Rule shift, repeated confirmation, and hypothesis subset sampling

TED G. SOMMERS and DENNIS H. HOLDING
University of Louisville, Louisville, Kentucky 40208

and

PAUL FINGERMAN
American Institutes for Research, Washington, D.C.

Three experiments were conducted to test the application of Levine's transfer hypothesis to arithmetical problem solving. In the first, subjects were run to a long criterion, or a short criterion, and were then instructed that a new problem had begun; testing then continued until the short criterion was reached on the second problem. In a second experiment, subjects were run under the same conditions but were given no instructions on reaching the first criterion. In both experiments, the new problem was either a reversal of the first problem (inside) or from another set of hypotheses (outside). It was found that an inside transfer problem was solved in fewer trials than an outside problem. Those subjects who were given an inside transfer problem preceded by shift instructions or by a long-criterion run solved the new problem in fewer trials than other corresponding groups. In the third experiment, two short-criterion groups were run as before, comparing the reversal transfer problem with a different inside transfer problem. Transfer to either inside problem was found equally easy. The data are consistent with the theory that new hypotheses are sampled from those subsets which yielded prior confirmation.

Sweller (1973) found that in an arithmetical rule-learning problem the number of postsolution trials strongly influenced performance in a succeeding problem. Eleven subjects were presented a series of two digit numbers and asked to respond with a related number according to a rule which they were to discover. Feedback (the correct number) was provided on every trial. In one group, the short-criterion group, subjects were presented stimuli until they made a second consecutive correct response. On this trial they were told "wrong" and given feedback consistent with the shift problem, which had the new but similar solution of reversing the first procedure. A second group, the long-criterion group, was treated identically, except that the shift was introduced after 10 consecutive correct responses. When the solution to the shift problem was a reversal of the solution in the first problem, the long-criterion group solved the second problem in a mean of 1.12 trials, while the short-criterion group required a mean of 14.25 trials.

In the second experiment, where multiple-shift problems were employed, almost all short-criterion subjects failed to solve even the first shift problem, while all but one long-criterion subject solved all the succeeding shift problems very quickly. Again, the solutions to the shift problems were similar to the solution to the first problem. Sweller (1973) interpreted his findings in terms of Levine's (1971) model of nonlearning. This model basically suggests that subjects in any problem-solving task are sampling hypotheses from a subset of the universe of hypotheses. Based on prior experience, this set is selected and used until a solution is found or all hypotheses in the set have been tested and rejected, at which point the set is empty and a new set may be selected. Sweller (1973) argued that the short-criterion subjects rejected the entire (correct) subset of hypotheses when the shift occurred, and failed to use it in subsequent problems. A similar explanation has been applied to sequence learning problems (Fingerman & Levine, 1974; Ress & Levine, 1966).

Levine (1974) has presented a general transfer hypothesis which includes the cases of nonlearning and insoluble problems. In the basic hypothesis model (1969), he postulated that subjects sample a hypothesis from some universe and respond on the basis of that hypothesis. They maintain their hypothesis when positive feedback is presented; when negative feedback is presented, the hypothesis is rejected and a new one is sampled. The transfer model (1974) added two new axioms. First, the hypothesis sampling is made from a subset of the entire universe. This subset is referred to as a domain, and is considered to be a set of "similar" hypotheses, that is, a set containing elements described and defined by a single grouping principle (perceptual similarity, common verbal label, etc.). "The second new element is the transfer hypothesis itself: When S receives a series of problems, he infers from the first n solutions the domain within the universe from which the n + 1st

solution will be taken. He will start the n + 1st problem by sampling Hs [hypotheses] from this domain" (p. 292). In the context of this transfer hypothesis, Sweller's (1973) argument may be modified as follows. The long-criterion group inferred from the solution to the first problem the correct domain for subsequent problems. The short-criterion group tested a hypothesis from this domain at the end of the first problem, but were denied the confirmation that this was the solution. Thus, they continued to sample from a wider universe of hypotheses during subsequent problems.

Several experiments were conducted, using Sweller's paradigm, in order to examine this explanation more thoroughly. In the first, short- and long-criterion groups were shifted either to a new problem of comparable difficulty from outside that domain. Control groups served to check the equivalence of the new problems.

EXPERIMENT 1

Method

Subjects. A final total of 84 students, serving for credit in introductory psychology courses, provided the data for analysis. Six other students were tested, but failed to solve the first problem and were replaced.

