Journal of Experimental Psychology:
Learning, Memory, and Cognition
1994, Vol. 20, No. 1, 245–249

0278-7393/94/$3.00

Plausible Models of Alphabetic Search: Reply to Scharroo, Leeuwenberg, Stalmeier, and Vos (1994)

David Klahr

D. Klahr, W. C. Chase, and E. Lovelace (1983) proposed a model of the cognitive processes involved in alphabetic retrieval in terms of a 2-level hierarchy of forward-linked associations. J. Scharroo, E. Leeuwenberg, P. F. M. Stalmeier, and P. G. Vos (1994) attempt to demonstrate that a simple associative model is more plausible, more parsimonious, and a better fit to the data than is Klahr et al.'s model. In this commentary I argue that Scharroo et al. misrepresent the way in which Klahr et al. evaluated their model and that they fail to demonstrate the superiority of a simple associative model. In addition, I suggest that a composite model that integrates the distinctive features of both models would advance our understanding of the process of alphabetic retrieval.

Klahr, Chase, and Lovelace (1983) proposed a model of the cognitive processes involved in alphabetic retrieval. They modeled both the structure of the alphabet (as a two-level hierarchy of forward-linked associations) and the retrieval process invoked in a variety of alphabetic retrieval tasks. They tested their model by creating a computer simulation (ALPHA) of the retrieval process operating on their proposed alphabetic structure and by regressing model effort against response times from their own experiments, as well as several that had been previously reported. Scharroo, Leeuwenberg, Stalmeier, and Vos (1994) attempted to demonstrate that a simple associative model is more plausible, more parsimonious, and a better fit to the data than is Klahr et al.'s ALPHA. Scharroo et al.'s article consists of four parts: (a) a summary and critique of ALPHA; (b) an alternative formulation of the process of alphabetic retrieval, in which response times are assumed to reflect the pattern of association strengths between probes and responses; (c) the results of an experiment in which Scharroo et al. attempted to replicate the Klahr et al. experiment; and (d) a comparison of the fit of Klahr et al.'s ALPHA and Scharroo et al.'s model of alphabetic search by association (ASA) to the data from their single experiment and another reported by Browman and O'Connell (1976).

In this commentary, I argue that Scharroo et al. (1994) misrepresented the way in which Klahr et al. (1983) evaluated ALPHA and that they failed to demonstrate the superiority of ASA over ALPHA. In addition, I suggest that a composite model that integrates the distinctive features of both models would advance the understanding of the process of alphabetic retrieval.

Scharroo et al.'s Summary and Critique of Klahr et al.

Scharroo et al. (1994) focused on Klahr et al.'s (1983) weak and strong tests. They concluded that the weak test is trivial because it is circularly defined and that the strong test is not passed with sufficient frequency to support ALPHA. Neither claim is correct.

1. The weak test is neither circular nor trivial because the alphabet segmentation is defined by an independent measure and then that same segmentation is used for all subjects. Klahr et al. were quite explicit on both of these points:

> Our decision about how to segment the alphabet [in Experiment 1] is based on an informal post hoc analysis of local extreme points. The fact that this segmentation is consistent with the phrasing in [the Alphabet Song] provides some additional basis for believing it to be correct, but we have no direct independent evidence that it is the segmentation used by our subjects. In the second experiment, we asked subjects to report directly their "entry points," if any. The independent assessment of the segmentation allowed us to perform a more rigorous evaluation of the model. (Klahr et al., 1983, pp. 465–466)

Scharroo et al. apparently ignored the analysis summarized in Klahr et al.'s Tables 2 and 3. Instead, they chose to use the aggregate curves (Scharroo et al., Figure 2) to count the frequency with which the strong test was passed. The whole point of Klahr et al.'s strong and weak tests, however, was to assess the extent to which "individual subjects have the same chunk boundaries as those determined by the aggregate analysis" (p. 473). Because Klahr et al. used the same set of chunk boundaries to evaluate each individual subject, the circularity argument is refuted.[1]

2. The results of Klahr et al.'s strong test are statistically significant. On the basis of a visual inspection of Klahr et al.'s data (reprinted in Scharroo et al., Figure 2), Scharroo et al. reported that the strong test "is only positive in 10 out of the 20 cases" (p. 237), and they claimed that this is what one would expect from a random RT pattern. What Klahr et al. actually showed in their Table 2 is that the frequency distributions for the strong test obtained from their 42 subjects are highly significant ($p < .002$ and $p < .0001$ for Experiments 1 and 2, respectively[2]).

I thank Jim Staszewski for his insightful comments and suggestions about alphabetic retrieval.

