7

Transfer and Expertise

DANIEL R. KIMBALL & KEITH J. HOLYOAK

Readers having a passing familiarity with research on transfer and expertise might think it somewhat odd to find these two topics covered in a single chapter. After all, one of the most common generalizations from research on expertise is that expertise is domain and task specific: experts in a domain are exceptional at performing familiar tasks within that domain, but often poor at transferring that expertise to other domains or even, in many cases, to novel tasks within the same domain (see, e.g., Ericsson & Charness, 1994). However, there are at least three ways in which transfer and expertise are related. First, transfer is a basic process in learning (perhaps synonymous with it; Gick & Holyoak, 1987), and therefore experts must use transfer in acquiring their expertise. Second, as we will see, level of expertise affects the degree and quality of subsequent transfer (Chi, Feltovich, & Glaser, 1981). Third, there appear to be types of expertise that *do* involve transfer of learning to novel tasks and domains; Hatano and Inagaki (1986) have termed such types "adaptive" expertise, which they contrast with the "routine" expertise associated with performance of familiar, domain-specific tasks. Most research to date has focused on routine expertise, perhaps because it is more conducive to exploration in the laboratory.

In this chapter, we touch on these related aspects of transfer and expertise in the context of a broader discussion of both fields. We first discuss transfer, providing a theoretical framework and outlining major research findings in the field. We then take up expertise, discussing the theories and empirical research regarding routine and adaptive expertise as they affect basic memory processes, knowledge representation, and transfer.

Transfer

Definitions and Taxonomy

Transfer is theoretically indistinguishable from learning, as can be seen by its definition: transfer is the degree to which prior performance of a task (the *training task*) affects performance on a second task (the *transfer task*) that varies in similarity to the training task (Gick & Holyoak, 1987). The two tasks may be identical (*self-transfer*), highly similar (*near transfer*), or very different (*far transfer*). While self-transfer may be regarded as the prototypical example of learning, it is also arguably rare to nonexistent, for at a minimum, the temporal, spatial, and other contextual conditions surrounding the two tasks always vary in some respect (Estes, 1955).

Closely related to the proximity of transfer is its generality. Transfer may be either *specific*, selectively influencing performance on a particular transfer task, or *general*, influencing a broader range of transfer tasks. Variations in generality of transfer are related to the distinction between the two types of expertise mentioned earlier: routine expertise, with its emphasis on domain- and task-specific knowledge, and adaptive expertise, with its emphasis on the development of general reasoning skills. We will return to this aspect of transfer in our discussion of expertise.

Transfer may be either positive or negative, depending on whether performance of the training task benefits or hinders performance of the transfer task. For example, learning to classify a small animal with a tail as a dog may result in positive transfer upon encountering a spaniel, but in negative transfer (overgeneralization) upon encountering a cat. Of course, often there will be no transfer at all, especially if the learning and transfer stimuli are very unrelated (e.g., learning to classify a small animal as a dog is unlikely to have any impact on classifying vehicles).

Factors Influencing the Magnitude and Direction of Transfer

In the preceding example, the existence and direction of transfer depend both on the characteristics of the two situations that are salient to the classifying person—the animal's size and shape of tail, perhaps, rather than the sound it may utter—as well as on the nature of the person's task—classifying animals into subclasses, rather than on some other basis (e.g., classifying them all as animals). These considerations exemplify the two main principles that determine magnitude and direction of transfer. First, the greater the overall *perceived* similarity between the training and transfer tasks, the greater will be the *magnitude* of transfer—that is, the more likely it is that transfer will be attempted. The theoretical basis for similarity is controversial, but the concept generally refers to the sharing of features and relations between the training and transfer tasks (see Tversky, 1977; Medin, Goldstone, & Gentner, 1993). However, perceived similarity may not correspond to objective similarity, because psychological factors such as context, knowledge, and expertise influence perceptions of similarity. For example, novice problem solvers in physics are often unaware of structural (functional, goal-relevant) similarities and dissimilarities between problems and therefore judge problems as similar on the basis of surface (nonfunctional, goal-irrelevant) similarities, resulting in inappropriate attempts to transfer solution methods from the training to the transfer problems (Chi et al., 1981).

Once transfer is attempted, the direction of transfer is determined by the degree of objective structural similarity between the training and transfer tasks—the extent to which the two tasks share features or relations that are causally relevant to the goal or required response. Thus, in the physics example, negative transfer will result to the extent two problems differ in objective structural characteristics yet surface similarities lead the physics novices to attempt to transfer the solution. On the other hand, physics experts accurately perceive the objective structural dissimilarities between the two problems and do not attempt transfer. By contrast, experts exhibit positive transfer on problems that do share structural components, as do novices when the surface and objective structural characteristics of the two problems are both similar.

Empirical Evidence Regarding Transfer

Because transfer depends on the application of previously acquired knowledge, it is inherently dependent on memory. Accordingly, the following discussion of transfer is organized into two major sections, each representing an aspect of memory: episodic encoding effects and episodic retrieval effects.