Procedure. The stimuli were the set of 64 numbers used by Sweller (1973), running from 11 through 88, omitting those containing 0 or 9. The numbers were printed in Columbia type, .8 in. (2 cm) high, on the backs of 5 x 3 in. cards. The cards were placed on a stand about 60 cm from the seated subject, one at a time, in a different random order for each subject. The subjects were asked to respond to each stimulus card with an answer between 0 and 110 and were informed that there would be one correct answer in each case, conforming to a rule which they were to discover by trial and error. After each response the experimenter told them whether or not they were correct and the correct answer, leaving the stimulus card in position until the subject was ready to proceed to the next trial. Subjects who did not solve the problem in 64 trials were eliminated from the experiment.

When the short-criterion subjects had made one correct response, they were told they were correct and another card was presented. If this card was correct too, they were told "Wrong—the problem has changed," and testing continued through the second stage of the experiment. Thus, as far as the experimenter was concerned, the criterion was for two successive correct responses. If, as occasionally happened, the second card was wrongly answered, the trials continued as before until two successive correct responses were reached. For the long-criterion groups, 10 correct trials with confirmation were given before the procedure was changed. All experimental groups were then taken to the short criterion on the transfer pattern.

Design. Subjects were assigned to six 14-member groups,

four of which were experimental groups confronted with two successive problems. The first two groups practiced to the short criterion on the first problem. The solution was to add the sum of the two digits to the stimulus number, so that the correct answer to 41 was 46 and the answer to 83 was 94. One of these groups (the inside group) was then shifted to a new problem compatible with the hyposhesis confirmed during their prior learning (the reversal problem: subtract the sum of the two digits from the stimulus number). The other short-criterion group (the outside group) was shifted to a problem whose solution was presumed to be outside the inferred domain, introducing an external constant (add 1 to the second digit and this total to the stimulus number). Groups 3 and 4 were also inside and outside groups, working on the same first and second problems in the same conditions, except that they were required to reach the long criterion before the rule shift was made. Control Groups 5 and 6 were run to the short criterion, each with one of the "new" problems as their first problem. Their function was to check the results of pilot work indicating that the two new problems were of equal difficulty, and that both were of the same level of difficulty for naive subjects as the first problem.

Results

The mean numbers of trials elapsing before the first correct response at criterion are shown in Table 1, together with the number in each group who failed to solve the second problem in 50 trials. It is important to establish that the scores by the two control groups are similar, and that the first problem scores of the experimental groups are comparable. An initial test on the two control means showed that they did not differ significantly $[t(26) = .17, p > .10]$. These two scores were combined for analysis of variance with the four experimental groups' scores on the first problem. As expected, there were no significant differences among the six scores on the first problem $[F(5,78) < 1]$.

The effects of the rule shift are seen in the scores on the second problem of the four experimental groups. The four sets of scores for those who solved the second problem were compared by analysis of variance for the effects of amount of repeated confirmation, or length of prior criterion, and the effects of having the second solution rule taken from inside or outside the previously confirmed hypothesis subset. The analysis showed that length of the prior criterion to which the groups were trained had no significant effect on the number of trials needed for solution of the second problem $[F(1,45) = 1.51, p > .10]$. However, the large difference apparent in Table 1 between the inside and outside pairs of scores is highly significant

Table 1
Mean Trials to Criterion by Inside, Outside, and Control Groups (Experiment 1)

	Short-Criterion Group					Long-Criterion Group					Control Group	
	Problem 1	Trials	Problem 2	Trials	NS*	Problem 1	Trials	Problem 2	Trials	NS*	Problem 1	Trials
Inside	+ (A + B)	20.3	− (A + B)	2.1	0	+ (A + B)	18.1	− (A + B)	1.6	1	− (A + B)	18.2
Outside	+ (A + B)	21.9	+ (B + 1)	17.8	2	+ (A + B)	20.7	+ (B + 1)	15.5	4	+ (B + 1)	17.4

Note—*Problems are symbolized by designating the stimulus number = 10 A + B.*
*NS = nonsolvers.

[F(1,45) = 33.10, p < .0001]. There was no significant interaction.

Discussion

It was expected that, although no subjects were actually told they had solved the problem, the 10 confirmatory trials experienced by the long-criterion subjects would cause them to infer a domain consisting only of solutions similar to the one they had previously found. In contrast, the short-criterion group should either infer no new domain or, as Sweller (1973) suggested, infer a wider domain excluding hypotheses of the type they had most recently tested. Although the control groups showed the two shift solutions to be of approximately equal difficulty, both the long- and short-criterion groups found the inside solution to be considerably easier than the outside solution.