Correspondence concerning this article should be addressed to David Klahr, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

[1] Indeed, if the weak test was trivial because of the circularity of its definition, then it should have been passed on 100% of the boundaries by all subjects.

[2] On this point of analysis, as well as several others, the consistency between Experiments 1 and 2 is quite remarkable given the fact—not

Table 1
First-Order Intercorrelation Matrix Between Klahr et al. Experiment 1 Mean Reaction Times and Different Models

Measure	1	2	3	4	5
1. After mean	—				
2. Before mean	.677	—			
3. ALPHA	.717	.670	—		
4. ASA	.453	.596	.702	—	
5. Position	.575	.410	.597	.153	—

Note. Pearson product–moment correlations. Reaction times are from Klahr, Chase, and Lovelace (1983). ASA = alphabetic search by association; ALPHA = Klahr et al. simulation.

Although Scharroo et al. (1994) noted that they find the backward search at Level 2 plausible, they neglected to comment on one of the crucial assumptions associated with the process. To quote their example, "for finding the letter preceding the letter N, one might review the series L, M, N within the chunk [L . . . P]" (Scharroo et al., 1994, p. 243). But why not review the series A,B,C or U,V,W? In other words, how does the subject, given the probe N, know which chunk to access and search? Where is the information stored that locates the correct chunk? A proponent of an associative model might argue that associative links to neighboring letters facilitate the location of a starting letter in these situations. However, an associative model that is extended to include the auxiliary assumption that, for some letters, there are stronger links to letters several steps prior to the probe than those immediately adjacent becomes increasingly similar to ALPHA. Furthermore, such extensions would reduce both the simplicity and the plausibility of the pure associative model proposed by Scharroo et al.

Scharroo et al.'s Experiment With Dutch Subjects

Scharroo et al. (1994) argued that the characteristic shape of the response time (RT) curves reported in Klahr et al. (1983) may be an artifact of averaging over individual subject RT curves with widely differing patterns of peak RTs. The most direct way to evaluate this hypothesis would have been to obtain Klahr et al.'s data for further analysis. Instead, Scharroo et al. ran their own version of the Klahr et al. experiments, using Dutch college students. The most interesting result from their experiment is that the RT curves from Dutch students look like no other results reported in the literature, either individually or in aggregate. The curves lack the characteristic slowly ascending, sawtooth pattern reported by Klahr et al. and all previous studies. The Scharroo et al. data also contain many more peaks and valleys.

Comparing ALPHA and ASA

These fundamental differences in the basic phenomenon notwithstanding, Scharroo et al. (1994) then proposed that the way to compare ALPHA and ASA is to fit them both to the results of their experiment. The comparisons are based mainly on the amount of variance accounted for by different regression models, which are based on a variety of aggregations and assumptions. Scharroo et al. did not attempt to replicate Klahr et al.'s (1983) analysis of strong and weak tests, even though, as suggested earlier, it is the method used by Klahr et al. to assess the robustness of both the segmentation of the alphabet and individual subjects' use of an ALPHA-like retrieval process.

When the models are compared at the same level of aggregation with the same number of degrees of freedom (to be explained shortly), they perform equally well. ASA's only advantage over ALPHA comes from regressing the distribution of reported segmentations against RTs (bottom row of Scharroo et al., 1994, Figure 8). Here ASA accounts for the same amount of variance for the forward task (60%) as does ALPHA, but it accounts for 81% of the variance for the backward task, compared with ALPHA's 44%. Note, however, that ASA has been transformed from a parsimonious 0/1 associative model to a 25-parameter model (the level of association for each adjacent letter pair, based on Scharroo et al.'s Figure 7.) Scharroo et al. did not address this problem, nor did they comment on the substantial differences between the R^2s multiple squared regressions for forward and backward tasks for all analyses involving ASA.

There are several additional problems with this procedure. First, there is much more between-subjects variance in alphabetic segmentation for Dutch subjects than for the American subjects whose responses ALPHA was designed to model. As noted, the Dutch students' curves are very unusual. This is because nearly all of the Americans have learned the alphabet through the same nursery rhyme, whereas—according to Scharroo et al. (1994)—there is no equivalently widespread method of training used in The Netherlands. If so, then it is difficult to know what segmentation to use when comparing the models, and yet they both depend crucially on the proper segmentation. Scharroo et al. used a very crude measure for segmentation, based on an arbitrary level of consistency among subjects' responses to vague questions about their "preferred segmentation of the alphabet."[3] Another problem with Scharroo et al.'s analysis is that they reported only the amount of variance (in RTs) accounted for, but they did not propose any parameter estimates for the components of the RTs. One of the strengths of the Klahr et al. (1983) model is the consistency between the range of search rates they estimated for their model and the extensive literature on similar search parameters in other cognitive models. Scharroo et al. were silent on

emphasized by Klahr et al. (1983)—that the two experiments were run in laboratories at two different universities having different subject populations, with different procedures, by experimenters who decided to collaborate on the Klahr et al. article long after the experiments had been run.

