

Predicting feedback effects from response-certitude estimates

THOMAS EMERSON HANCOCK
Grand Canyon University, Phoenix, Arizona

and

WILLIAM A. STOCK and RAYMOND W. KULHAVY
Arizona State University, Tempe, Arizona

To test a control model that related feedback and response-certitude estimates, subjects responded to multiple-choice items, received either verification or elaboration feedback, and answered the items again on both immediate and 1-week-delay tests. Feedback study time was linearly related to a discrepancy index that combined initial item status with certitude estimates. For both initial corrects and errors, conditional probabilities of subsequent correct responding were related in a theoretically relevant manner.

In this study, we examined how response feedback affects behavior in an item-learning paradigm. Our previous work had indicated that feedback effects were a concomitant function of subjects' rated certainty in their initial responses to items (Kulhavy & Stock, 1989). In that case, ratings of "certitude" were conceptually related to other knowledge estimates such as feelings of knowing and the tip-of-the-tongue phenomenon (e.g., Leonesio & Nelson, 1990; Nelson & Narens, 1990).

In the current study, we collected certitude ratings within an experimental sequence in which each subject (1) attempted to answer a multiple-choice item, (2) rated his/her certainty in the response selected, (3) received feedback regarding the correctness of his/her response, and (4) responded to the same item a second time. Within this paradigm, feedback affects the discrepancy associated with each individual response. When a response and its feedback match one another, discrepancy decreases. Conversely, discrepancy increases for responses that do not match their feedback. Discrepancy will be at a maximum for initial errors when certitude is high, fall off for other incorrect responses, and be lowest when certitude is high and the selected response and feedback match each other. We treat discrepancy as a control variable (Powers, 1978), where increases in its magnitude cause subjects to initiate actions that reduce it. Hence, subjects should spend more time on feedback associated with higher discrepancy, and such additional processing time should be linked to a higher probability of their correcting initial error selections.

We conceive of the feedback that follows an item response as consisting of two components, which we label

and denote as "verification" (f_v) and "elaboration" (f_i). Verification refers to the "yes-no" match between feedback and response, and elaboration refers to any additional content contained in the feedback message. Our measure of discrepancy (d) combines the verification component (correct vs. incorrect initial response choice) with the subject's certitude rating (c), to produce the index:

$$f_v \times c = d. \quad (1)$$

In keeping with our previous work, f_v is set equal to $(-1)^m$, where $m = 1$ for initial correct responses and $m = 0$ for initial errors. This substitution has the effect of redistributing certitudes along a scale (d) that increases continuously from high-certitude correct to high-certitude errors. The discrepancy scale indexes the mismatch between internal and external states, and our servocontrol position predicts that subjects will work to minimize such differences. The addition of elaborative material (f_i) to the feedback message should act to increase both feedback study times and probability of initial error correction.

METHOD

Two types of item feedback were varied across two test occasions in a completely within-subject design. The basic arrangement was a 2-feedback (verification vs. elaboration) \times 2-occasion (immediate test vs. 1-week delay) factorial with repeated measures on both factors. The subjects were 26 student volunteers between the ages of 15 and 18 who received \$25 for their participation.

Materials

We selected 200 five-alternative, multiple-choice items from those available in the 1986 College Board Achievement Tests. One hundred items were selected in about equal numbers from History, Social Studies, and World Cultures, with a mean difficulty of .53 ($SD = .17$). The second 100 items were chosen from Biology, Chemistry, and Physics, with an average difficulty of .57 ($SD = .15$). The mean number of lines per item was 8, with an average of 5.4 words per line, and no item included visual materials such as graphs, pictures, or illustrations.

Verification feedback consisted of the statements "Your choice was RIGHT" or "Your choice was WRONG," which appeared above the

Correspondence should be addressed to R. W. Kulhavy, Applied Cognition Facility, Payne Hall 0611, Arizona State University, Tempe, AZ 85287-0611.

stem of the item just completed. Item alternatives were not present during verification feedback. Elaborated feedback consisted of the item stem and all response alternatives (not numbered), with the correct choice designated by an asterisk. In addition, the same right-wrong statement used in verification feedback again appeared directly above the item stem. The subjects rated their certitudes using a 5-point Likert scale with response options ranging from (1) *random guess* to (3) *somewhat sure* to (5) *absolutely sure*, with points 2 and 4 unlabeled. The subjects completed a training module prior to the experiment in which they learned and practiced the precise meanings assigned to the rating-scale points.

