memory testing.

Rather than duplicate our published review of LTWM in chess, we illustrate the process-based LTWM account and show how these phenomena present genuine challenges for Vicente and Wang's (1998) product theory. Their theory could possibly explain the small, yet reproducible, advantage of chess masters over novices for brief presentations of randomly rearranged chess positions by slight deviations from complete randomness because the advantage amounts to only two or three chess pieces, or approximately 10% of the presented pieces (Gobet & Simon, 1996a). However, when Saariuluoma (1989, Experiment 1) presented regular and random positions by giving the location of each of the associated pieces, the advantage for chess masters over medium players on random positions was comparable to that for regular positions, roughly 10 pieces (or about 40%) of the presented pieces. The same general finding was replicated in two additional experiments. Furthermore, with an even slower presentation rate (4 s per chess piece), the chess masters and the medium players could recall around 82% and 70%, respectively, of the pieces in random positions (Saariuluoma, 1989, Experiment 2). A similar uniform recall advantage for experts on both randomized and regular positions was found by Lories (1987), who presented a regular and a randomized position with the normal procedure, but with a longer presentation time of 1 min each.

Within the LTWM framework, chess experts can rely on their greater knowledge and more refined representations to encode even random combinations of chess pieces by generating complex transformations and partial matches to familiar patterns. There are well-known examples of these types of encodings (Montague, 1972): memorizing nonsense syllables by their relations to real words, such as KOZ by cozy and PYM by payment (Prytulak, 1971). However, reliable formation of these complex encodings requires extended processing with sufficient presentation time for experts to be able to tap their greater knowledge and show large recall advantages for randomized materials. With sufficiently fast presentation rates of random stimuli, such as arbitrary digits and nonsense syllables, individual differences in recall are dramatically reduced (Ericsson, 1985). Whereas most process theories, such as Gobet and Simon's (1996b) and Ericsson and Kintsch's (1995), allow for different encoding processes to become effective at sufficiently slow rates of presentation, there appears to be no simple way for Vicente and Wang's (1998) product theory to explain the interaction between the magnitude of expertise effect and presentation rate.

There is a related finding for the recognition of random chess positions that should be equally problematic for Vicente and Wang's (1998) product theory. It was surprising to us that Vicente and Wang (1998, p. 42) cited the chess study by Goldin (1979) in support of their argument that recognition tests replicate findings of recall tests outside the domain of medicine, because this particular study uncovered a pattern of results that seems to challenge the constraint attunement hypothesis. Goldin found that recognition memory for both random and regular chessboards increased with chess skill but that there was no interaction with the amount of structure. The recognition results from Goldin were obtained with self-paced presentation and may be constrained by potential ceiling effects. In a subsequent study with brief (8-s) presentation times, Saariuluoma (1984) replicated the finding of a uniform improvement in recognition for both random and regular positions without any main effect or interaction with the amount of structure (see Figure 2). Within our framework, the most plausible explanation for this interaction with type of memory test is that chess players adapt their encoding of the regular and random positions to optimize their performance on the particular types of memory test within the available presentation time. This account is consistent with retrospective comments by Goldin's (1979) participants in her study of recognition, with a follow-up experiment by Saariuluoma (1984) in which he varied the structure of the random position, and with laboratory research comparing recognition and recall tests (Hunt & Einstein, 1981).

Finally, the line of research in chess that most closely has explored perception and the attunement to natural constraints has asked chess players to make judgments about chess positions, such as "How many bishops and knights are there?" or "Is the king in check?" Chess experts are uniformly faster than less skilled players in accurately making these judgments for both regular and random positions (see Saariuluoma, 1995, for a review). Somewhat surprisingly, the speed advantage for regular positions over random positions is quite small at all levels of skill and does not show the expected interaction with skill level. The lack of interaction between chess skill and random versus regular structure for perceptual processing of chess positions (Saariuluoma, 1984, 1985) might be explained by the lack of correspondence between the "contrived" perceptual tasks and the representative conditions for playing chess and studying chess positions. In the representative 6The position was taken from Holding and Reynold's (1982) study, in which certain constraints had been imposed on the randomization to avoid illegal relations between chess pieces.
From bars) and random (white bars) chess positions for chess players at four levels of chess skill. The data are from Saariluoma (1984, Experiment 6). From Commentationes Scientiarum Socialium (Vol. 23, p. 63), by P. Saariluoma, 1984, Societas Scientiarum: Fennica, Turku, Finland. Copyright 1984 by author. Adapted with permission.