[3] Note that the Scharroo et al. (1994) probe for segmentation is very different from the procedure used by Klahr et al. (1983). Scharroo et al. asked subjects at the end of the experiment to draw lines indicating their preferred subdivision of the alphabet, whereas Klahr et al. asked subjects to report following each trial what, if anything, they had done to "think of the correct response." In other words, the Klahr et al. subjects were asked to report on the contents of working memory from the immediately preceding trial, whereas the Scharroo et al. subjects were, in effect, asked to propose a theory of alphabetic segmentation. See Ericsson and Simon (1984) for comments on the relative validity of these two types of verbal reports.

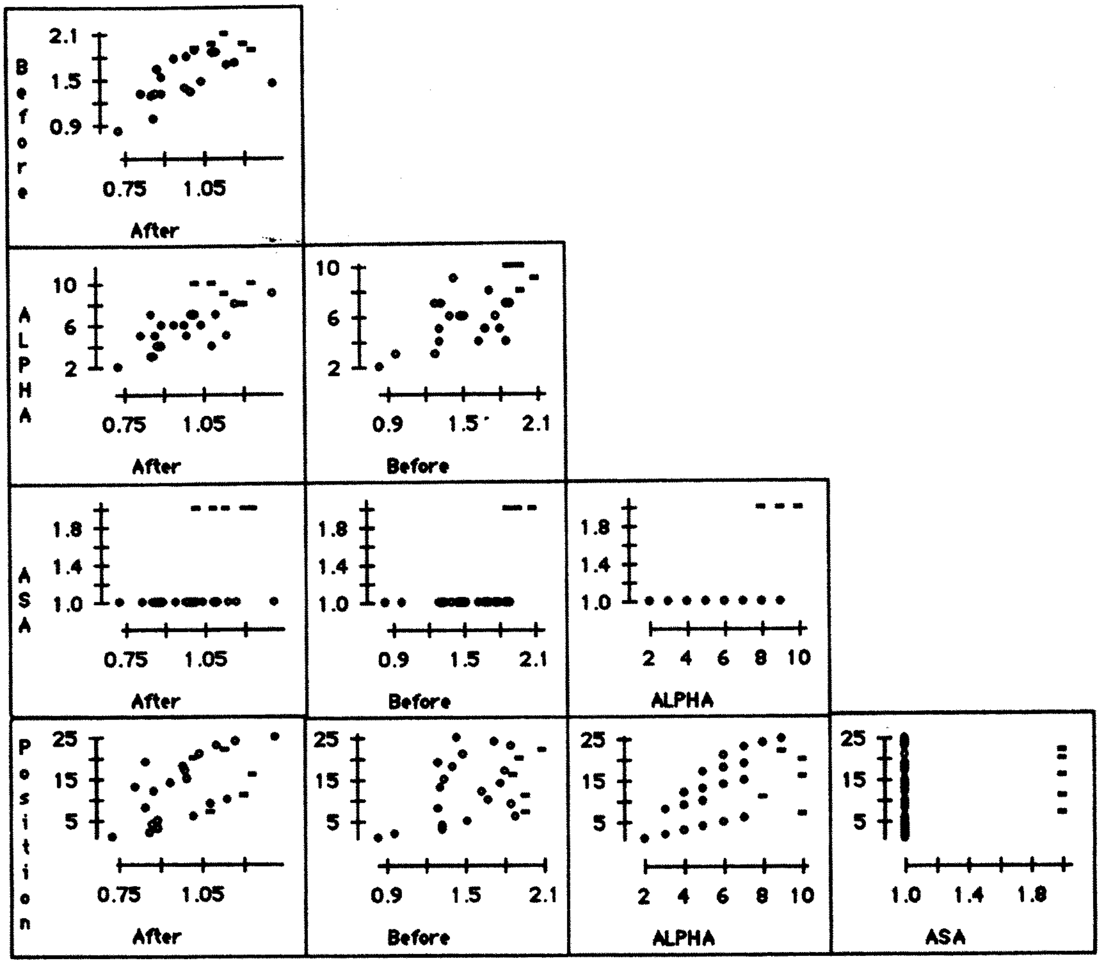

Figure 1. Scatterplot matrix for before and after reaction times (Klahr et al., 1983, Experiment 1) and structural variables from ALPHA, alphabetic search by association (ASA), and letter position. Dashed points correspond to chunk boundary crossings; open circles, to all other letters. ASA has been modified from 0/1 weights to 2/1 association strengths for display convenience.

how their associative strengths are transformed into specific response times.