This departure from Sweller's results might be due to the effect of the shift instructions upon the "solution experience." Sweller's results implied that the short-criterion group had almost no solution experience and so made no inference, or a wrong one, about a new domain; in the present study, both short- and long-criterion subjects behaved as though they had experienced a solution. Since Sweller apparently shifted his subjects from one problem to the next with no signal, his long-criterion subjects were no doubt surprised after a run of 10 correct feedbacks and assumed that a new problem had begun. His short-criterion subjects, on the other hand, had no such pattern enabling them to detect the shift and thus lacked a solution experience. However, Polson, Dunham, and Reeve (1971) have shown that the form taken by the shift instructions plays a major part in determining concept identification performance in a new problem. In the present study, both groups were told that a new problem had begun. If this constitutes a sufficient signal to the short-criterion subjects that the problem was solved, they should infer a narrow domain and the results would be consistent with Levine's model.

The next experiment was therefore designed to compare transfer performances in the absence of shift instructions. Short- and long-criterion groups were run as before, to inside and outside problems, but the change-over instructions were omitted. Removing the instructions should differentially affect performance by the short-criterion subjects.

EXPERIMENT 2

Method

The subjects were 56 undergraduates, again serving for credit. These subjects were run in four groups of 14 members each, to the long or short criterion on the first problem and to inside or outside solutions for the second problem. The stimulus materials and procedure were identical to those in Experiment 1, except that the shift instructions were omitted. Thus, on reaching criterion on the first problem, a subject was merely told "Wrong" and trials continued until the second problem was solved.

Results

The mean trials to criterion for both problems and the number of nonsolvers are shown in Table 2. The pattern of results resembles Experiment 1, although the number of trials required by the short-criterion group for the inside problem increased, as predicted. Scores for the short-criterion/inside groups in Experiments 1 and 2 were compared by t test to examine the predicted effect of instructions. The two scores differed significantly [t(26) = 2.48, p < .05].

Table 2
Mean Trials to Criterion for Uninstructed Groups (Experiment 2)

Group		Problem 1	Trials	Problem 2	Trials	NS
Short Criterion	Inside	+ (A + B)	15.2	− (A + B)	7.4	0
	Outside	+ (A + B)	14.7	+ (B + 1)	19.1	1
Long Criterion	Inside	+ (A + B)	15.0	− (A + B)	1.1	0
	Outside	+ (A + B)	16.4	+ (B + 1)	16.6	4

Note—NS = nonsolvers.

An overall analysis of variance included all four groups from Experiment 2 and the four corresponding groups from Experiment 1. The effect of presenting inside vs. outside problems was again highly significant [(1,92) = 57.82, p < .0001]. Neither the main effects of criterion length [F(1,92) = 2.4] nor the main effect of instructions [F(1,92) = 2.5] was significant. None of the two-way interactions, nor the three-way interaction was significant.

Discussion

Presumably, removing the shift instructions had no effect on the long-criterion groups because the extended criterion trials had served as solution confirmation, rendering the instructions redundant. The short-criterion/outside groups were also unaffected by the instructions which, although signaling that a new problem had begun, only reduced the indefinitely large number of possible hypotheses by one. Andrews, Levinthal, and Fishbein (1969) have shown, by inserting hypothesis probes before and after a first feedback trial, that after a first disconfirmation subjects still choose a next hypothesis from the same domain. In the present case, adding the shift instructions to the disconfirmation on the second criterion trial led to little relative change in the probability of selecting correct hypotheses from outside the original domain, so that in either group the outside problem remained difficult.

The differences obtained by removing shift instructions from the short-criterion/inside group, although numerically small, confirm the Levine model in principle. In this case, it may be assumed that subjects taken to the short criterion, without shift instructions to confirm a solution, continue to select from a relatively broad domain compared with the instructed subjects, who infer a narrower subset. The uninstructed group clearly made some inferences concerning the domain, since their scores were lower than those for the outside or control groups. However, it seems possible that the reversal problem selected by Sweller (1973) and adopted in the present experiments constitutes a special case, since reversal learning often has atypical properties. Hence, it seemed desirable to avoid the implications of a reversal explanation by testing a neutral inside problem. Experiment 3 therefore compares instructed and uninstructed groups on an equally difficult, nonreversal problem.

