The Nature of Expertise in Anagram Solution[1]

Laura R. Novick & Nathalie Coté
Dept. of Psychology & Human Dev.
Box 512 Peabody
Vanderbilt University
Nashville, TN 37203
novicklr@vuctrvax.bitnet
cotenc@vuctrvax.bitnet

Abstract

Second-generation theories of expertise have stressed the knowledge differences between experts and novices and have used the serial architecture of the production system as a model for both expert and novice problem solving. Recently, Holyoak (1991) has proposed a third generation of theories based on the idea of expertise-related differences in the processing of solution constraints. According to this view, the problem solving of experts, in contrast to that of novices, often is better characterized as a process of satisfying multiple solution constraints in parallel than as a process of serially testing and rejecting hypotheses. We provide data from three experiments that are consistent with this hypothesis for the domain of anagram solution.

Background

How does the problem solving behavior of experts differ from that of novices? According to Holyoak (1991), the first generation of research on expertise was based on Newell and Simon's (1972) theory that experts are distinguished by their superior ability to employ general heuristic search methods. Subsequently, research in domains such as chess and physics discovered that heuristic search was a weak method actually used more often by novices than by experts. This research spawned a second generation of theories that focussed on expertise-related differences in knowledge. According to the new theories, what distinguished experts from novices was the larger size, superior organization, and greater accessibility of their knowledge base within the domain of expertise. Regardless of level of expertise, however, problem solving was conceptualized as a serial process based on the architecture of the production system.

Recently, Holyoak (1991) has challenged the second-generation view of expertise. In addition, he has proposed a third generation of research that would be focussed on processing differences rather than knowledge differences. According to this new view, theories of expert performance would be based on the parallel architecture of connectionism (e.g., Rumelhart, McClelland, & the PDP Research Group, 1986). The hypothesis is that problem solving in experts, in contrast to that of novices, often is better characterized as a process of attempting to satisfy multiple task constraints in parallel than as a process of serially testing and rejecting hypotheses.

A complete understanding of expertise most likely will require reference to both knowledge and processing components of performance. In this paper, we highlight the issue of processing differences. Studying processing differences in knowledge-rich domains such as chess and physics is difficult, however, because solvers at different levels of expertise also will differ in terms of their knowledge bases. Given this potential confound, we chose anagram solution as the domain of study, because expertise-related differences in knowledge could be minimized.

Anagram Solution

Anagram solution involves unscrambling a string of letters into an English word (e.g., "iasyd" becomes "daisy"). Clearly, anyone literate in English has the knowledge necessary to solve most anagrams (e.g., knowledge of spelling

[1] The research reported in this paper was supported by a grant from the University Research Council of Vanderbilt University.

constraints). To further minimize any impact of knowledge differences in our research, we used only five-letter anagrams of common words (such as the example just given), and our subjects were college students at a highly selective university.

A second important criterion for selecting a domain is that there is some *a priori* reason for expecting experts to be more likely than novices to engage in parallel processing of solution constraints. Anagram solution meets this criterion as well. An intriguing phenomenon is that sometimes the answer to an anagram seems to "pop out" very quickly without any conscious awareness of a solution attempt. Anecdotal reports from self-proclaimed experts suggest that they solve many five-letter anagrams (e.g., "erjko", "dnsuo", "rcwdo", "iasyd") in less than 1.5-2 seconds. In contrast, novices do not often report pop-out solutions.

The hypothesis that experts attempt to solve anagrams by trying to satisfy in parallel the multiple, often conflicting, constraints on the rearranged order of the letters suggests a potential mechanism for the occurrence of the pop-out phenomenon. Moreover, experts' intuitions concerning pop-out solutions contrast with the conclusions of the experimental literature (which presumably is heavily weighted by data from non-expert solvers), which characterizes anagram solution as a deliberate, serial process of testing and rejecting hypotheses.

Although we believe that serial processing of solution constraints characterizes much of anagram solution (even by experts), we do not believe that it provides a complete account of expert behavior. Consistent with research in other domains (e.g., chess), we suspect that anagram experts often unscramble anagrams by a parallel rather than a serial process.

Serial and Parallel Models of Anagram Solution

How would serial and parallel models of anagram solution differ? Numerous answers to this question are possible. Our goal here is simply to provide a brief description of what each type of model could look like, not to construct detailed simulations of such models. To understand the experimental predictions we make later, it will help to have in mind a concrete example of each type of solution process.