The models can be compared in a more straightforward fashion than the approach used by Scharroo et al. (1994). Scharroo et al. discussed three structural components: the number of steps in ALPHA, the association strength in ASA (0/1), and the alphabetic position of the probe. Each of these variables is correlated with the other two and with the mean RTs from Klahr et al. (1983)[4] (see Table 1). ALPHA's correlation with the after data is much higher than ASA's (.72 vs. .45) and slightly higher on the before data (.67 vs. .60). However, as shown in the scatterplot matrix in Figure 1, similar correlations derive from qualitatively different underlying distributions (cf. scatterplots for ALPHA-before and ASA-before).

A more precise understanding of the relation among the three components can be obtained by running simple and multiple regressions of the components against the before and after data from Klahr et al (1983). These regressions reveal not only how much variance is accounted for, but also what the parameter estimates are for the amount of time each component contributes to the overall RT (based on the beta weights in the regression equation). The results are shown in Table 2. Each row in the table can be read as the regression equation for the simple or multiple regression indicated in that row. For example, Row i shows the following regression equation for RTs on the after task: RT = 718 + 46*(next), where *next* is the number of basic steps executed by ALPHA in running through the alphabetic structure. This model accounts for 49% of the variance in mean RTs for Klahr et al.'s Experiment 1.[5] The

[4] To keep this commentary brief, I use only Klahr et al.'s (1983) Experiment 1 here. It was actually the tougher of the two experiments in Klahr et al. because it gave a slightly poorer fit to ALPHA in the original Klahr et al. article.

[5] The regression results reported here are slightly different from those in Klahr et al. (1983) due to a different (and more accurate) regression program.

Table 2
Regression Results for Several Models and Klahr et al.'s Experiment 1

Row	Constant[a]	ALPHA β[b]	ALPHA *SE*	ASA β[c]	ASA *SE*	Position β[d]	Position *SE*	Adjusted R^2 (%)
				After				
i	718	46****	9					.49
ii	808			162*	67			.17
iii	733	50***	13	−36	74			.48
iv	854					11*	3	.30
v	709	35	18	9	83	5	4	.48
				Before				
vi	988	95***	22					.43
vii	1,011			474**	133			.33
viii	905	70*	31	196	172			.43
ix	1,344					18*	8	.13
x	872	50	43	259	197	7	10	.41

Note. All entries except R^2 multiple correlations are shown in milliseconds.
[a]Intercept of regression line. [b]Beta weights for each *next* in ALPHA. [c]Beta weights for modified associative model in which all associations equal 1, except for boundary crossings, which are equal to 2. Boundary crossings are based on the Klahr, Chase, and Lovelace (1983) segmentation. [d]Beta weights on alphabetic position of probe letter.
*$p < .05$. **$p < .01$. ***$p < .001$. ****$p < .0001$.

ASA model, by itself, accounts for only 17% of the variance (Row ii), which is even less than a simple model that predicts increasing RTs with alphabetic position (Row iv). Combining components (Rows iii and v) does not improve the overall adjusted multiple squared regression R^2. ASA does somewhat better with the before data (Row vii), but it does not do as well as ALPHA (Row vi).[6]

Toward a New Model of Alphabetic Retrieval

The alphabet is "a common long list, with little explicit structure, learned very early and used throughout life" (Klahr et al., 1983, p. 462). Have Scharroo et al. (1994) increased our knowledge about how that list is structured and processed? I think not. For one thing, although they dislike the hierarchical structure of ALPHA so much that they have repeatedly disparaged it as "implausible," they failed to account for a substantial literature on the hierarchical storage of both short and long serial lists (Anderson & Bower, 1973; Broadbent, 1975; Chase & Ericsson, 1982; Estes, 1972; Lee & Estes, 1977; Staszewski, 1988). Furthermore, their model is silent on two issues that bear on the validity of RT models such as ALPHA. The first neglected issue is the absolute magnitude of underlying processes as revealed by the parameter estimates in the regression models. For example, Klahr et al. noted that their parameter estimates are consistent with many other estimates of the rate of speeded memory search. The other missing component of Scharroo et al.'s analysis is a discussion of error rates. Scharroo et al. reported neither the rate of error nor their procedure for dealing with error trials. Furthermore, they ignored an obvious implication of their associative model: that errors should occur disproportionately at the weakly associated letters.