Encoding of the Training Task

Degree of Learning. In some cases, subjects fail to induce the operative rule during training. For example, in a problem-solving paradigm, the training task may be underconstrained, in that subjects are able to solve a category of problems by using a strategy such as trial and error that is simpler than the complex rule that the experimenter intends to define the category (Sweller, 1980). In such cases, positive transfer to a problem requiring use of the complex rule is extremely unlikely.

Barring failure of learning, so long as the training and transfer tasks require structurally similar responses, positive transfer increases with the degree of original learning (Ellis, 1965). However, when the tasks differ structurally, the relationship between degree of learning and transfer is more complex: transfer initially becomes more and more negative as the degree of learning increases, but as learning continues to grow stronger, transfer begins to become less negative and eventually even becomes positive (Mandler, 1962). This reversal in the direction of transfer with increasingly strong learning may seem to contradict the principle that negative transfer results when training and transfer tasks are structurally inconsistent. However, what seems to happen in some cases is that additional training boosts learning of very general components of the training task that in fact are structurally similar to those of the transfer task, such as general problem-solving strategies common to both tasks (e.g., "win-stay, lose-shift" in discrimination learning) and higher order similarities between training and transfer responses (e.g., the relevance of color rather than other stimulus dimensions, even though the specific colors differ in the two tasks). Thus learning of the training task occurs at different levels of abstraction, and the content at some of the levels may be structurally consistent with the transfer task.

Learning Strategies. The optimal learning strategy for maximizing transfer depends upon characteristics of the training task. For example, when a category of problems is defined by a group of interdependent rules, none of which alone is sufficient for solving the problem, some form of implicit learning (e.g., observing or memorizing instances in the absence of instructions to learn rules) may sometimes be superior to an explicit hypothesis-testing strategy (for reviews, see Reber, 1993; Seger, 1994). Another example of the influence of learning strategy on transfer involves the use of means-ends versus forward-search problem-solving strategies. A means-ends strategy involves working backwards from the goal, whereas a forward-search strategy involves working forward from the givens. Sweller, Mawer, and Ward (1983) manipulated the specificity of the goal stated in a problem and found that more specific goal statements encouraged use of a means-ends strategy, and less specific goal statements encouraged forward search. Subjects who used forward search showed better transfer to new problems, apparently having induced rules connecting givens to problem-solving operations (also see Vollmeyer, Burns, & Holyoak, 1996). As we will see later, in many problem domains forward search is one of the hallmarks of expertise.

Number and Variability of Examples. Positive transfer increases with the number of examples provided during training (e.g., Homa & Cultice, 1984), but optimizing transfer also depends on the representativeness and variability of the training examples. Varying the surface features of the examples during training, thereby more completely representing a category, permits abstraction of increasingly accurate rules for determining category membership on the basis of shared structural components (e.g., Anderson, Kline, & Beasley, 1979). For example, Bassok and Holyoak (1989) found that students who learned arithmetic progression problems in algebra exhibited more positive transfer to an isomorphic problem category—constant acceleration problems in physics—than did students who learned the physics problems and were tested for transfer with the algebra problems. Bassok and Holyoak attributed this asymmetry in transfer in part to the exposure of the algebra-trained students to a wider range of examples during training, in contrast to the physics-trained students, whose training problems all involved objects in motion. However, Bassok (1990) found evidence that a more basic cause of the failure of transfer from physics to algebra was that subjects represented the acceleration problems as rate problems rather than more generally as constant-increase problems. This view is consistent with the principle discussed earlier, according to which transfer will fail if the perceived similarity of the training and transfer tasks is low.

As another example, Gick and Holyoak (1983) manipulated the number of problems used as training examples. They used analogous "convergence" problems as training and transfer tasks. Convergence problems are based on Duncker's (1945) "radiation problem," in which the reasoner must find a way to destroy a stomach tumor without destroying surrounding healthy tissue, by using a type of ray that at sufficiently high intensity will destroy the tumor, but at that same intensity would also destroy the surrounding tissue through which the ray must pass to reach the tumor. The convergence solution requires directing multiple, converging rays toward the tumor at different angles, with the intensity of each ray being sufficiently low to avoid destruction of the surrounding tissue, but the combined intensity of the rays being sufficiently high to destroy the tumor.

Gick and Holyoak (1983) trained subjects on either one or two convergence problem analogs and tested for transfer with a second analog. An example of an analog for the radiation problem involves a general who wants to amass his forces to attack a fortress, but all the roads leading to the fortress contain mines that will detonate if a sufficiently large group traverses the road. Subjects provided with two source analogs exhibited substantially more transfer than did subjects seeing only one source analog. Gick and Holyoak argued that providing multiple-source analogs permitted abstraction of a generalized schema for the problem category, improving the likelihood that subjects will spontaneously recognize the structural similarities among the problems, thereby facilitating transfer (see also Ross & Kennedy, 1990).

Order of Examples. Positive transfer may also depend on the ordering of low-variability and high-variability examples, as well as the use of an implicit or explicit mode of learning. Elio and Anderson (1984) found that when people use an explicit strategy, actively seeking a single, deterministic rule, they benefit from presentation of high-variability examples first. Presentation of high-variability examples early makes it less likely that learners will unduly restrict the rule they induce to a limited set of features. However, when people adopt an implicit mode of learning, they seem to benefit more from early presentation of low-variability examples, which may serve to establish a strong memory representation for the examples most central to the category.

