Higher-order dimensions in concept identification[1]

MADELEINE SCHLAG-REY AND PATRICK SUPPES, *INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES, STANFORD UNIVERSITY, Stanford, Cal. 94305*

In a complex concept identification task, solvers and nonsolvers tended to diverge in overall but not in trial-by-trial changes of hypothesis. Neither error frequency nor specific content of feedback was a major factor in concept attainment. The results suggest that Ss sample from subsets of hypotheses structured by higher-order dimensions.

Tests of all-or-none hypothesis-testing models of concept identification have usually been confined to predictions of overt responses rather than specific hypotheses (Trabasso & Bower, 1966). On the other hand, experiments inferring hypotheses from sequences of responses obtained on blank trials (Levine, Miller, & Steinmeyer, 1967) appear limited in scope because of the small number of hypotheses available to S.

We have attempted to overcome both of these limitations by observing changes of hypotheses based on fairly abstract dimensions, under no memory strain.[2]

Method. On each trial, S was requested to classify correctly in two groups 62 patterns made up of all possible combinations of Ds and Ks in strings of 1, 2, 3, 4, and 5 letters. The patterns were typed together on single sheets of paper, one sheet being used per trial. The sequence of patterns was constant for the individual, randomized for the group in 31 different lists.

After S had classified all patterns by either circling or crossing out, S selected one of them for which E indicated the correct response (circling or crossing) on a new sheet which included the past information too. S kept all sheets completed. S was told that the experiment would stop upon his statement of the solution rule (circling all patterns ending with DD or KK and only those). S could attempt to formulate the solution rule at any point, but if the formulation was wrong the experiment continued until the correct solution or the time-limit (1 h) was reached.

Group 1 (155 college students) was run according to this procedure. Group 2 (155 identically recruited Ss) was run concomitantly under the same conditions except that each S received from the start the information available by trial 7 to a matching S of Group 1.

Results. Seventy-two Ss in Group 1 and 80 in Group 2 solved the problem. Assuming that Group 2 started on trial 7, the mean number of trials, respectively, for solvers of Group 1, of Group 2, nonsolvers of Group 1, of Group 2 was: 8.56 (SD = 2.87), 9.60 (SD = 2.29), 11.75 (SD = 2.68), 12.93 (SD = 2.32). Fig. 1 shows the cumulative frequency of finding the solution as a function of the trials. Obviously, concept attainment was not a simple increasing function of the amount of information received. Other factors investigated were the following.

Error frequency. Information could infirm S's hypothesis in two ways: by (a) conflicting with the response to one pattern, or by (b) conflicting with a statement of the rules. On the first block of seven or less presolution trials of Group 1, the average proportion of (a) was .49 for solvers, .51 for nonsolvers, the average proportion of either (a) or (b) was .60 for solvers and .63 for nonsolvers; computed over the last three trials of that block, they were .41, .48, .59, and .64, respectively. Clearly, the solvers did not profit from being informed about more errors than were the nonsolvers.

Specific information. Forty-nine per cent of solvers of Group 2 were matched with solvers of Group 1, 56% of nonsolvers matched with nonsolvers (X^2 = .340, df = 1, p > .50). Had specific collections of instances and noninstances of the concept been instrumental in solving the problem, the Ss of Group 2 who started with a solver's information would have shown this to be beneficial. The conclusion that it was not is tempered by the fact that 17 out of the 22 Ss of Group 2 who solved on their first trial had been matched with solvers.

Specific hypotheses. All obtained classifications (C) were tested through a multi-stage program for the consistent use of some rule. It appeared interesting to distinguish four types of hypotheses based on the combinations of two independent binary properties of the rules as follows: Property A: C depends on specific letters, i.e., on D as opposed to K; $\bar{A}$: C is unaltered when Ds are converted into Ks and Ks into Ds; Property B: C depends on perceiving the whole pattern as relevant for the recognition of some attribute (a sequential arrangement of a count of letters, absolute or relative to the length of the pattern); $\bar{B}$: C depends on abstracting either one or both ends of the pattern as the sole parts containing relevant information. Thus, a classification based on a count of Ks was included in Type AB; on the homogeneity of letters, in $\bar{A}B$; on the beginning letter, in $A\bar{B}$; on the criterion, in $\overline{AB}$. A residual category, O, subsumed inconsistent classifications and complicated rules. (However, when the length of the pattern was used in conjunction with another rule, the latter determined the Type.)