Mendelsohn (1976) has proposed a serial, hypothesis-testing model of anagram solution. The first phase of solution involves forming hypotheses about the correct letter order based on the judged likelihood of each possible bigram (two-letter combination). These hypotheses are formed in decreasing order of the bigrams' frequencies in the language. For example, given the anagram "dnsuo," the first three hypotheses to be tested concerning the initial bigram of the solution would be, in order, "un", "so", "do" (Mayzner & Tresselt, 1965). The second phase of solution involves testing each hypothesis by retrieving from memory words that match the hypothesized partial reorganization of the anagram. As each hypothesis fails to match a word, the next most probable one is tested until a solution is found. An alternative serial model of anagram solution might propose that the second phase involves rearranging the remaining (i.e., non-initial) letters of the anagram and attempting to find a match between the candidate solutions and entries in one's mental lexicon.

Now consider a parallel model based on a connectionist architecture. In one such model (Novick, in progress), the (symbolic) processing units correspond to hypotheses about possible combinations of letters and positions (e.g., D in position 1, denoted here by D1, for the anagram "dnsuo"). Excitatory and inhibitory links (denoted by "<+>" and "<≠>", respectively) between the letter/position units embody constraints on the rearrangement of the letters. Inhibitory links instantiate the constraints that each letter can occupy only a single position (e.g., D1<≠>D2) and each position can contain only a single letter (e.g., D1<≠>N1). Excitatory links enforce English spelling rules by favoring bigrams that are more common in the language (e.g., U1<+>N2 would be greater than U1<+>S2). Constraints on individual letter/position units also can be modeled. Such constraints might include a bias to begin words with consonants, to put vowels in the middle of a word, and to keep the letters in the same positions in which they occur in the anagram. Mayzner and Tresselt (1958) provided experimental evidence for the last constraint. The letter/position units accumulate activation over time, as a function of the other units to which they are connected and the weights on the connections, until a steady state is reached. To a first approximation, a solution is achieved if the five units with the highest activations form an English word.

The models just described represent extremes along a continuum of degree of serialism versus parallelism. Hofstadter (1983) has proposed a model of anagram solution that incorporates both types of processing. In this model, letters float in a "cytoplasm" looking for other letters with which to form clusters of increasing size until a word is created. Although the progression from isolated letter to bigram to syllable to word occurs serially, the clustering of units at different levels happens in parallel. Different clusterings are explored to different depths, depending on such constraints as rules of spelling, and this also occurs in parallel. As structures coalesce, the model gradually makes a transition from primarily parallel to primarily serial processing.

Experiments 1a and 1b

Our first goal was to provide scientific evidence for the existence of very fast solutions (i.e., the pop-out phenomenon) and for the association of pop-out with expertise. In addition, for the design of Experiment 2, we needed to identify words that experts were more likely to solve quickly than novices. We will refer to such stimuli as discriminating anagrams.

Method

Subjects. The subjects in Experiment 1a were 17 psychology graduate students and 2 undergraduates who were selected to represent the full range of self-reported ability on a 1 (awful) to 9 (excellent) scale. In the experimental session, subjects completed an objective test of anagram-solving ability (the Scrambled Words Test) in which they were given 10 min to solve 20 difficult five-letter anagrams selected from Arnold and Lee (1973). The subjects in Experiment 1b were 20 undergraduates who were preselected based on their Scrambled Words Test scores: eight high solvers had scores of 12 or more, four intermediate solvers had scores between 7 and 11 inclusive, and eight low solvers had scores of 6 or less.

Materials, design, and procedure. Subjects in Experiment 1a solved 110 core anagrams. An additional 40 filler anagrams that required only one letter move for solution (e.g., "pkoer") were interspersed throughout this list to ensure that all subjects would have some success at the task. Subjects in Experiment 1b solved 120 core anagrams. Some of the anagrams were the same as those used in Experiment 1a. Others were new scramblings of the old words because the earlier results did not enable us to identify a sufficiently large set of discriminating anagrams. Such anagrams are difficult to identify, because numerous anagram problems can be constructed for a word, and the different anagrams are not equivalent in difficulty (e.g., "rcwdo" results in more fast solutions than does "dwcor"). The anagrams were divided into two blocks. Subjects also completed a block of 10 practice anagrams. All anagrams were printed in lower-case letters.