Roles of Abstract Training and Examples. Presentation of examples and abstract rules or schemata during training both appear necessary to optimize transfer. Even when abstract rules or schemata are presented, providing examples appears to facilitate transfer by showing how an abstract concept can be instantiated, especially if the concept is not part of the intuitive repertoire of the learner. Conversely, presenting an abstract rule or schema along with training examples appears to facilitate transfer to novel examples, especially when the rule is difficult to induce from examples alone (see Nisbett, Fong, Lehman, & Cheng, 1987), or when the training and transfer examples are superficially dissimilar (Gick & Holyoak, 1983). Surface and structural similarity also appear to jointly affect the efficacy of presenting abstract rules, but in a complex way. Ross and Kilbane (1997) found that when an abstract principle is presented separately from a training problem, transfer is worse for a transfer problem that reverses the structural roles of objects in the problem than for a transfer problem that uses dissimilar objects. However, when the abstract principle is instead embedded in the training problem, transfer is worse for the dissimilar-object problem than for the role-reversed problem. Thus, presenting an abstract principle separately from the training problem benefits transfer to a problem that has different surface features altogether, but impairs transfer to a problem using the same surface components in different structural roles.

Summary. In general, the conditions at encoding influence whether or not the training task is encoded in terms of structural components that are shared with the transfer task. Manipulations that foster the acquisition of generalized rules, sufficiently abstract to characterize both the training task and the subsequent transfer task, will increase positive transfer. The rules acquired must be well learned, and based on an overall set of examples diverse enough to allow generalization mechanisms to abstract the common structural components from surface differences. Direct abstract training in rules embodying appropriate solution procedures is likely to be useful, but rules for classifying novel instances into the category must also be acquired to ensure successful transfer. Unless such rules have been acquired earlier, it will be necessary to augment training in abstract rules with exposure to concrete examples.

Retrieval of Training Task at Transfer

Similarity of Surface and Structural Components. As might be expected from the principles discussed earlier, transfer is affected by manipulating similarity of training and transfer tasks. Varying the similarity of surface and structural components has different effects. For example, using the convergence-problem paradigm described earlier, Holyoak and Koh (1987) increased surface similarity to the radiation problem by using another analog involving the same type of critical object (rays), in which converging lasers were used to repair the filament of a light bulb. They found that increased surface similarity significantly improved spontaneous transfer (i.e., retrieval of the relevant analog), but did not affect use of the analogy after a hint to access the source analog was given (see also Keane, 1986; Ross & Bradshaw, 1994).

Holyoak and Koh also manipulated structural similarity by using two different versions of the light bulb analog, one in which the rationale for convergence was to avoid damaging the glass surrounding the filament—structurally analogous to avoiding destruction of healthy tissue in the radiation problem—and the other in which the convergence rationale was the unavailability of a sufficiently strong single laser source. Transfer was greater for the structurally analogous problem both before and after a hint was provided (see discussion below of uninformed versus informed transfer). The positive effect of structural similarity on spontaneous transfer suggests that similarity of relations as well as similarity of objects guides retrieval and hence transfer (also see Wharton, Holyoak, & Lange, 1996).

Procedural learning can also be influenced by the degree to which the training and transfer tasks can be characterized by structurally similar production rules. For example, Kieras and Bovair (1986) trained subjects to perform a series of procedures on a control panel device, manipulating the order of training to vary the degree to which previously learned production rules were included in new procedures either in original or modified form. Multiple regression analysis revealed that a new rule added approximately 70% more time to total training time on average than a previously encountered rule, and that the number of new production rules required by the new procedure was the best predictor of total training time for it. Thus, learning is faster and presumably easier to the extent that the training and transfer tasks are based on shared production rules.

Structural similarity between training and transfer tasks can in turn be influenced by pre-experimental knowledge, thereby affecting transfer. Bassok, Wu, and Olseth (1995) found that transfer in problem solving depends on how subjects use pre-experimental knowledge to interpret the structure of the training and transfer problems. Using isomorphic problems involving random assignment of elements from one set to another, Bassok et al. found that subjects induced a symmetric structure ("pair") if the two sets were similar types of people (e.g., doctors and doctors), but they induced an asymmetric structure ("get") if one set was objects and the other people (e.g., prizes and students). Transfer was facilitated if the interpreted structures matched, but was impaired if they mismatched (see also Bassok & Olseth, 1995).

Similarity of Processing. Positive transfer is more likely when the training and transfer tasks require use of similar processing, a phenomenon that has been termed "transfer-appropriate processing" (Morris, Bransford, & Franks, 1977). For example, McDaniel and Schlager (1990) trained subjects either by having them both generate a problem-solving strategy and implement it with specific operations (discovery condition), or by providing them with the strategy and simply having them implement it (implementation-only condition). Subjects were then tested on transfer problems that required either applying the learned strategy in a new context or generating a new strategy. Discovery subjects outperformed implementation-only subjects on the transfer problem requiring generation of a new strategy, but the two groups did not differ in performance of the transfer problem requiring only application of the learned strategy. Thus, transfer is improved when the training and transfer problems both require the same type of processing.