Figure 2 shows the proportions of AB, $\overline{AB}$, $A\bar{B}$, $\bar{A}B$, and O hypotheses pooled on presolution trials 1 through 7 for Group 1 and presolution trial 7 (beginning trial) for Group 2. On trial 1, solvers resembled nonsolvers except that they selected a smaller proportion of Type O hypotheses. As a function of the trials, other discrepancies appeared and increased so that, at trial 7, solvers of Group 2 were closer to solvers of Group 1 than to nonsolvers of Group 2. The rising curves of Type O hypotheses reflect an increasing tendency to use disjunctive rules to hold partially relevant hypotheses. The upward trend of $\overline{AB}$ and downward trend of AB are better seen in the backward learning curves of Fig. 3. Curiously enough, on trial n-1 (last incorrect hypothesis) the proportion was .05 for AB while approximately .20 for each of $A\bar{B}$, $\bar{A}B$, and $\overline{AB}$. A finer description of the hypotheses would further disclose their tendency to draw nearer to the criterion by shifts to its values on various dimensions. For instance, given Type B, the conditional probability of hypotheses based on ordered subsets of letters vs unordered subsets increased from .45 on trial 1 to .78 on trial n - 1. Similarly, given $\bar{B}$, the conditional probability of hypotheses based on ending letters rose from .04 to .65.

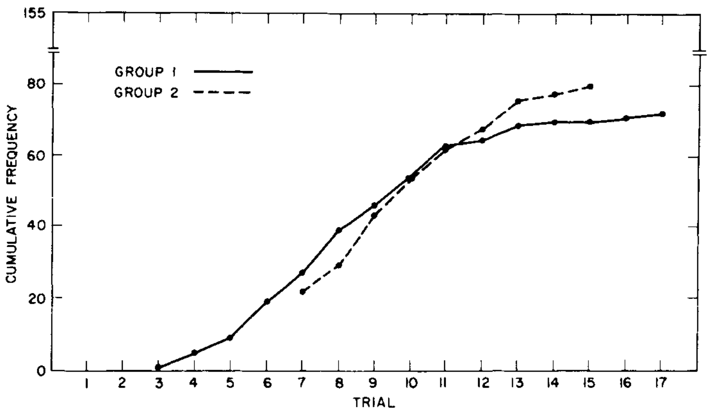

Fig. 1. Cumulative frequency of solution-trial.

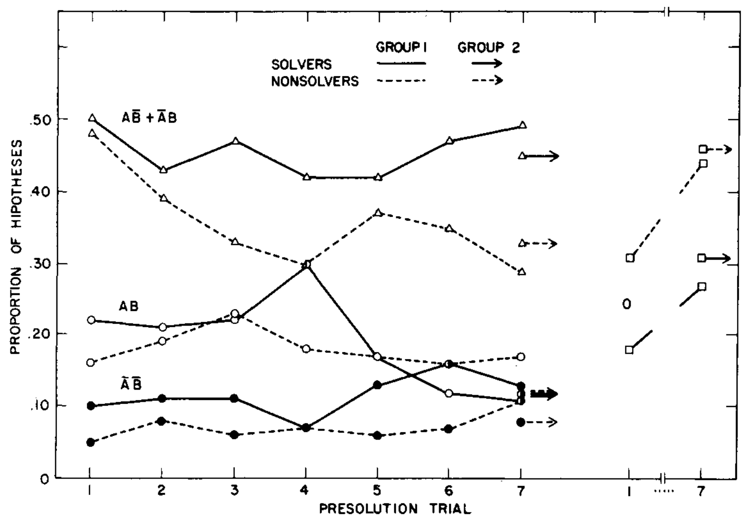

Fig. 2. Proportions of hypotheses of Type AB, $\bar{A}\bar{B}$, $A\bar{B}$ + $\bar{A}B$, and O.