EXPERIMENT 3

Method

The subjects were 42 undergraduates, recruited as before, assigned to three groups. The procedure remained as in the previous two experiments. Two groups performed the original addition problem, to the short criterion, and were then shifted to a new inside problem. Only one of these groups received the shift instructions. The new problem (add the absolute difference between the two digits to the original number) was selected as equivalent

in difficulty to the earlier problems during pilot work. The third group again served as a control, performing only the new problem.

Results

The data are shown in Table 3. The control group score is comparable with those of the equivalent groups in Experiment 1. Analysis of variance showed no significant differences among the three control groups $(F < 1)$. The effect of instructions upon solution of the new problem was examined by t test, which showed a significant difference between the instructed and uninstructed groups $[t(25) = 2.92, p < .005]$.

Further checks were made between the comparable groups in Experiments 2 and 3. There was no significant difference between the instructed $(t = .37)$ nor the uninstructed $(t = .48)$ short-criterion/inside groups in the two experiments. Confirming the earlier inside-outside difference, the uninstructed group score in Experiment 3 differed from the uninstructed/short-criterion group score in Experiment 2 $[t(25) = 2.32, p < .025]$.

Discussion

The results of Experiment 3 suggest that it makes little difference whether the solution to the transfer problem is a reversal of the previous rule or a different manipulation of the presented number. Provided that no outside terms are introduced into the solution, subjects appear equally likely to select nonreversal hypotheses involving adding or subtraction of the component digits as to test reversals. Such hypotheses seem to form a domain of the type envisaged by Levine (1974).

The earlier two experiments, together with the corresponding data of Sweller (1973), seem consistent with this type of analysis. Subjects who are given overlearning confirmation trials on one hypothesis seem "set" for its domain, requiring few trials to discover a similar, new solution but many to discover an outside solution. Subjects transferred after reaching the short criterion are likely to sample from the same domain if confirmation is provided by a shift instruction, but otherwise find relatively greater difficulty with a new problem from that domain. In these experiments the difficulty for uninstructed/short-criterion subjects was nevertheless greater when outside, rather than inside, problems were presented. Presumably, as Huesmann and Cheng (1973) report, subjects given little information begin by considering simple mathematical hypotheses, and even without extended confirmation trials or shift instructions, are less likely to select hypotheses involving outside constants. In fact, all groups transferring to the outside problem found it almost equally difficult. Since the control groups

Table 3
Mean Trials to Criterion: Nonreversal Problem (Experiment 3)

Group	Problem 1	Trials	Problem 2	Trials	NS
Instruction	+ (A + B)	11.4	+ I A − B I	2.5	1
No Instruction	+ (A + B)	12.7	+ I A − B I	8.7	0
Control	+ I A − B I	18.0			

Note—NS = nonsolvers.

found all the problems comparable when given first, it must be assumed that the experience of even the uninstructed/short-criterion group made some contribution toward specifying a domain.

Although the data do not provide a critical test of the Levine model, they form a coherent pattern qualitatively consistent with it. In particular, it appears that the notion of hypothesis testing by domains permits useful predictions to be made for arithmetical rule learning, despite the theoretical problems posed by the indefinitely large number of potential mathematical hypotheses.

REFERENCES

ANDREWS, O. E., LEVINTHAL, C. F., & FISHBEIN, H. D. The organization of hypothesis-testing behavior in concept-identification tasks. *American Journal of Psychology*, 1969, **82**, 523-530.

FINGERMAN, P., & LEVINE, M. Nonlearning: The completeness of the blindness. *Journal of Experimental Psychology*, 1974, **102**, 720-721.

HUESMANN, L. R., & CHENG, C. M. A theory for the induction of mathematical functions. *Psychological Review*, 1973, **80**, 126-138.

LEVINE, M. Neo-noncontinuity theory. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation.* New York: Academic Press, 1969.

LEVINE, M. Hypothesis theory and nonlearning despite ideal S-R reinforcement contingencies. *Psychological Review*, 1971, **78**, 130-140.

LEVINE, M. A transfer hypothesis whereby learning-to-learn, einstellung, the PREE, reversal-nonreversal shifts, and other curiosities are elucidated. In R. L. Solso (Ed.), *Theories in cognitive psychology: the Loyola symposium.* Potomac, Md: Lawrence Erlbaum, 1974.

RESS, F. C., & LEVINE, M. Einstellung during simple discrimination learning. *Psychonomic Science*, 1966, **4**, 77-78.

SWELLER, J. Effect of amount of initial training on concept shift problems. *Journal of Experimental Psychology*, 1973, **99**, 134-136.