The anagrams were presented one at a time on a computer screen for a maximum of 10 sec each. Subjects pressed a button on a response box as soon as they solved the anagram, and then they reported their solution out loud to the experimenter. The computer recorded the solution time. Feedback was given concerning the correct solution for each trial.

Subjects in both experiments also completed the Concealed Words Test from the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963). This is a speeded test in which subjects have to identify words that have been partially erased (see Figure 1). French, Ekstrom, and Price indicate that performance on this test is correlated with anagram solution. More important for our purposes, successful performance on the Concealed Words Test would seem to require parallel processing, because each partial letter is ambiguous in isolation, and in fact McClelland and Rumelhart (1981) used their computer model's ability to "read" such items as supporting evidence for the use of parallel processing in the identification of letters.

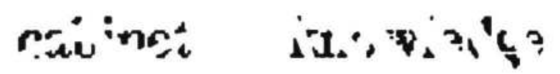

Figure 1. Two examples of the type of item that appears on the Concealed Words Test. The solutions are "cabinet" and "knowledge," respectively.

Results

Because the literature provides no guidance in defining fast solutions, any criterion is somewhat arbitrary. We chose 2 sec as our cut-off time for a fast solution, because below about 2 sec experts get clear intuitions of the solution popping out, whereas above about 2 sec experts get clear intuitions of using a serial strategy. To control for the fact that by definition better anagram solvers solve more anagrams than poorer solvers overall, our measure of proficiency at solving anagrams quickly was the percent of each subject's solutions that occurred in under 2 sec. We will refer to this measure as %rapid.

Not surprisingly, accuracy on our relatively unspeeded test of anagram expertise (the Scrambled Words Test) was highly correlated with accuracy on the core anagrams in our speeded experimental task: $r = .79, p < .01$, and $r = .54$, $p < .05$, for Experiments 1a and 1b, respectively. More importantly, expertise was highly correlated with *speed* of solution of the experimental items, as defined by the %rapid measure: $r = .71, p < .01$, and $r = .46, p < .05$, respectively, for Experiments 1a and 1b. Although the correlation between expertise and fast solutions does not illuminate directly the issue of strategy use, the data from the Concealed Words Test provide evidence consistent with the hypothesis that experts are more likely to use a parallel solution strategy than are novices. In both experiments, accuracy on this test was highly correlated with the %rapid measure: $r = .61$ and $r = .57$ for Experiments 1a and 1b, respectively, both $p < .01$.

Experiment 2

Gathering experimental evidence for parallel versus serial processing is difficult, because typically it is possible to construct models of the two types that mimic each other. Nevertheless, it is important to distinguish the two models (Townsend, 1990). The purpose of Experiment 2 was to begin to explore the consequences of the hypothesized greater frequency of parallel processing of solution constraints among experts than novices. We used a methodology described by Townsend in which subjects' processing was interrupted prior to completion. After any fixed amount of time, more partial information will be available to those engaged in parallel rather than serial processing, because in the former case multiple pieces of information are processed simultaneously, whereas in the latter case processing proceeds sequentially. For example, assume that processing of a five-letter anagram is interrupted after *N* ms. With parallel processing, one has *N* ms of partial information on each letter, including information on the various constraints concerning the locations of the letters and bigrams. With serial processing, however, one has either *N* ms of partial information on one letter or *N*/5 ms of partial information on each letter (or something between these two extremes; in any case, less information is available than with parallel processing). Information about the constraints on the letter positions would be sparser, because that information cannot be accessed until all of the letters have been encoded.

Method

Solvability judgment task. The subject's task was to judge whether a string of letters could be unscrambled to form an English word (e.g., "dnsuo" forms "sound," but "rusyb" does not form a word no matter how the letters are rearranged). The stimuli were presented at one of three display durations: short, intermediate, or long. Even at the relatively long durations, however, the time allotted for processing was so brief that most subjects were not expected to be able to solve the anagrams. Thus subjects were forced to make their solvability judgments based on partial information. Because the unsolvable items were very similar to the anagrams, considerable information about the possible positions of *all* letters would be needed to distinguish the two types of stimuli. If subjects have considerable information on only a few of the letters, or very little information on all letters, their performance should be near chance level.