**Figure 2.** Percentage of correct recognition of presented regular (black bars) and random (white bars) chess positions for chess players at four levels of chess skill. The data are from Saariluoma (1984, Experiment 6).

Ericsson and Kintsch (1995) considered only the statistically reliable findings of Coughlin and Patel’s (1987) study, as it customary in most areas of science. Experts were better able than medical students to identify the correct diagnosis, but accuracy of diagnosis did not reliably interact with the presentation format. When experts later recalled information about the case, they relied on the standard categories of the regular descriptions even for the scrambled descriptions (Groen & Patel, 1988), whereas the medical students tended to preserve the order in which information was presented for both scrambled and regular versions (Coughlin & Patel, 1987). The latter findings, combined with the superior recall of the experts, provide important evidence for experts' ability to encode and organize presented information relevant in LTWM, even when the order of presentation does not match the standard format. Ericsson and Kintsch's proposal for how LTWM allows flexible access to information about patients is based on a proposal by Claessen and Boshuizen (1985), which was later refined by Groen and Patel as well as Patel et al. (1994). According to this proposal, medical experts acquire a schemalike structure to encode information about patients in terms of higher level diagnostic facts that remain invariant across different diagnostic alternatives, that is, disease states.

Vicente and Wang’s (1998) claim that LTWM theory cannot account for this result is curious. They appear unwilling to accept that Coughlin and Patel (1987) did not find a reliable interaction between expertise and format. In fact, the original reviewers of the article were also surprised by this result. They asked Coughlin and Patel to provide a detailed description of their results and to speculate on some potential trends and differences between the two cases in spite of the absence of statistically reliable differences between text formats. In one of the two sample cases (acute endocarditis), a correct diagnosis could not be made unless participants determined the sequence of events leading up to the acute symptoms. Scrambling the sequence of sentences describing the actual events forced the reader to mentally reconstruct the actual sequence. Vicente and Wang saw supporting evidence for their viewpoint in one of the two recall measures for relevant information reported by Coughlin and Patel (1987), namely, the number of recalled information units, "that corresponded exactly to the information in the original text" (p. 822). This variable showed a statistically unreliable pattern of results that was consistent with Vicente and Wang's constraint attunement hypothesis. We have reproduced this pattern in Figure 3A. On the basis of this nonsignificant finding, Vicente and Wang (1998) proclaimed that "it explicitly contradicts the recent LTWM theory of Ericsson and Kintsch (1995)" (p. 44). This is the only explicit contradiction between LTWM theory and empirical evidence cited in their review.

It is rare in scientific psychology to refute a theory on the basis of a finding that does not even meet the minimum criterion of statistical reliability. Furthermore, we offer two independent refutations of their claim. First, more careful examination of Coughlin and Patel's (1987) results reveals an opposite pattern—still not reliable—for the other recall measure, namely, the number of

**Medical**

Medical diagnosis has been successfully simulated in the laboratory by presenting medical doctors and students with information about specific patients and then comparing their diagnoses with the correct diagnoses. The most frequent type of diagnosis requires medical experts to review and integrate a large body of evidence, such as patients' descriptions of their symptoms and results from physical examinations, as well as data from laboratory tests and other ancillary procedures, before final diagnoses are made. Although the accuracy of medical diagnosis for frequent and representative medical problems typically reaches a stable level after the completion of residency, the most experienced experts demonstrate superior diagnosis of more difficult and infrequent cases (Norman et al., 1992; Schmidt et al., 1990). Patel and her colleagues (Patel, Arocha, & Kaufman, 1994; Patel & Groen, 1991) have reported reliably superior performance associated with higher levels of medical expertise.

Coughlin and Patel's (1987) study of recall of descriptions of medical patients. Vicente and Wang (1998) selected a result from a study by Coughlin and Patel (1987) that they felt "LTWM cannot account for" (p. 48). In this study, second-year medical students (novices) and family medicine practitioners (experts) were presented with two patient problem cases. Each case was presented in two formats: a standard format with a familiar patient narrative structure and a random format in which sentence order had been scrambled. The standard format included information in the following sequence: patient's personal data, medical history, physical findings, and laboratory findings (Coughlin & Patel, 1987, p. 821).