In another problem-solving study, Weisberg, Di Camillo, and Phillips (1978) had subjects attempt to solve Duncker's (1945) candle problem, which requires finding a way to attach a candle to a wall using given materials. On an earlier paired-associate task, some subjects had studied the word pair *box-candle*, which paired elements of the correct solution (emptying a box of tacks, tacking it to the wall, and putting the candle on the box). However, these subjects were no better at solving the problem than were subjects who had not seen the word pair. Weisberg et al. suggested that subjects use the goal, and not the separate problem elements, as the primary cue when searching memory. The processing required for the two tasks in this case, memorization and problem solving, were dissimilar, so transfer was poor.

Informed vs. Uninformed Transfer. Using the convergence problem paradigm, Gick and Holyoak (1980, 1983) established the general finding that informed transfer—transfer after subjects are given a hint as to the structural similarity of the tasks—is markedly more frequent than uninformed, spontaneous transfer, demonstrating that people can transfer learning if they are made aware of the structural similarities between training and transfer tasks. However, the hint may not be as effective after the subject has failed to transfer learning from the training task during an unsuccessful attempt to perform the transfer task. In such a case, the representation of the transfer task may become sufficiently dissimilar from that of the training task that retrieval of information about the training task after the hint will not increase the perception of structural similarity (Perfetto, Bransford, & Franks, 1983).

Expertise also affects transfer. For example, as mentioned earlier, experts generally tend to classify tasks on the basis of structural similarity more often than do novices, who use surface similarity more often (Chi et al., 1981). We now turn to expertise, examining it in a broader context before returning at the end of the discussion to consider the mutual effects of transfer and expertise in more detail.

Expertise

Focusing on the role of memory in expertise, we first examine research on routine expertise, involving memory for information related to typical tasks within the expert domain, and then the more limited body of research on adaptive expertise, involving the transfer of expert knowledge to novel tasks and domains.

Routine Expertise: Memory for Expert Domain Information

Routine experts differ from novices both in the structure of domain-related knowledge and in the ability to retrieve episodic domain-related information. Evidence indicates that experts represent domain information in long-term memory (LTM) in the form of schemas, knowledge structures that are hierarchically organized and highly interconnected semantically. With certain exceptions, experts exhibit superior episodic memory for domain-typical information largely because they can directly access LTM to rapidly and reliably encode and retrieve the information, rather than maintaining it in short-term memory (STM) alone. We first discuss the evidence regarding schematic organization of knowledge within the expert domain, and then the evidence and theories regarding the respective roles of STM and LTM in experts' domain-specific episodic memory performance.

Schematic Representation of Expert Domain Knowledge

Evidence supporting the schematic representation of knowledge in experts' LTM is largely indirect, arising from studies using episodic memory tasks, and is based on instances of both superior and inferior memory performance. Experts generally exhibit better recall and recognition than do novices for episodic information related to the expert domain, but this superiority is even more exaggerated for information that is related to the central goals within the domain. For example, Spilich, Vesonder, Chiesi, and Voss (1979) found that baseball experts recalling a baseball passage included more propositions and a higher proportion of both goal-related propositions and relations among the propositions than did novices. Spilich et al. interpreted these data as indicative of the experts' superior situation model, an instantiation of a well-developed schema. There is also evidence that experts use retrieval structures—instantiations of schemas—in episodic processing, as assumed in the theories of Ericsson and Kintsch (1995) and Gobet and Simon (1996c), discussed further below.

Even in cases in which expertise does not yield uniformly superior memory performance, performance may depend on the use of schemas in LTM. For example, Adelson (1984) found that, following a programming task, expert computer programmers had worse incidental recall than novices for details of code. This performance decrement resulted from the experts' paying greater attention to the goal structure of the programming task than to code details. Similarly, Schmidt and Boshuizen (1993) found that recall of patient information following a medical diagnosis task varied nonmonotonically with the level of expertise, exhibiting an inverted U function: subjects with an intermediate level of expertise recalled more than did those with either more or less expertise. Schmidt and Boshuizen attributed the poorer recall of the most expert subjects to an increase in selectivity and abstraction, consistent with the hierarchical, goal-driven nature of schemas (also see Patel & Groen, 1991). Studies such as these show that experts tend to pay more attention to goal-relevant information, consistent with processes of schema abstraction and instantiation, which often results in poorer memory for goal-irrelevant details.

Roles of LTM and STM in Expert Episodic Memory Performance

Theories seeking to explain the mechanisms underlying experts' generally superior episodic memory within the expert domain have differed in the degree to which they assume experts store and retrieve episodic information in STM versus LTM. We discuss three of these theories—the chunking theory of Chase and Simon (1973), the mutiple template hypothesis of Gobet and Simon (1996c), and the long-term working memory theory of Ericsson and Kintsch (1995)—and the evidence regarding STM and LTM involvement in experts' episodic memory.