Our hypothesis of greater parallel processing among experts than novices leads to the prediction that high solvers will be more accurate than low solvers at distinguishing solvable from unsolvable items. At the shortest display durations, we expect that the low solvers' performance will be at or near chance. In contrast, high solvers may be above chance even at the short durations.

Materials. Based on the results of Experiments 1a and 1b, we chose 30 anagrams that discriminated high and low solvers in terms of solution time, with pop-out being more likely for the high solvers. The items were selected to meet the following criteria: (a) the correlation between solution time and expertise was at least as extreme as -.24 (M = -.39, ranging from -.24 to -.67), (b) there was a gap in the solution time distribution of at least 200 ms at or before 2 sec (M = 476 ms, ranging from 204-1313 ms), and (c) the difficulty of solution was moderate, defined as 20-65% of subjects solving the anagram at or below the lower limit of the break in the solution time distribution (see (b) above; henceforth referred to as the rapid solution cut-off time; M = 41%). In addition to the 30 experimental anagrams, there were 45 training and 15 warm-up anagrams.

Each anagram was matched to a "nonanagram" (i.e., an item that could not be unscrambled to form an English word; e.g., "clnai"), which was constructed as follows: First, a word was selected that began with the same letter as the anagram. Then, one letter of that word was replaced by another letter (vowel for vowel or consonant for consonant) such that the resulting set of five letters could not be rearranged to form a word. The letter that was substituted met the restriction that its frequency of occurrence in English differed from that of the replaced letter by a ratio of no more than 2:1 (see Pratt, 1942). Finally, the letters of the nonword were scrambled such that the absolute difference between the summed bigram frequency (SBF) of the nonanagram and that of its matched anagram was no more than 25 points (M = 10.38). The SBF of "clnai," for example, is 48, which is obtained by summing the frequency of CL in positions 1 and 2, LN in positions 2 and 3, etc. (see Mayzner & Tresselt, 1965).

Subjects, design, and procedure. The subjects were 30 undergraduates who were preselected based on their Scrambled Words Test scores: 15 high solvers had scores of 12 or more, and 15 low solvers had scores of 6 or less. We crossed the two levels of expertise with three levels of display duration. Because the rapid solution cut-off times differed for the 30 experimental anagrams (ranging from 1318-1999 ms), the short, intermediate, and long display durations were defined as percentages of the cut-off times: 45% (M = 803 ms), 70% (M = 1249 ms), and 85% (M = 1517 ms), respectively. Three stimulus lists were constructed so that each subject would see a given experimental item at only a single display duration, but across lists each item would appear at each of the three durations. The nonanagram display durations were yoked to their matched anagram times.

A deadline procedure was used to force subjects to make their solvability judgments based on partial information. The stimulus was displayed for a predetermined duration. Coincident with the offset of the stimulus, a beep sounded indicating that subjects were to respond. If subjects did not respond within 250 ms, another beep sounded to indicate the end of the response period. Responses were recorded up to 125 ms after the second beep. Feedback on each trial included the response time for that trial, the average response time, and the solution (or "not a word" for the nonanagrams). Subjects completed 90 training trials prior to the experimental items. Responses to only 5% of the items were lost due to subjects failing to respond within the 375 ms deadline.

Results

Our primary measure of performance was d', a sensitivity measure from signal detection theory. Applying the theory to our task, we assume that each stimulus yields an impression of solvability at some value along a "sensory" continuum. Then, d' is defined as the distance between the means of the distributions for the anagrams and the nonanagrams (in standard score units). The more sensitive subjects are to the solvability of the items, the farther apart their two distributions and the higher their d' scores. A d' of 0 indicates chance performance.

A d' score was computed for each display duration for each subject. An ANOVA on these data indicated that high solvers were more sensitive to item solvability than were low solvers, $F(1, 28) = 18.65$, $p < .001$. In addition, sensitivity increased as display duration increased, $F(2, 56) = 2.88$, $p < .07$. There was no interaction between level of expertise and display duration, $F(2, 56) < 1$. The mean d' scores are shown in Table 1. For the high solvers, all of the means were reliably above chance. In contrast, low solvers' performance was reliably above chance only at the longest (85%) duration. In further support of the hypothesis that good performance on the solvability judgment task involves parallel processing, mean d' scores (collapsed across display duration) were positively correlated with Concealed Words Test scores, $r = .43$, $p < .05$.