Specific transition processes. The very small probability of going in one trial from Type AB to the solution could mean that S scanned a list on which, for some reason, the criterion was more distant from AB than most other hypotheses. It could also mean that S tended to sample in a restricted space centered on the last held hypothesis, h_{n-1}, the distance between h_n and h_{n-1} being a function of the number of independent dimensional shifts. Were this true for actual changes of hypotheses, one would expect the transition probabilities involving O shift (change within Type), 1 shift, and 2 shifts to be in decreasing order. Pooling the appropriate transitions yielded the corresponding mean probabilities: .45, .41, .14 for solvers and .51, .37, .12 for nonsolvers (Group 1, first six transitions); on the basis of chance alone, given the number of ways in which O-, 1-, or 2-shift transitions could occur, the expected values were .25, .50, .25.

Discussion. Solving a complex identification problem seems to depend more on the type of wrong hypotheses entertained than on the number of hypotheses discarded. Moreover, it does not necessarily follow from the shrinkage of the space of tenable hypotheses since S tends to resort to elaborate disjunctions to accommodate conflicting information. Our tentative typology has revealed the influence of higher-order dimensions; although not overwhelming, there is evidence of gradual learning in using them.

Guy, Van Fleet, & Bourne (1966) have proposed that rather than sampling randomly from a pool of tenable hypotheses S may select one or more according to a "noticing order." We suggest instead that S samples from a subset that varies both according to the past reinforcement history and higher-order dimensions. A detailed formulation of these ideas will appear elsewhere.

REFERENCES

GUY, D. E., VAN FLEET, F. M., & BOURNE, L. E., JR. Effects of adding a stimulus dimension prior to a nonreversal shift. *J. exp. Psychol.*, 1966, 72, 161-168.

LEVINE M., MILLER, P., & STEINMEYER, C. H. The none-to-all theorem of human discrimination learning. *J. exp. Psychol.*, 1967, 73, 568-573.

SUPPES, P., & SCHLAG-REY, M. Observable changes of hypotheses under positive reinforcement. *Science*, 1965, 148, 661-662.

TRABASSO, T., & BOWER, G. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. *J. math. Psychol.*, 1966, 3, 163-173.

NOTES

1. This research has been supported by the U. S. Office of Education.

2. A preliminary and partial report of this experiment has appeared in Suppes & Schlag-Rey (1965).

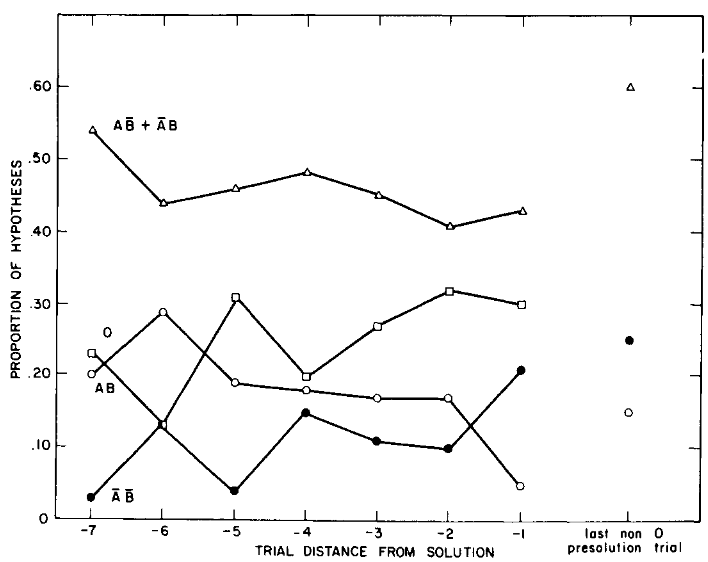

Fig. 3. Backward learning curves. Group 1.