Ericsson and Kintsch (1995) considered only the statistically reliable findings of Coughlin and Patel's (1987) study, as it customary in most areas of science. Experts were better able than medical students to identify the correct diagnosis, but accuracy of diagnosis did not reliably interact with the presentation format. When experts later recalled information about the case, they relied on the standard categories of the regular descriptions even for the scrambled descriptions (Groen & Patel, 1988), whereas the medical students tended to preserve the order in which information was presented for both scrambled and regular versions (Coughlin & Patel, 1987). The latter findings, combined with the superior recall of the experts, provide important evidence for experts' ability to encode and organize presented information relevant in LTWM, even when the order of presentation does not match the standard format. Ericsson and Kintsch's proposal for how LTWM allows flexible access to information about patients is based on a proposal by Claessen and Boshuizen (1985), which was later refined by Groen and Patel as well as Patel et al. (1994). According to this proposal, medical experts acquire a schemalike structure to encode information about patients in terms of higher level diagnostic facts that remain invariant across different diagnostic alternatives, that is, disease states.

Vicente and Wang's (1998) claim that LTWM theory cannot account for this result is curious. They appear unwilling to accept that Coughlin and Patel (1987) did not find a reliable interaction between expertise and format. In fact, the original reviewers of the article were also surprised by this result. They asked Coughlin and Patel to provide a detailed description of their results and to speculate on some potential trends and differences between the two cases in spite of the absence of statistically reliable differences between text formats. In one of the two sample cases (acute endocarditis), a correct diagnosis could not be made unless participants determined the sequence of events leading up to the acute symptoms. Scrambling the sequence of sentences describing the actual events forced the reader to mentally reconstruct the actual sequence. Vicente and Wang saw supporting evidence for their viewpoint in one of the two recall measures for relevant information reported by Coughlin and Patel (1987), namely, the number of recalled information units, "that corresponded exactly to the information in the original text" (p. 822). This variable showed a statistically unreliable pattern of results that was consistent with Vicente and Wang's constraint attunement hypothesis. We have reproduced this pattern in Figure 3A. On the basis of this nonsignificant finding, Vicente and Wang (1998) proclaimed that "it explicitly contradicts the recent LTWM theory of Ericsson and Kintsch (1995)" (p. 44). This is the only explicit contradiction between LTWM theory and empirical evidence cited in their review.

It is rare in scientific psychology to refute a theory on the basis of a finding that does not even meet the minimum criterion of statistical reliability. Furthermore, we offer two independent refutations of their claim. First, more careful examination of Coughlin and Patel's (1987) results reveals an opposite pattern—still not reliable—for the other recall measure, namely, the number of
endocarditis

Type of case

70 60 50 40 30 20 10 0

endocarditis arteritis

Figure 3. Mean percentage of relevant “information units” recalled by physicians for regular (black bars) and scrambled (white bars) versions of text for an endocarditis case and a temporal arteritis case from Coughlin and Patel (1987). Panel A shows relevant “information units” recalled exactly—the only data reported by Vicente and Wang (1998). Panel B shows relevant “information units” recalled with changes or interpretations. Panel C shows total “information units” recalled correctly. From “Processing of Critical Information by Physicians and Medical Students,” by L. D. Coughlin and V. L. Patel, 1987, Journal of Medical Education, 62, pp. 825–826. Copyright 1987 by Vimla L. Patel. Adapted with permission.

Second, a further problem with Vicente and Wang’s (1998) argument is that Ericsson and Kintsch (1995) would not have made the counterintuitive prediction that scrambling sentences describing the critical order of a sequence of events would have no impact on the comprehension of the text describing the case or on subsequent recall. Ericsson and Kintsch proposed that experts are able to acquire LTWM mechanisms that allow them to encode independent units of information when these units of information are presented in varied orders in the “natural” environment. However, an independent unit of information would not necessarily correspond to a single sentence, and when it corresponds to several consecutive sentences, then rearranging the order of sentences will lead to additional problems for the reader.

In conclusion, the most striking finding of Coughlin and Patel’s (1987) study was clearly medical experts’ competence at successfully integrating randomly arranged independent units of informa-
tion in LTWM without reliable decrements in diagnostic accuracy and recall. Similarly, Norman, Brooks, and Allen (1989) found that experts were able to encode independent units of information from lab tests in many different presentation orders without any reliable or dramatic decrement in memory performance. Both of these findings seem to be inconsistent with Vicente and Wang's (1998) exploitation of their constraint attenuation hypothesis, rather than supporting it.