	Sensitivity (d')			
Exposure	High Solvers		Low Solvers	
Duration	M	SD	M	SD
45%	0.79	0.84	0.20	0.55
70%	1.17	0.73	0.31	0.84
85%	1.34	0.68	0.43	0.64
mean	1.10	0.57	0.31	0.42

Table 1. Sensitivity to Solvability (d') at Each Display Duration for High and Low Solvers.

We also analyzed subjects' criteria for choosing a response. "Sensory" values above and below the criterion lead to responses of "solvable" and "not solvable," respectively. We used C as a measure of the location of the response criterion, because it is independent of d' (Snodgrass & Corwin, 1988). Mathematically, C is the distance (on the sensory continuum) of the criterion from the intersection of the anagram and nonanagram distributions. Positive C scores indicate a strict criterion (bias to respond "not solvable"), and negative scores indicate a lenient criterion (bias to respond "solvable"). $C = 0$ means that the subject is unbiased. An ANOVA indicated a marginally reliable effect of expertise, $F(1, 28) = 3.49$, $p < .08$, with high solvers setting a stricter criterion than low solvers (see Table 2). There also was a main effect of display duration, $F(2, 56) = 3.79$, $p < .03$, indicating that response criteria became stricter as display duration increased. It makes sense that subjects would require more evidence before deciding that an item is solvable if they already have been working for a relatively long time without getting an answer. There was no interaction between level of expertise and display duration, $F(2, 56) < 1$.

Exposure Duration	Response Criterion (C) High Solvers M	SD	Low Solvers M	SD
45%	0.00	0.60	-0.13	0.64
70%	0.39	0.50	0.02	0.39
85%	0.36	0.43	0.08	0.58
mean	0.25	0.38	-0.01	0.40

Table 2. Response Criterion (C) at Each Display Duration for High and Low Solvers.

Discussion

The research reported here has been conducted within the framework of the third generation of expertise theories (Holyoak, 1991), which proposes that expert performance is based on parallel processing of solution constraints, in contrast to the serial processing of novices. Our data on anagram solution are consistent with this hypothesis. First, we provided evidence of a pop-out phenomenon of very fast anagram solution (within 2 sec) that is highly correlated with expertise. Second, when subjects had to judge whether letter strings were solvable based on incomplete processing of those strings, performance was directly related to expertise. In fact, at the shorter display durations, only the more expert solvers performed reliably above chance. Finally, both the frequency of pop-out solutions and the ability to discriminate solvable and unsolvable stimuli based on partial information were highly correlated with scores on the Concealed Words Test, which prior research would suggest requires parallel processing for good performance.

In sum, the research reported here has provided evidence concerning fundamental processing differences as a function of expertise. Clearly, any attempts to facilitate the acquisition of expertise must be based on a solid understanding of what is to be acquired. Although anagram solution probably is not a skill at which most people wish to become expert, it is an excellent domain for testing the hypothesis under consideration. Once the nature of processing differences is understood in simpler domains (such as anagram solution), the work can be extended to more knowledge-intensive and "messier" domains, such as physics problem solving or medical decision making.

References

Arnold, H., & Lee, B. (1973). *Jumble #7*. NY: The Berkeley Publishing Group.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.

Hofstadter, D. R. (1983). The architecture of Jumbo. *Proceedings of the International Machine Learning Workshop* (pp. 161-170). Monticello, IL.

Holyoak, K. J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 301-335). NY: Cambridge University Press.

Mayzner, M. S., & Tresselt, M. E. (1958). Anagram solution times: A function of letter order and word frequency. *Journal of Experimental Psychology, 56*, 376-379.

Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplement, 1*, 13-32.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

Mendelsohn, G. A. (1976). An hypothesis approach to the solution of anagrams. *Memory & Cognition, 4*, 637-642.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Novick, L. R. (in progress). [Alfred: An implementation of parallel constraint satisfaction as a model of anagram solution by experts].

Pratt, F. (1942). *Secret and urgent: The story of codes and ciphers*. Garden City, NY: Blue Ribbon Books.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1*, 46-54.