Procedure

The experiment was conducted in a computer laboratory using IBM PC XTs, where response selections, certitude ratings, and response latencies were all written automatically to disk. Type of feedback was counterbalanced across subjects, with each subject seeing two sets of 100 items each, with the specific items in each set randomly selected for every person. The subjects participated in groups of 5 to 7, with each feedback order about equally represented at every session. The subjects were assigned to feedback orders by being given a set of disks from a randomly shuffled stack as they appeared for the experiment. The subjects completed on-line practice screens to familiarize themselves with the keyboard and the item-learning task; the training screens were consistent with the type of feedback encountered in the first phase of the study. Following training, the subjects completed the certitude training module, practiced the rating task on the computer, and asked any procedural questions that related to the experiment.

In the first phase, the subjects received the first 100 test items, each on a different screen, with the stem and numbered alternatives at the top of the screen and the rating scale at the bottom. When an item appeared, the subjects were instructed to study it, type in a number (1-5) designating the correct response selection, and then type in a number (1-5) indicating their degree of certitude in the response just selected. A new screen containing the appropriate feedback appeared following the second response. The subjects were allowed unlimited time on all screens. After they had completed the first 100 items, they were allowed a 15-min rest before returning to the terminal to complete the immediate test. For the test, items were presented in a new random order in exactly the same fashion as before, except that feedback screens were omitted. Following another 15-min rest, the subjects completed a new series of practice screens containing the appropriate feedback for the second phase and then completed the remaining 100 test items in exactly the same manner as before. One week following the first session, the subjects returned to the computer laboratory and answered all 200 items in a new random order without feedback, with a 15-min rest interpolated between 100-item blocks.

RESULTS

Items were scored by allowing 1 point for each correct response, and response latencies were recorded in milliseconds. All tests of significance were evaluated at $\alpha = .05$. A 2 (order of counterbalancing) \times 2 (test occasion) analysis of variance on correct items yielded no significant effects, which indicated that order of feedback presentation had little influence on performance across items. All analyses below were also conducted separately by item domain, and in no case did the data for the separate domains differ appreciably from the overall results. The individual domain analyses are available from the authors.

Our main interest in this study was to examine both feedback study times and second responses to items as a function of the discrepancy measure (d) developed in Equation 1 above. To accomplish this task, initial responses were sorted into 10 categories (corrects and errors \times 5 levels of initial certitude). Feedback study times were averaged, and the conditional probabilities reported

were the proportion of items in each category correctly answered on subsequent tests. We required a minimum of five items to be in each category before computing an average feedback time or a probability, and for cases where this standard was not met, adjacent items were pooled according to procedures detailed in Hancock (1990). These points appear individually at noninteger values (the average of items pooled) of the discrepancy measure in the plots below.

Figures 1 and 2 depict average feedback study time plotted by discrepancy for both verification and elaboration feedback, respectively. Contained in each figure are box-plots for data that occurred at integer values of discrepancy, a number of individual data points at intermediate values of discrepancy as obtained by our pooling criterion, and two separate regression lines. In each figure, one regression analysis was conducted for correct responses and one for incorrect responses. In all four cases, only a linear component was significant, with $F(1,90) = 9.95$, $MS_e = 1.88$, $R^2 = .10$ for verification corrects, $F(1,100)$

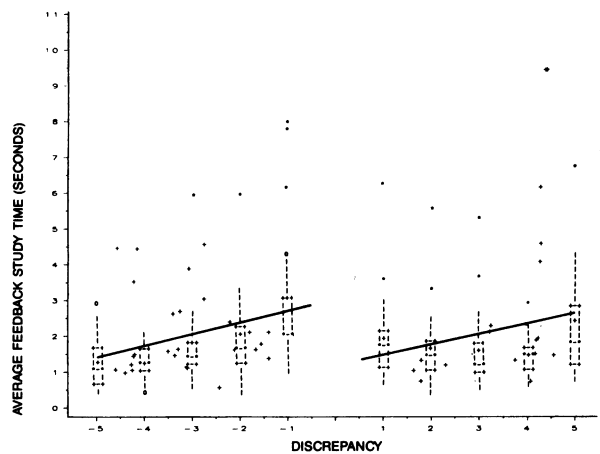


Figure 1. Average feedback study time as a function of discrepancy for items followed by verification feedback.

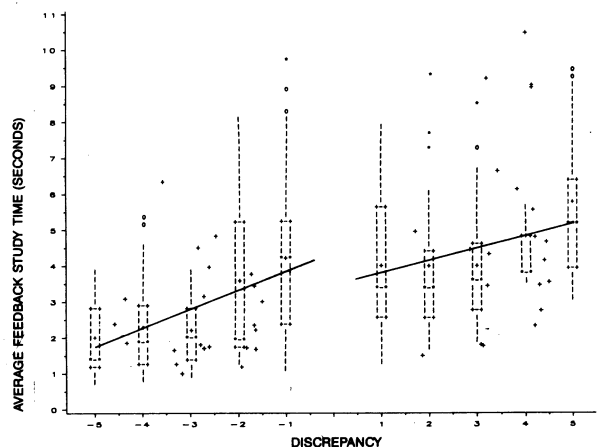


Figure 2. Average feedback study time as a function of discrepancy for items followed by elaborated feedback.