Chunking Theory. A frequently cited example of superior domain-related memory is the finding by Chase and Simon (1973) that chess masters recall more piece locations from briefly presented boards reflecting actual middle-game and end-game positions than do novices, but that when the pieces are arranged randomly, this advantage disappears (or at least is extremely small; Gobet & Simon, 1996a). The masters' superior memory solely for game positions appears to be largely attributable to recognition of patterns previously encountered by the masters and stored in LTM (Gobet & Simon, 1996b). The finding of superior memory of experts for domain-typical but not domain-atypical stimuli holds in many domains other than chess, including bridge, music, medicine, computer programming, and open motor skills (see Ericsson & Lehmann, 1996, for a review).

Chase and Simon (1973) proposed a chunking theory of expertise to explain their findings. As elaborated by Gobet and Simon (1996c), this theory assumes that a chunk consists of a recurring pattern reflecting a characteristic relation (e.g., attack, defense) among a set of a few pieces. When a chunk is recognized in a newly encountered board position, a pointer is placed in STM referencing the LTM representation of the chunk, thereby reducing the effects of STM capacity limitations. The theory assumes, however, that there is no direct storage of the memory trace in LTM. Because chess masters have stored more and larger chunks in LTM, they can use the pointers in STM to access more information than novices about piece locations from a given board—at least when the board contains chunks corresponding to representations stored in LTM. However, when the pieces are randomly arranged, eliminating all but the occasional random chunk, masters cannot use their storehouse of chunks effectively, and their recall advantage over novices virtually disappears. Thus, the chunking theory accounts for the Chase and Simon results, and similar findings.

Evidence Challenging the Chunking Theory. Evidence has accumulated contesting the chunking theory's assumption that experts store episodic traces solely in STM (see Ericsson & Kintsch, 1995, for a review). For example, Charness (1976) found that chess masters' memory for briefly presented chess positions was not adversely affected by a delay, regardless of whether subjects rehearsed during the interval or performed a distractor task, even a visual chess task. Mnemonists capable of extraordinary digit spans also exhibit negligible decrements in recall after interference tasks (see Ericsson & Staszewski, 1989). As STM is presumably emptied of studied material during such delays, these findings call into question the assumption that STM alone is used to store the episodic trace.

Other evidence against exclusive STM storage includes the finding by Gobet and Simon (1996c) that chess masters who briefly viewed multiple boards recalled more chunks than STM is usually assumed to hold, and exhibited not only recency effects, indicative of STM storage, but also primacy effects, typically viewed as reflecting LTM storage. In addition, the largest chunks were larger than assumed by chunking theory and differed in size depending on the number of boards presented, contrary to the chunking theory's assumption that chunk size should not vary. Other studies indicate that experts also appear to store episodic information in LTM during incidental memory tasks, indicating that episodic storage in LTM takes place in the normal course of their activities. For example, Lane and Robertson (1979) found that incidental memory of chess positions is related to level of expertise, and so long as the study-phase task involves domain-relevant goals, incidental memory is as good as intentional memory.

The evidence indicating a direct role for LTM in experts' episodic storage and retrieval has given rise to two recent, similar theories, to which we now turn.

Multiple Template Hypothesis. Gobet and Simon (1996c) have proposed a theory of chess expertise that assumes that chess masters accumulate multiple "templates" in LTM during an episode involving presentation of multiple boards. The templates are essentially instantiations of LTM schemas and each comprises a substantial number of core piece locations and a number of variable slots. The slots can be filled by different configurations of less central pieces, as well as other information such as possible future moves and antecedent opening strategies, and may have revisable defaults. The authors note the similarity between their templates and other schematic memory structures such as frames (Minsky, 1977) and scripts (Schank & Abelson, 1977).

Gobet and Simon (1996c) view their hypothesis as a modification of the chunking theory, with the templates being more complex chunks in that they include variable slots. The hypothesis retains the chunking theory's assumption that pointers to LTM chunks and templates are placed in STM at encoding. Gobet and Simon note that filling the variable slots requires the use of some STM capacity, which they offer as an explanation for the decrease they observed in maximum chunk size as the number of studied boards increases. The hypothesis thus assumes that STM capacity limits the number of templates that can be reliably stored and retrieved in a given episode, barring the use of a deliberately acquired mnemonic retrieval structure that subsumes several templates. Gobet and Simon describe an example of a retrieval structure, comprising the list of world chess champions, that was used by a master in deliberate training to expand his memory for multiple boards.

Richman, Staszewski, and Simon (1995) applied a theory similar to the multiple template hypothesis to simulate extraordinary digit span using a revised version of the Elementary Perceiver and Memorizer computer program, EPAM IV. The program included retrieval structures in LTM similar to those reported by mnemonists in verbal protocols, as well as a discrimination net for recognition processes and an associative semantic memory. During encoding, the program associated the presented digits with aspects of both the retrieval structure and semantic memory. The program was able to simulate a mnemonist subject's learning curve in extending the digit span, as well as his free and cued recall performance, including overall accuracy, proactive inhibition with multiple lists, and the timing pattern for various operations. Especially critical to the simulation performance was the redundant storage of information in both the retrieval structure and semantic memory during encoding and rehearsal.