Lack of superior memory by medical experts on some memory tasks. Vicente and Wang (1998) criticized the theoretical framework of LTWM for its inability to explain "the lack of expertise effects in several memory recall studies in the domain of medicine" (p. 48). Vicente and Wang (1998, p. 46) proposed that experts encode and recall the "gist" and the high-order relations of an encountered stimulus and that in domains such as medicine, they are unable "to reconstruct the lower level details," which leads to an unrepresentatively low recall score. This account does not differ significantly from Ericsson and Kintsch’s (1995) published account: "The increased selectivity and frequency of abstract encoding of medical information by experts leads to a lower level of recall than for the intermediate participants when the number of all recalled pieces of presented information is counted" (p. 236). Our proposed application of LTWM mechanisms to medical diagnosis relies on higher level concepts (Schmidt & Boshuizen, 1993) or facets (Patel et al., 1994) "that can be induced from data on patients and allow for more effective reasoning about medical diagnosis" (Ericsson & Kintsch, 1995, p. 236). New concepts of higher level information have allowed investigators to go beyond the important distinctions made by Patel, Groen, and Fredriksen (1986) in the analysis of recall of medical information based on relevance and the inferences drawn from this information under conditions of uncertainty. Vicente and Wang explicitly acknowledged this as an important advance in understanding experts' memory for cases. However, they found that a subsequent study by Patel et al. [Patel, Groen, & Arocha (1990)] failed to replicate the expertise advantage for higher order information" (p. 47), thereby renewing their doubt about the sufficiency of that type of account.

It should be noted that in this cited study, Patel et al. (1990) attributed the absence of relations between expertise and recall in their studies to a methodological artifact: "It is reasonably clear that recall encounters a ceiling effect, and that this ceiling occurs at relatively low levels of expertise" (p. 404). When the clinical problems are routine in nature, it is easy to remember and recall all relevant information.

Our careful reading of Vicente and Wang's (1998) article revealed no empirical evidence for the relation between superior diagnostic performance and recall that we cannot account for by the encoding of higher level information about the patients' medical condition to facilitate reasoning and evaluation of diagnostic alternatives, which is specified in LTWM theory. The closer one can mimic the conditions of medical diagnosis, the better the subsequent encoding and recall of essential information by medical experts. In fact, Norman et al. (1989) found that medical experts' memory performance decreased when they were told to intentionally memorize, rather than merely encode, presented medical information in the context of a diagnosis. In contrast, medical students' recall performance was best when they intentionally memorized, under the same experimental conditions. A similar interaction between incidental-intentional memory and the expertise effect was observed in T. Noice and Noice's (1997) study, in which professional actors' memory for lines in a play was superior to that of college students only during the incidental study of the script. The experts' advantage disappeared when all participants were instructed to memorize the lines.

General Summary and Discussion

We share Vicente and Wang's (1998) goal of explaining expertise as a result of extended adaptation to demands of representative tasks and their environment. Furthermore, we enthusiastically support their goal of going beyond product theories toward process theories of both the structure and the acquisition of expert performance. In the body of this article, we have described a theoretical framework for the acquisition of superior expert performance that specifies how expert performers are able to sustain improved performance by engaging in designed training activities (deliberate practice) and acquiring LTWM structures that are critical to their ability to plan, monitor, and evaluate their performance. Knowledge patterns, representations, and skills are acquired to increase experts' performance and its adaptation to representative task demands. And, contrary to Vicente and Wang, we have shown that some of these mechanisms can transfer to unfamiliar "contrived" memory tasks, such as those developed by Chase and Simon (1973). We have rejected Vicente and Wang's claims of the insufficiency for our LTWM framework by successfully proposing a more detailed account of the findings cited by them in support of their product theory. In addition, we have proposed accounts of interactions of memory performance with level of expertise, rate of presentation, and specific type of "contrived" memory test. We have also shown that the more specific criticisms leveled by Vicente and Wang against Ericsson and Kintsch's (1995) LTWM framework are unwarranted and, in some cases, were based on an unfortunate misinterpretation of the original scientific studies. We now move on to a discussion of Vicente and Wang's constraint attenuation hypothesis and our respective views on how expertise can be integrated with the study of everyday activities in natural environments.