= 5.86, $MS_e = 2.57$, $R^2 = .06$ for verification incorrects, $F(1,88) = 16.99$, $MS_e = 2.90$, $R^2 = .16$ for elaboration corrects, and $F(1,98) = 4.84$, $MS_e = 4.19$, $R^2 = .05$ for elaboration incorrects. These data support our initial hypothesis that feedback study time (effort) increases as a function of the discrepancy index. Furthermore, the positive relation between study time and discrepancy is not restricted to errors and is actually stronger for initially correct responses. As expected, the subjects who saw elaborated feedback spent more time studying the additional information, and the form of the boxplots indicate greater variability of responses associated with elaborated feedback.

Figures 3, 4, 5, and 6 display conditional probabilities of correct responding as a function of discrepancy, for verification and elaboration feedback, on both test occasions. A least squares polynomial regression line has been

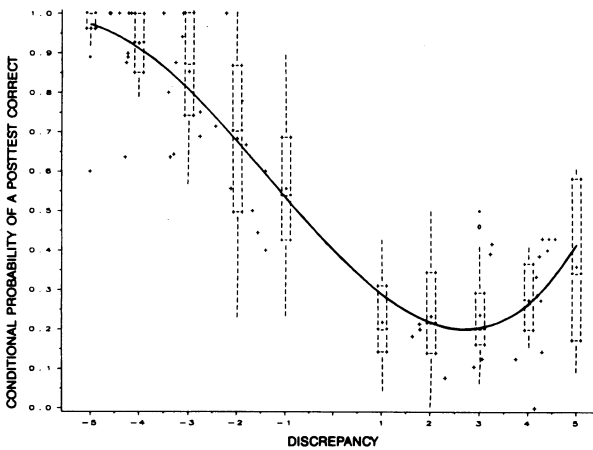


Figure 3. Average conditional probability of a correct response on the immediate test as a function of discrepancy for items followed by verification feedback.

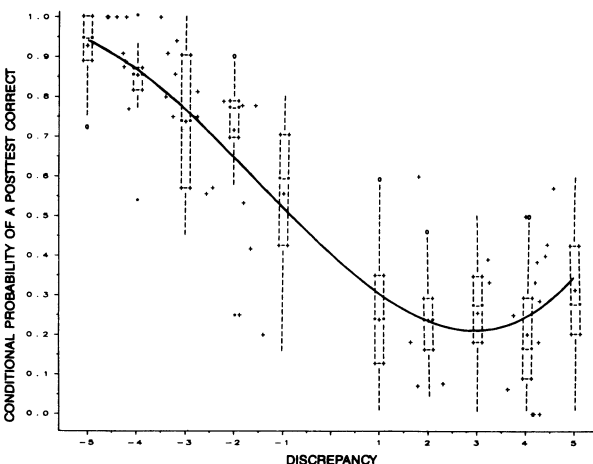


Figure 4. Average conditional probability of a correct response on the delay test as a function of discrepancy for items followed by verification feedback.

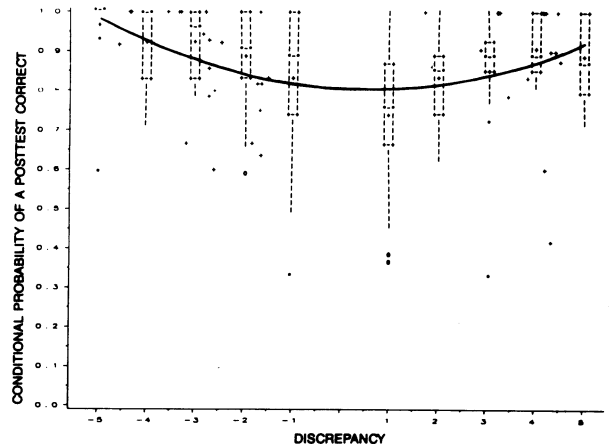


Figure 5. Average conditional probability of a correct response on the immediate test as a function of discrepancy for items followed by elaborated feedback.

