

# The effects of labeling upon hypothesis-testing behavior

RICKARD A. SEBBY

*Kearney State College, Kearney, Nebraska 68847*

and

KENNETH L. WITTE

*University of Arkansas, Fayetteville, Arkansas 72701*

The effects of labeling upon the consistency and the efficiency with which 40 college students test hypotheses were investigated. Subjects labeled after positive, negative, or positive and negative feedback, or they did not label. Negative feedback was experienced following the first, second, or third outcome trial of each problem. More extensive labeling was found to facilitate the consistency with which subjects tested hypotheses. However, hypothesis-testing efficiency was reduced by more extensive labeling. The results were interpreted as reflecting differential processing demands.

One conceptualization of human discrimination learning, proposed by Levine (1966, 1969), posits that adults engage in verbal coding and rehearsal of stimulus cues when testing hypotheses (Hs) in a simultaneous discrimination learning task. Following positive feedback, the subject is assumed to covertly label and repeat those cues that are in the functional H set (coding). When negative feedback is experienced, subjects presumably eliminate previously coded cues and covertly label a new set of cues (recoding).

Recently, Gholson and Beilin (1979) have further defined the role of stimulus coding (verbal and non-verbal) as one of a variety of cognitive subprocesses that are integrated by a central processor. Adult subjects are assumed to engage in verbal coding when testing Hs. However, since this coding is usually covert, one does not know if the subject completely codes the stimuli (all the dimensional cues) on each trial or whether the stimuli are coded consistently across trials.

Previous research by Eimas (1970) and by Gholson, Phillips, and Levine (1973) has attempted to affect the coding and recoding of H information by using perceptual aids. Both Eimas and Gholson et al. found that subjects' consistency in not selecting Hs that had been previously eliminated was facilitated by aids that reduced memory demands associated with coding and recoding. Similarly, Gholson, O'Connor, and Stern (1976) found that stimulus differentiation training enabled children to learn more about the relevant cues in the task environment. This knowledge affected their use of particular H-testing strategies or systems. Thus,

Reprint requests should be sent to Rickard A. Sebbly, Department of Psychology, Kearney State College, Kearney, Nebraska 68847.

H-testing efficiency was increased following such training. These data indicate that one way to aid the coding and recoding processes employed by subjects would be to utilize perceptual aids.

Another possible way of directly influencing the extent of verbal coding and recoding would be to require subjects to overtly label the stimuli. In relation to that possibility, previous research by Sebbly and Witte (Note 1) indicated that overt verbal labeling facilitated the consistency with which Hs were selected for test. This facilitation occurred only when recoding was minimal: Subjects recoded only on one of three outcome trials; negative feedback occurred only on one of three outcome trials, and subjects thus had to recode just once. However, Sebbly and Witte also found that labeling interfered with efficient H testing. Subjects in that study were either required to label only following negative feedback or not to label at all. Results of a system analysis (see Gholson, Levine, & Phillips, 1972, for a description) indicated that subjects required to engage in overt labeling used a less efficient H-testing system most frequently.

On the basis of these findings by Sebbly and Witte (Note 1), this study sought to further examine the effects of verbal labeling upon the consistency and the efficiency with which adult subjects test Hs. This purpose was accomplished by requiring labeling following either positive or negative or positive and negative feedback, in addition to the inclusion of a nonlabeling control condition. Discrimination learning problems involving minimal recoding (negative feedback occurring on one of three trials) were utilized, since such problems had demonstrated the effect of labeling previously. The hypotheses investigated were that extensive verbal labeling (labeling following both positive and negative

feedback), relative to less extensive labeling, would facilitate H-testing consistency and interfere with H-testing efficiency.

## METHOD

### Subjects and Design

Forty undergraduates served as subjects. Ten subjects were randomly assigned to each of the four labeling conditions.

### Stimuli

Each subject was presented a series of 12 bivalued four-dimensional discrimination problems containing blank-trial probes. Stimuli were constructed in a manner similar to those used by Gholson et al. (1972). The four stimulus dimensions were alphabetic letter (A vs. T), color (red vs. blue), size (6.3 and 2.5 cm in height), and line position (above or below the letter).

### Procedure

Three pretraining problems introducing each subject to possible solutions, feedback delivery, blank trials, and the labeling instructions were presented initially. The experimenter said either "correct" or "wrong" when delivering feedback following the subject's choice response. Beginning on the third pretraining problem, and for all problems thereafter, subjects in the three overt labeling conditions verbalized the particular letter, color, size, and line position present in the correct stimulus following either positive or negative or both positive and negative feedback. Experimental problems were each 16 trials in length with blank trial probes, 4 trials in length, inserted between consecutive feedback trials. Experimental problems were arranged in three blocks of three problems; each block contained negative feedback occurring on the first, second, or third outcome trial. The order in which subjects experienced each of these three types of problems within each block was counterbalanced so that each problem type occurred equally often in each position across blocks.

## RESULTS AND DISCUSSION

A procedure similar to that used by Eimas (1970) was employed to determine the consistency with which Hs were tested. The procedure involved determining (1) the proportion of Hs on the first blank-trial probe that were consistent with Outcome Trial 1 (no intervening trials), (2) the proportion of Hs on the second and third blank-trial probes that were consistent with Outcome Trials 1 and 2 and with Outcome Trials 2 and 3, respectively (one intervening trial), and (3) the proportion of Hs on the third blank-trial probe that were consistent with all of the preceding outcome trials (two intervening trials). These data are presented in Table 1. Following an arcsin transformation (Winer, 1971), the three sets of data were examined by separate analyses of variance having labeling as a between-subjects factor and type of feedback received (positive or negative) as a within-subjects factor.