Our disagreements with Vicente and Wang (1998) can be organized around a single central issue: How can one best study individuals' skill acquisition and successful adaptation to the demands of everyday activities in a manner that includes expert performance? However, we first comment on the strengths and weaknesses of Vicente and Wang's product theory that were alluded to in our earlier discussion.


The simple and elegant formulation of the constraint attenuation hypothesis must appeal to many scientists studying expertise, which is by definition complex and has traditionally been very difficult to communicate to outsiders. One reason for its simplicity is that Vicente and Wang (1998) constrained their product theory to answer two empirical questions regarding the expertise effect in recall: "Under what conditions will there be an expertise advantage?" and "What factors determine how large that advantage will be?" (p. 35). In our first section in this article, we showed that using a straightforward application of terms such as expert and
novice does not allow us to accurately predict when the expertise advantage will be found among existing studies of the expertise effect in memory recall for “natural” domain-specific material. Consequently, we find that Vicente and Wang’s product theory in its published form cannot accurately predict when an expertise advantage in recall will be found.

To assess how successfully Vicente and Wang’s (1998) product theory can be used to make predictions about factors influencing the expertise advantage, we avoided the problem of predicting when the advantage would be found by restricting our review to the small number of domains of expertise that Vicente and Wang explicitly selected for review in their original article and thus condemned. Therefore, we assess their product theory’s ability to answer their second question and successfully predict what factors determine the relative magnitude of the expertise advantage in a given situation. The primary strength of the constraint attunement hypothesis is its ability to predict the relative magnitude of the expertise effect as a function of experimental destruction of the inherent constraints of natural stimuli. However, their theory cannot easily be extended to predict what other types of factors will influence how large the expertise advantage will be.

In our first section, we showed that familiarity with the contrived task influenced the memory performance of chess masters and that age was a stronger predictor than chess rating for the “contrived” memory performance (for an account of this age effect within our framework, see Krampe & Ericsson, 1996, and Ericsson, 1996). In the third section of this article, we reported several findings of additional factors that influence the expertise advantage. First, we found that the type of encoding condition, using incidental versus intentional memory instructions, for example, produced an interaction with the presence and magnitude of the expertise effect. Second, the type of memory test appears to interact with the expertise effect. With standard recall tests, the superiority of experts is large for regular chess positions and very small for random positions, but with recognition tests, the advantage for the experts shows a comparable magnitude for both regular and random positions. Finally, and most important, we showed the classic effects of increased presentation time. Most important, we showed that the advantage for random chess positions by experts was small for brief presentations and quite substantial for longer presentation times. With sufficiently long study times, both experts and novices are able to reproduce the random positions equally well (ceiling effect). The mere demonstration that random chess positions can be recalled virtually perfectly after sufficiently long study time would seem to rule out an account by product theories based on constraints, especially the slight deviations from “completely random positions” proposed by Vicente and Wang (1998, p. 45). We argued that only a process theory that distinguishes between different types of encoding processes for brief and long presentation times would be able to provide a satisfactory account of these interactions.

As part of their product theory, Vicente and Wang (1998) made two theoretically important supplemental assumptions that have empirical implications. First, experts learn how to extract and adapt to only a subset of the existing constraints in the environment, namely, those relevant to the goal of their primary task. Second, experts have to be attuned to goal-relevant constraints at the time of presentation in order to display the expertise effect. Our central concern is how Vicente and Wang can empirically assess the validity of these assumptions. If their product theory is going to be able to make a priori predictions about the superior memory performance of experts for goal-relevant constraints, then one needs to be able to extract and describe them in advance. The only empirical method that occurs to us for identifying those aspects would entail studying experts performing representative tasks (capture the essence of their expertise by representative tasks; Ericsson & Smith, 1991) and then assessing these aspects reflecting goal-relevant constraints by process-tracing and experimental methods. When we get down to the pragmatics of conducting actual research, Vicente and Wang’s theoretical position may not be that different from ours.

We recommend that experts be given representative tasks in which inference that the goal-relevant constraints mediate the experts’ performance can be derived from analyses of the actual performance. This implies that both approaches should study WM associated with representative performance, rather than performance on “contrived” memory tasks. If we were able to agree on the need for representative tasks, then the primary difference remaining between Vicente and Wang’s (1998) ideas and our ideas concerns the mechanisms mediating superior expert performance that are responsible for the expertise effect in memory recall. In general, we found that Vicente and Wang’s supplemental assumptions were in themselves two major steps toward a complete process theory that offers promise for an ultimate integration of our respective theoretical frameworks.