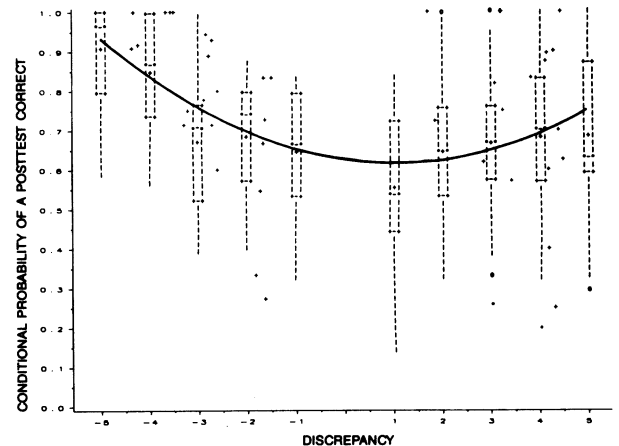


Figure 6. Average conditional probability of a correct response on the delay test as a function of discrepancy for items followed by elaborated feedback.

overlaid on each figure. Polynomial regression analyses simultaneously fitting linear, quadratic, and cubic terms for discrepancy were significant for all four figures, with verification—immediate test, $F(3,190) = 262.61$, $MS_e = .02$, $R^2 = .81$; verification—delay test, $F(3,190) = 192.22$, $MS_e = .02$, $R^2 = .75$; elaboration—immediate, $F(3,186) = 9.83$, $MS_e = .02$, $R^2 = .14$; and elaboration—delay, $F(3,186) = 17.42$, $MS_e = .03$, $R^2 = .22$. However, decomposition of these models revealed significant cubic terms only for verification feedback on both immediate [$F(1,190) = 41.23$] and delay [$F(1,190) = 18.32$] tests. For elaboration feedback, significant quadratic terms were obtained for both immediate [$F(1,187) = 22.71$] and delay [$F(1,187) = 27.98$] tests.

DISCUSSION

Data on both feedback study times and conditional probabilities of correct test responding generally support our initial hypotheses. The

amount of effort a subject expends on a feedback message is a direct function of the discrepancy between external and internal knowledge states. For both verification and elaboration feedback, high-certitude corrects are well maintained to both immediate and delay tests. In the verification functions, conditional probabilities drop off steeply for lower certitude corrects, level off near chance for errors, and increase again for high-certitude errors. Viewing elaborative feedback results in a more gradual decrease, with errors replaced by corrects at a high rate, especially for high-certitude errors. The fact that high-certitude errors tend to reappear following a 1-week delay implicates an interference mechanism, where the certitude is an index of the strength of the original response association. Our conclusion is that a servocontrol model based on certitude estimates provides a useful way for conceptualizing the effects of feedback within an item-learning paradigm.

REFERENCES

COLLEGE ENTRANCE EXAMINATION BOARD (1986). *The college board achievement tests: 14 tests in 13 subjects*. New York: Author.

HANCOCK, T. E. (1990). *Assessing the verification-elaboration components of feedback*. Unpublished doctoral dissertation, Arizona State University, Tempe, AZ.

KULHAVY, R. W., & STOCK, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1, 279-308.

LEONESIO, R. J., & NELSON, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 464-470.

NELSON, T. O., & NARENS, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 1-51). New York: Academic Press.

POWERS, W. T. (1978). Quantitative analysis of purposive systems: Some spadework at the foundations of scientific psychology. *Psychological Review*, 85, 417-435.

(Manuscript received November 15, 1991.)

Notices and Announcements

International Congress of Psychology Brussels, Belgium July 19-24, 1992

The next International Congress of Psychology will be held in Brussels, July 19-24, 1992, under the auspices of the International Union of Psychological Science (IUPsyS).

Congress brochures may be obtained from Brussels International Conference Centre, Place des Expositions, Place de Belgique, B-1020 Brussels, Belgium (FAX 32 2 478 80 23).

Proposals for the Scientific Program, including symposia, may be sent to the U.S. member of IUPsyS, the National Academy of Science. They should be addressed to Pamela Ebert Flattau, USNC/IUPsyS, National Research Council, GH 186, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

APA Applies for Travel Grant for Brussels Congress

The American Psychological Association is applying for funding from the National Science Foundation, the National Institute of Mental Health, and other federal agencies to administer a block travel grant that would allow individual travel awards to be made to US participants in the scientific program of the 25th International Congress of Psychology to be held in Brussels, July 19-24, 1992. The travel grant program will be cosponsored by APA and the US National Committee of the International Union of Psychological Science.

Applicants for awards will be judged according to criteria established by the APA Committee on International Relations in Psychology.

Requests for application packets should be directed to the APA International Affairs Office (750 First Street, NE, Washington, DC 20002-4242). The deadline for receipt of completed application forms is April 1.