Long-Term Working Memory. Ericsson and Kintsch (1995) have proposed a theory, the long-term working memory theory, that is similar to the multiple template hypothesis in several respects, but is intended to apply more generally to a broader range of routine expertise. The theory assumes that experts develop skill in the rapid storage and retrieval of domain-specific information in and from LTM, using what Ericsson and Kintsch have described as long-term working memory (LT-WM). The experts' extensive domain knowledge facilitates the identification of the studied items that are most likely to require retrieval and the association of those items with the most effective retrieval cues. Experts are assumed to develop retrieval structures to aid in this process, either in the natural course of performing domain-related tasks or as a consequence of deliberate mnemonic effort. For example, one mnemonist with an extraordinary digit span associated the presented digits with long-distance running times (e.g., 3596 became 3 minutes 59.6 seconds or just under 4 minutes for running a mile; Chase & Ericsson, 1981). The association in LT-WM of the cues in these retrieval structures with the studied items during encoding facilitates the reinstatement of the cues in STM at retrieval, which in turn facilitates retrieval of the studied items from LTM. Because this association of cues and items takes place in LTM, interference tasks that are effective in preventing retrieval from STM have little or no effect on experts' memory.

Ericsson and Kintsch (1995) propose that two mechanisms, recency and elaboration, help experts overcome the proactive and retroactive interference that one might expect from association of multiple items with the same retrieval structures and cues. If the intertrial spacing at encoding is long enough, experts can take advantage of temporal distinctiveness by using recency in encoding and retrieving the information. Elaboration involves interassociation of items processed during a single trial or task, sometimes through use of higher order relations among chunks, thus providing redundant retrieval pathways. For example, Chase and Ericsson (1982) reported that a digit-span mnemonist elaborated higher order relations among sets of digits within a list to avoid interference effects from processing multiple lists of digits.

Ericsson and Kintsch (1995) surveyed research in five expert domains—mental abacus calculation, mental multiplication, dinner-order memory, medical expertise, and chess. All domains except one (mental abacus calculation) yielded evidence of substantial incidental recall and/or postsession recall by experts, indicating the encoding of new structures in LTM. For example, as noted above, Lane and Robertson (1979) found that incidental recall of chess positions is substantially equivalent to intentional recall. The exception to this pattern, mental abacus calculation, is a task requiring continuous updating of digits, resulting in severe retroactive inhibition (Hatano & Osawa, 1983). Experts in all five domains used retrieval structures. For example, medical experts reorganized randomly presented medical diagnosis information into a standard format for recall (e.g., Groen & Patel, 1988), and similarly a waiter with extraordinary memory for multiple dinner orders recalled the orders in a clockwise fashion regardless of input order (Ericsson & Polson, 1988).

Ericsson and Kintsch (1995) argue that LT-WM is a hallmark of expertise generally. To underscore that point, they offer text comprehension as an example of a domain in which expertise is common and in which they assert use of LT-WM is instrumental. They discuss the importance of LT-WM in Kintsch's (1988) computational construction-integration model, which assumes that both a propositional text base and an elaborated situation model are constructed during text processing through an alternating sequence of unconstrained semantic association and inference generation (construction) followed by constraint satisfaction (integration). Retrieval structures in LT-WM are used to store and continuously update the text base and situation models during text processing. Experienced readers become adept at building efficient retrieval structures based on accumulated domain knowledge that allows accurate anticipation of future retrieval requirements. Such retrieval efficiency is particularly evident in text processing within an expert domain, as revealed, for example, in the study of baseball expertise by Spilich et al. (1979).

Comparison of Template Hypothesis and LT-WM Theory. Both the LT-WM theory and the multiple template theory assume experts encode episodic information directly into LTM, using retrieval structures to increase effective memory capacity and processing efficiency. However, the two theories differ in some respects. The multiple template hypothesis contemplates two varieties of retrieval structures, operating at different levels: templates that serve as chunks with variable slots; and deliberately acquired, mnemonic retrieval structures that subsume several templates and that experts use to overcome interference effects. Use of the latter is the only mechanism Gobet and Simon (1996c) offer for circumventing STM capacity limits when storing multiple templates. The LT-WM theory does not explicitly distinguish between these types of retrieval structures, but is more explicit in suggesting that retrieval structures may develop naturally in the course of acquiring domain experience, as well as through deliberate mnemonic acquisition. Ericsson and Kintsch (1995) also offer temporal and elaborative encoding as two additional mechanisms that allow experts to overcome proactive and retroactive interference when using the same retrieval structure repeatedly, and therefore they apparently assume that use of an overarching retrieval structure subsuming several retrieval structures is not necessary, although its use is possible. While the template hypothesis also contemplates elaborative associations based on semantic memory, Gobet and Simon have not explicitly suggested that such elaborations would be sufficient to overcome interference effects.

Summary

The research reviewed here indicates that experts use schemas to represent expert domain information in LTM. The evidence is also consistent with the direct encoding into LTM by experts of episodic information within the expert domain, and the association of that information with elements of retrieval structures. These structures, which amount to instantiated schemas, are acquired by experts naturally through domain-related experience or deliberately through mnemonic effort, and assist in the maintenance of rapid, flexible, and reliable access to the episodic information.