Two Different Approaches to the Study of Expertise

Vicente and Wang’s (1998) product theory and our process-based LTWM framework offer alternative accounts for most instances of “contrived” memory performance discussed in Vicente and Wang’s (1998) original article. However, there are certain important related phenomena that are well explained by only one of the theories. In the first section of this article, we discussed the study of process control by Vicente (1992), whereby “experts” and “novices” were shown complex states of the system corresponding to the combination of 34 numerical process variables that Vicente and Wang featured as a prime example of their theoretical framework. The “experts” were able to generate numerical estimates for both “natural” and “random” states of the system with less error than the novices, but the expertise advantage was reliably greater for “natural” states. Vicente and Wang elegantly explained the expertise advantage by the “experts’” superior knowledge of the structure of the environment and their reliance on constraints and relations between process variables for “natural” states of the system. We believe that the “experts’” greater attunement to relevant constraints of “natural” states will be more readily observed when the presented states contain redundant variables and when the “experts’” explicit memory of the states is poor, as in Vicente (1992). In comparison, an LTWM account of that type of phenomenon is severely hampered by the low level of accuracy and absence of reliable superiority by “experts” for the representative task of diagnosing the presented states of the complex system because the explanation of this diagnostic performance does not appear to depend on extensive access to systems information in WM. Furthermore, the lack of prior experience with that particular system even among the “experts” would inevitably lead to an unsatisfactory and complex account in the LTWM framework.
where above chance performance was based on partial memory and guessing constrained by the "experts'" superior knowledge of the structure of the system and its normal operation.

The strength of our LTWM framework is clear when the "contrived" memory tests approximate the natural conditions for representative performance in the domain of the expertise. For example, when experts are unexpectedly asked to recall task-related information after the completion of a representative task, the model of recall can be derived directly from the LTWM model of WM for the representative task, essentially without any additional mechanisms. Ericsson and Kintsch's (1995) LTWM framework was designed to account for WM in skilled activities, in particular superior performance on representative tasks that capture the essence of expertise in the domain (Ericsson & Smith, 1991). We believe that this view acknowledges the essential need to reproduce the naturally occurring situation, including a simulation of all the essential aspects of the "natural" materials required to successfully capture the superior expert performance. Furthermore, it may require simulating the dynamic context of a critical event rather than presenting only a static snapshot of a complex situation.

Probably the most important consideration is to use tasks that capture the ecologically valid temporal-processing constraints. As the time available before an action has to be initiated increases, the range of potentially useful sources of information increases as well. At extremely brief exposures, such as hitting a pitched baseball, the challenge is to uncover information that can be extracted in time to guide the selection and execution of actions. At long exposures, such as the selection of the next move in a chess game, it is possible to go beyond retrieval of appropriate actions from memory to planning and exploration of alternative possibilities before committing to an action. We believe that Ericsson and Kintsch's (1995) theoretical framework of LTWM provides an improved explanation of how expert performers acquire different mechanisms to adapt to the available temporal constraints that affect cognitive processing in task domains where accuracy is the primary challenge, such as chess and medical diagnosis, as well as in domains where speed is the primary challenge, such as mathematical calculation.

In conclusion, when two theoretical frameworks approach the same empirical phenomenon, there is frequently little cross-fertilization and productive interchange, and it is often difficult to communicate. In this article, we have made an effort to show that it is possible to develop a theoretical framework that is sufficiently general to represent our different views and eventually help us resolve the genuine disagreements through further empirical evidence. Even more exciting is the prospect of having identified a robust empirical phenomenon in natural environments that is of mutual interest to scientists from different theoretical and methodological backgrounds. Of particular interest is the fact that this naturally occurring phenomenon can be regenerated under laboratory conditions by reproducing the essential natural ecological constraints. We believe that expert performance, with its complexity of mediation and high degree of reproducibility, provides an unusual opportunity to study the results of extreme adaptation, allowing insights into the general mechanisms of the normal adaptation and learning that govern development and skill acquisition in everyday life.

References


Members of Underrepresented Groups:
Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to Demarie Jackson at the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In your letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1-4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.