One might have hoped that, coming more than 10 years after Klahr et al. (1983) first proposed their model, an article purporting to increase our knowledge about the process of alphabetic retrieval would extend and integrate the literature on memory retrieval rather than simply proposing an either–or test between two models, both of which are obviously oversimplifications of the cognitive processes involved in these tasks. It is not surprising that many letters of the alphabet have strong associations with their immediate neighbors, particularly in the forward direction. But what happens when those associations are weak? Although Scharroo et al. (1994) concluded that ALPHA is "unnecessarily complex," they also believe that when "associative strength is low, direct retrieval may fail. In those cases the letter at issue is retrieved by active search" (pp. 243–244). This is precisely the process that Klahr et al. attempted to explicate in ALPHA. It is clear that a complete model of this process will require a combination of the features of both ALPHA and ASA.

In recent years, explicit models of this type of two-stage decision process have been proposed for a variety of speeded retrieval tasks (Reder, 1987; Reder & Ritter, 1992; Siegler, 1988; Siegler & Shrager, 1984). For example, in Siegler and Shrager's model of the retrieval of the sum of two single-digit arguments, the first step is to try a direct association, but if the association strength between the arguments (m and n) and the result ($m + n$) is weak, the model enters a second stage in which an iterative process is executed. It would seem that such a two-stage model, incorporating both a distribution of direct associations as well as an ALPHA-like hierarchical search, could best describe what people do when they attempt alphabetic retrievals.

[6] Although Scharroo et al. (1994) did not address it directly, one might ask why one should take seriously a model that accounts for this modest amount of variance. My hunch is that the reliable variance in the data is not much more than this and that ALPHA is capturing just about all there is. However, that remains to be demonstrated.

References

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* New York: Holt, Rinehart & Winston.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In R. A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory* (pp. 3–18). New York: Wiley.

Browman, C. P., & O'Connell, D. C. (1976). Sequential phonological effects in recitation times. *Bulletin of the Psychonomic Society, 8,* 37–39.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *Advances in learning and motivation,* pp. 2–58 (Vol. 16). San Diego, CA: Academic Press.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory,* (pp. 161–190). New York: Holt, Rinehart & Winston.

Klahr, D., Chase, W. C., & Lovelace, E. (1983). Structure and process in alphabetic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 462–477.

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior, 16,* 395–418.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19,* 90–138.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 435–451.

Scharroo, J., Leeuwenberg, E., Stalmeier, P. F. M., & Vos, P. G. (1994). Alphabetic search: A comment on Klahr, Chase, and Lovelace (1983). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 236–244.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117,* 258–275.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.

Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 71–128). Hillsdale, NJ: Erlbaum.

Received January 11, 1993
Revision received May 7, 1993
Accepted May 8, 1993 ■

Six Editors Appointed, 1995–2000

The Publications and Communications Board of the American Psychological Association announces the appointment of six new editors for 6-year terms beginning in 1995. As of January 1, 1994, manuscripts should be directed as follows:

- For the *Journal of Abnormal Psychology*, submit manuscripts to Milton E. Strauss, PhD, Department of Psychology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7123.

- For the *Journal of Applied Psychology*, submit manuscripts to Philip Bobko, PhD, Editor, *Journal of Applied Psychology*, P.O. Box 130, Skillman, New Jersey 08558.

- For the *Journal of Comparative Psychology*, submit manuscripts to Charles T. Snowdon, PhD, Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, Wisconsin 53706-1696.

- For the Attitudes and Social Cognition section of the *Journal of Personality and Social Psychology*, submit manuscripts to Arie W. Kruglanski, PhD, Department of Psychology, University of Maryland, College Park, Maryland 20742.

- For *Professional Psychology: Research and Practice*, submit manuscripts to Patrick H. DeLeon, PhD, JD, Editor, *Professional Psychology: Research and Practice*, APA, Room 3084, 750 First Street, NE, Washington, DC 20002-4242.

- For *Psychological Review*, submit manuscripts to Robert A. Bjork, PhD, *Psychological Review*, Department of Psychology, University of California, Los Angeles, California 90024-1563.

Manuscript submission patterns make the precise date of completion of 1994 volumes uncertain. The current editors, Susan Mineka, PhD; Neal Schmitt, PhD; Gordon G. Gallup, PhD; Abraham Tesser, PhD; Ursula Delworth, PhD; and Walter Kintsch, PhD, respectively, will receive and consider manuscripts until December 31, 1993. Should any 1994 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 1995 volumes.