Adaptive Expertise: Transfer to Novel Tasks and Domains

Relative to the substantial amount of empirical and theoretical work that has been done on routine expertise within a task domain, considerably less research has focused on adaptive transfer of expertise. Nonetheless, some general themes have emerged. The key difference between routine and adaptive experts is the greater capacity of the latter to transfer learning to novel tasks within and beyond the initial domain (Hatano & Inagaki, 1986). Adaptive experts have a deeper conceptual understanding of the domain, possessing not just "know-how" and "know-what" but also "know-why." One might expect adaptive expertise to develop to a greater degree for tasks that are variable rather than stereotyped in nature, and to emerge from free exploration more than from direct focus on achieving highly specific goals (Sweller et al., 1983). Such conditions would be conducive to the development of more abstract, structural representations of domain knowledge, which would in turn better enable application of domain knowledge to novel situations that vary in surface characteristics from previously encountered situations.

It is readily apparent that experts in one domain do not necessarily exhibit comparable performance levels in other domains (see, e.g., Chiesi, Spilich, & Voss, 1979). There is even evidence that experts in a domain can be impaired relative to novices when both attempt a task outside the domain. For example, Wiley (1998) found that people with a high degree of baseball knowledge were impaired, relative to people with less baseball knowledge, at solving a "remote associates" task that required accessing non-baseball-related associates of baseball terms. For example, people with high knowledge of baseball were less likely than those with low knowledge to find a common associate linking the words *plate, broken, shot* (where the intended answer is *glass*), apparently because the competing baseball associations to the first word, *plate*, interfered with finding non-baseball associations.

Nonetheless, other studies have found positive transfer effects. For example, Gott, Hall, Pokorny, Dibble, and Glaser (1993) found that avionics technicians who represented the functions of testing devices more abstractly showed greater flexibility in transferring their knowledge to new testing devices. Barnett and Koslowski (1997) presented unusual restaurant management problems to novice undergraduates, restaurant managers, and general business consultants. The authors coded the verbal protocols of these three groups for evidence of "deep reasoning," which was construed as an indicator of adaptive expertise. Deep reasoning involved the use of theoretical business concepts, the use of justifications and explanations to support recommendations, and discussion of complementary alternative solutions. Barnett and Koslowski found evidence of more deep reasoning among the general business consultants, compared to the restaurant managers and undergraduates, who did not differ in their use of deep reasoning. The authors suggested that the variation in business scenarios previously encountered by the consultants was crucial to the development of their adaptive expertise. The business consultants would have had more opportunity to abstract a schema applicable to general business problems, rather than those applicable only to restaurants.

Adaptive expertise may also involve shifts in problem-solving strategies. Experts often use forward search of the problem space, reasoning forward from the givens, whereas novices use backward search, reasoning backwards from the goal (Chi et al., 1981). However, some studies have found that experts adapt their search strategies to the constraints of the task. For example, expert computer programmers use backward search because the initial state places few constraints on the task (Anderson, Farrell, & Sauers, 1984; Jeffries, Turner, Polson, & Atwood, 1981). Nevertheless, the experts do still use a breadth-first search strategy in search of the goal structure, unlike novices, who use a depth-first strategy. More generally, expertise in complex tasks often is distinguished not by some single canonical search strategy but by flexible switching among alternative strategies (Dorner & Scholkopf, 1991). The determinant of strategy selection appears to be the goal structure of the task, which is consistent with schematic representation of expert knowledge and with adaptive transfer of that knowledge to novel situations.

These types of studies show that adaptive transfer can occur, given the appropriate training conditions. It also appears likely that certain abstract skills are candidates as widespread mediators of adaptive transfer. Metacognitive skills are perhaps the most important of these, as they facilitate reasoning from first principles and play a key role in assessing when understanding is lacking, when the strategy currently in use is unlikely to succeed, and when the task requires restructuring (Dorner & Scholkopf, 1991). Abstract mathematics skills may also have broad applicability in facilitating transfer (Novick, 1988; Novick & Holyoak, 1991). Abstract training in statistics and everyday deductive reasoning has been shown to facilitate transfer to novel problems (see Nisbett et al., 1987). Causal reasoning is another potential candidate as a mediator of adaptive transfer (e.g., Cheng, 1997).

In summary, there is evidence that expertise can sometimes be transferred to novel tasks within and beyond the initial domain. Broad-based experience with a variety of related situations allows the development of abstract knowledge representations and skills, which in turn facilitate transfer of the knowledge to novel situations.

Conclusions

Transfer generally, and transfer of expertise in particular, is influenced by a number of factors. The likelihood that transfer will be attempted at all is determined by the overall perceived similarity between the training and transfer tasks. Given that transfer is attempted, the degree of positive transfer is determined by the objective structural similarity of the tasks. The degree of transfer is determined jointly by whether encoding conditions permit abstraction of rules that are sufficiently general to cover both tasks (e.g., as a result of experiencing a variety of examples); whether the two tasks share surface and/or structural components; whether similar processing is used in the tasks; and whether prior experience affects the perception of the tasks.