Analyses of the consistency of subjects' responding following no and one intervening trial indicated that labeling significantly interacted with type of feedback in each case [ $F(3,36) = 3.65, p < .05$ , and  $F(3,36) = 3.24, p < .05$ , respectively]. Following no intervening trials, labeling did not significantly affect the consistency of

Table 1  
Mean Percentage of Hypotheses Consistent with Positive (F+) and Negative (F-) Outcome Trials Zero, One, and Two Outcome Trials Removed

Label Condition	Number of Intervening Outcome Trials						Overall
	Zero		One		Two		
	Feedback F+	Feedback F-	Feedback F+	Feedback F-	Feedback F+	Feedback F-	
Positive	88	65	75	69	40	76	69
Negative	92	89	65	73	33	89	74
Both	87	99	68	93	64	90	84
No Label	78	74	60	72	30	82	66
Overall	86	82	67	77	42	84	

response when positive feedback was experienced. However, following negative feedback, the group labeling after both positive and negative feedback was significantly more consistent than any of the other groups. A similar pattern of results was obtained when the consistency of H selection following one intervening trial was inspected. Subjects who labeled following both types of feedback were most consistent following negative feedback.

Analysis of the data following two intervening outcome trials revealed a significant effect of type of feedback received [ $F(1,36) = 29.48, p < .001$ ]. Subjects were more consistent following negative feedback (84%) than following positive feedback (42%), which was attributable to the fact that when negative feedback was two outcome trials removed, only positive feedback intervened. Subjects had only to maintain a particular H following initial negative feedback. Although labeling did not significantly affect the consistency of H-testing behavior, inspection of Table 1 reveals that for positive feedback consistency was greatest for the group that labeled following positive feedback. For negative feedback consistency was greatest for the two conditions that required labeling following negative feedback.

Examination of the efficiency with which subjects test Hs was accomplished by determining the frequency with which subjects used the systems outlined by Gholson et al. (1972). The systems are, in order of most efficient to least efficient, focusing, dimensional checking, H checking, and stereotypes. Due to the restrictions imposed by the Gholson et al. analysis procedure, only 60 of 360 problems were used. The number of problems classifiable in each of the four conditions was found not to differ significantly [ $\chi^2(3) = 3.02, p > .05$ ]. These data are depicted in Table 2.

Stereotypic responding did not occur and was omitted from the table. Examination of the data reveals that, overall, dimensional checking systems occurred most often. This tendency was especially pronounced in the condition that required labeling after both positive and negative feedback. This same condition, along with the condition that required labeling after negative feedback, evidenced the least amount of focusing. Subjects in the

**Table 2**  
**Percentage of Hypothesis-Testing Systems: Focusing (F),**  
**Dimensional Checking (DC), and Hypothesis Checking**  
**(HC) Occurring in Each Condition**

Condition	Systems		
	F	DC	HC
Positive	39	50	11
Negative	20	62	18
Both	20	80	0
No Label	45	45	10
Overall	31	59	10

nonlabeling condition produced the greatest amount of focusing. The percentage of focusing and dimensional checking (45% in each case) found for subjects in the nonlabeling condition was comparable to that found by Gholson et al. (1972) for adult subjects.

The findings of this study are consistent both with previous research (Sebby & Witte, Note 1) and with the Hs advanced at the outset of this study. The seemingly paradoxical finding that labeling facilitates consistency while interfering with efficiency can be understood if one considers that consistent H-testing behavior underlies not only focusing, but also dimensional checking and H checking. Such consistency is necessary if one utilizes any of the prediction systems (systems sensitive to the effects of feedback) discussed by Gholson et al. (1972). However, focusing requires more extensive processing than any of the other systems. These already heavy processing demands may be increased appreciably by the requirement to label; thus, the use of focusing may become very difficult and therefore occur less frequently. Support for the notion that labeling may interfere with the use of more sophisticated strategies comes from research investigating serial recall by Hagen, Meacham, and Mesibov (1970).

Thus, as hypothesized, extensive labeling facilitated H-testing consistency while reducing the efficiency with

which adult subjects tested Hs. One area for further research would be to examine whether labeling affects older and younger age groups in a manner different from that observed in this study.

**REFERENCE NOTE**

1. Sebby, R. A., & Witte, K. L. *The effects of overt verbal coding upon adult hypothesis behavior*. Unpublished manuscript, University of Arkansas, 1977.

**REFERENCES**

EIMAS, P. D. Effects of memory aids on hypothesis behavior and focusing in young children and adults. *Journal of Experimental Child Psychology*, 1970, **10**, 319-336.

GHOLSON, B., & BEILIN, H. A developmental model of human learning. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 13). New York: Academic Press, 1979.

GHOLSON, B., LEVINE, M., & PHILLIPS, S. Hypotheses, strategies, and stereotypes in discrimination learning. *Journal of Experimental Child Psychology*, 1972, **13**, 423-446.

GHOLSON, B., O'CONNOR, J., & STERN, I. Hypothesis sampling systems among preoperational and concrete operational kindergarten children. *Journal of Experimental Child Psychology*, 1976, **21**, 61-76.

GHOLSON, B., PHILLIPS, S., & LEVINE, M. Effects of the temporal relationship of feedback and stimulus information upon discrimination-learning strategies. *Journal of Experimental Child Psychology*, 1973, **15**, 425-441.

HAGEN, J. W., MEACHAM, J. A., & MESIBOV, G. Verbal labeling, rehearsal, and short-term memory. *Cognitive Psychology*, 1970, **1**, 47-58.

LEVINE, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331-338.

LEVINE, M. Neo-noncontinuity theory. In G. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3). New York: Academic Press, 1969.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

(Received for publication January 7, 1980.)