An important way in which prior experience affects transfer is by the development of expertise. Existing theories provide a better account for routine expertise—the performance of familiar, domain-related tasks—than for adaptive expertise—characterized by a deeper conceptual understanding of the domain and transferability of knowledge and skills to novel tasks within and beyond the initial domain. Routine expertise appears to involve the organization of domain knowledge into relatively specific schemas and the use of schema instantiations and domain-specific retrieval structures to expand working memory by facilitating the rapid and reliable encoding and retrieval of episodic domain information into and from LTM. By contrast, adaptive expertise seems to require the development of more flexible and abstract learning mechanisms and schemas to promote a deeper conceptual understanding of the expert domain and the transfer of knowledge to novel tasks and domains.

Acknowledgment Preparation of this paper was supported by NSF Grant SBR-9729023.

References

Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 483–495.

Anderson, J. R., Farrell, R., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science, 8*, 87–129.

Anderson, J. R., Kline, P. J., & Beasley, C. M. (1979). A general learning theory and its application to schema abstraction. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13) (pp. 277–318). New York: Academic Press.

Barnett, S. M., & Koslowski, B. (1997). Deep processing and expertise: Etiology and applicability. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (p. 858). Mahwah, NJ: Erlbaum.

Bassok, M. (1990). Transfer of domain-specific problem solving procedures. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 522–533.

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 153–166.

Bassok, M., & Olseth, K. L. (1995). Object-based representations: Transfer between cases of continuous and discrete models of change. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1522–1538.

Bassok, M., Wu, L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory & Cognition, 23*, 354–367.

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

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–190). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 1–58). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 257–274.

Dorner, D., & Scholkopf, J. (1991). Controlling complex systems; or, expertise as "grandmother's know-how." In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 218–239). New York: Cambridge University Press.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58* (Whole No. 270).

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition, 12*, 20–30.

Ellis, H. C. (1965). *The transfer of learning.* New York: Macmillan.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725–747.

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

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273–305.

Ericsson, K. A., & Polson, P. G. (1988). An experimental analysis of a memory skill for dinner-orders. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 305–316.

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. 235–267). Hillsdale, NJ: Erlbaum.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369–377.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of training: Contemporary research and applications* (pp. 9–46). New York: Academic Press.

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

Gobet, F., & Simon, H. A. (1996b). The roles of recognition processes and look-ahead search in time-constrained expert problem-solving: Evidence from grand-master-level chess. *Psychological Science, 7*, 52–55.

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

Gott, S. P., Hall, E. P., Pokorny, R. A., Dibble, E., & Glaser, R. (1993). A naturalistic study of transfer: Adaptive expertise in technical domains. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 258–288). Norwood, NJ: Ablex.

Groen, G. J., & Patel, V. (1988). The relationship between comprehension and reasoning in medical expertise. In M. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 287–310). Hillsdale, NJ: Erlbaum.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). San Francisco: Freeman.

Hatano, G., & Osawa, K. (1983). Digit memory of grand masters in abacus-derived mental calculation. *Cognition, 15*, 95–110.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*, 332–340.

Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 83–94.

Jeffries, R., Turner, A. T., Polson, P. G., & Atwood, M. E. (1981). The processes involved in designing software. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 255–283). Hillsdale, NJ: Erlbaum.

Keane, M. T. (1986). On retrieving analogues when solving problems. *Quarterly Journal of Experimental Psychology, 39A*, 29–41.

Kieras, D. E., & Bovair, S. (1986). The acquisition of procedures from text: A production-system analysis of transfer of training. *Journal of Memory and Language, 25*, 507–524.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95*, 163–182.

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

Mandler, G. (1962). From association to structure. *Psychological Review, 69*, 415–427.

McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction, 7*, 129–159.

Medin, D., Goldstone, R., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254–278.

Minsky, M. (1977). Frame-system theory. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 355–376). Cambridge: Cambridge University Press.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238*, 625–631.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 510–520.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 398–415.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93–125). New York: Cambridge University Press.

Perfetto, G., Bransford, J., & Franks, J. (1983). Constraints on access in a problem solving context. *Memory & Cognition, 11*, 24–31.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* New York: Oxford University Press.

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

Ross, B. H., & Bradshaw, G. L. (1994). Encoding effects of remindings. *Memory & Cognition, 22*, 591–605.

Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 42–55.

Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 427–440.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Erlbaum.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21*, 338–351.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115*, 163–196.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 275–290.

Sweller, J. (1980). Hypothesis salience, task difficulty and sequential effects on problem solving. *American Journal of Psychology, 95*, 455–483.

Sweller, J., Mawer, R. F., & Ward M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112*, 639–661.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*, 75–100.

Weisberg, R., Di Camillo, M., & Phillips, D. (1978). Transferring old associations to new situations: A non-automatic process. *Journal of Verbal Learning and Verbal Behavior, 17*, 219–228.

Wharton, C. M., Holyoak, K. J., & Lange, T. E. (1996). Remote analogical reminding. *Memory & Cognition, 24*, 629–643.

Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition, 26*, 716–730.