OPTIMIZING SUMMARY KNOWLEDGE OF RESULTS FOR SKILL LEARNING *

Richard A. SCHMIDT and Claudia LANGE

* University of California, Los Angeles, USA

Douglas E. YOUNG

California State University, Long Beach, USA


Summary knowledge of results (KR) consists of presenting KR for each of a set of trials only after the last trial in the set has been completed. In experiment 1, motivated by a guidance hypothesis for KR, we searched for an optimal number of trials to be contained in the summary KR. Using a coincident-timing task with summary lengths of 1, 5, 10, and 15 trials, in a delayed no-KR retention test of learning, we found an inverted-U relationship between summary length in acquisition and retention performance, the 5-trial group being most effective. Experiment 2 did not support the hypothesis that the enhanced retention performance of this group relative to the 1-trial group was due merely to a similarity between acquisition and retention conditions (specificity of learning). Both experiments are discussed in terms of a guidance hypothesis to explain how optimizing KR summaries can enhance learning.

The study of knowledge of results (KR) – generally defined as augmented, verbalizable, post-response information about goal achievement in the environment (Schmidt 1988) – has been the chief means of investigating effects of feedback on human performance and learning (Adams 1987; Bilodeau 1966; Newell 1976; Salmoni et al. 1984). Traditional theories of motor learning (Adams 1971; Schmidt 1975; Thorndike 1927) have long emphasized the informational aspects

* This research was supported by Contract No. MDA903-85-K-0225 from the U.S. Army Research Institute (Basic Research) to R.A. Schmidt and D.C. Shapiro.

Requests for reprints should be sent to R.A. Schmidt, Dept. of Psychology, University of California, Los Angeles, CA 90024-1563, USA.
of KR, such that more trials with KR, more precise information contained in each KR, and more immediate KR presentations, should result in a better memory representation of the movement. However, recently this view has been challenged by numerous findings in the KR literature in which the informational quality, or ‘usefulness’ of KR has been altered experimentally.

One of these examples involves a type of KR scheduling referred to as summary KR, and has yielded particularly surprising effects on retention performance. Lavery (1962) had subjects practice ball-propelling tasks during six consecutive days of acquisition, with 20 trials per day. He gave one group of subjects error feedback with respect to movement outcomes at the end of every session, in the form of a graph on which scores for all 20 trials were plotted (Summary group). A second group received KR after every trial (Immediate group), and a third received both forms of scheduling (Both group). The Summary group had lower percent correct scores during practice relative to the Immediate and Both groups. Presumably, Summary KR was somewhat confusing to subjects because all responses in a given summary string, except the last, were followed by at least one other response instead of the appropriate KR, thus degrading the effectiveness of KR when it was eventually received. However, on a series of delayed no-KR retention tests, subjects who had received only summary KR scheduling were more accurate than those in the other two conditions. That is, even though summary KR degraded performance during the acquisition phase when KR was present (and being manipulated), it enhanced learning as measured by no-KR retention performance.

Lavery’s findings, and those obtained from other KR scheduling paradigms such as relative frequency (Ho and Shea 1978; Winstein and Schmidt 1988) and trials delay, whereby KR from a given trial is provided after one or more intervening trials (Lavery and Suddon 1962; Suddon and Lavery 1962), are drawn upon heavily in the guidance hypothesis for KR (Salmoni et al. 1984). This hypothesis assumes that, due to its powerful informational role, especially during the early stages of learning, KR guides the subject toward the appropriate movement pattern. However, the subject may learn to rely upon KR to maintain trial-to-trial performance instead of processing other features of task-related information (e.g., environmental or task cues, response-produced feedback) which promote better task learning. Perhaps this reliance degrades the development of subjective error
detection and correction capabilities. Relying on KR as guidance may also prevent subjects from developing a more stable response pattern, because after receiving KR, they will be more likely to alter their response on the next trial (Bilodeau 1966), thus failing to develop the capability to produce stable, consistent movement patterning. By contrast, conditions that provide less guidance in acquisition — such as summary KR — may force the learner to engage in some additional (or different) information processing during acquisition, resulting in more effective learning as measured by retention performance. The guidance hypothesis challenges the traditional views of learning (Adams 1971; Bilodeau 1966; Newell 1976; Schmidt 1975), which state that any variation of KR which provides more precise, frequent, or accurate information on movement outcome has a positive effect on learning.

The guidance hypothesis predicts that very short summary lengths (i.e., a minimal number of trials contained in one KR summary) are detrimental to learning because they are too guiding, or produce too much dependency on KR. On the other hand, excessively long KR summaries may encourage processes leading to the development of internal error-detection and correction mechanisms, or increased response consistency, but do not provide enough guidance to ensure adequate achievement of the response pattern. Thus, an optimum summary-KR length is predicted, one which balances the positive effects of guidance against its negative consequences.

To test these predictions, Schmidt et al. (1989) began to search for an optimal number of trials contained in KR summary reports (termed summary length) using a ballistic-timing task. Subjects received summary lengths of 1, 5, 10, or 15 trials, and performance in acquisition increased at a slower rate with increased summary length. However, in a delayed no-KR retention test, summary length in acquisition was inversely related to magnitude of mean absolute constant errors. Those subjects with the largest summary length (15 trials) had the lowest mean absolute constant errors, and their performance was less subject to decay across no-KR retention blocks than that of subjects in the other conditions. This finding is not in line with traditional views of learning, which would have predicted most efficient performance for the 1-trial summary condition. Schmidt and his colleagues failed to establish an optimum summary length, as they could not determine conclusively whether 15 trials represented an optimum or whether even longer summary lengths would have been more beneficial.
Experiment 1 was motivated by two issues. First, we were concerned that the summary-KR effect on retention might be limited to 'simple' tasks where only a minimal number of kinematic dimensions are involved in control (Lavery 1962; Schmidt et al. 1989). Second, and most important, we were concerned that no clear optimal summary-KR length (as predicted by the guidance hypothesis) was found in the previous study (Schmidt et al. 1989). It seems reasonable that, given the relatively 'simple' task used, such an optimum might have lain past the 15-trial summary condition. However, if a task with more kinematic dimensions were used, the guiding properties of KR should be relatively more important. Therefore, the optimum summary length should be smaller because more guidance is needed, and this optimum should fall within the range of summary-KR lengths examined (i.e., from 1 to 15 trials).

Experiment 1

Method

Experiment 1 repeated the procedures used in the previous study by Schmidt et al. (1989), using summary lengths of 1, 5, 10, and 15 trials. However, in this experiment we used a task with considerably more kinematic dimensions to be controlled. As before, relative amount learned was evaluated by performance on no-KR retention tests after 10 min and 2 days.

Subjects

Right-handed undergraduate and graduate students (28 female and 32 male) from the University of California, Los Angeles participated. Most received extra credit toward a class in Kinesiology, but 8 subjects were paid $10. Each of the four treatment groups contained 8 male and 7 female subjects, and each group had 2 paid subjects. All were naive to the task used here.

Apparatus and task

The task involved coincident timing, and was a laboratory analog of many common ball-bat tasks (e.g., baseball) where a moving object is to be struck with a bat at some coincidence-point (fig. 1). The appara-
Fig. 1. The coincident-timing apparatus used in experiment 1 and 2.

The apparatus (a detailed description appears in Schmidt and Young 1989, or Young and Schmidt 1990) consisted of a display of 32 red LEDs (1 mm in diameter) aligned 4.5 cm apart in a row, modified from a commercial Bassin timer task. The LEDs were pulsed sequentially, giving the impression of a moving object. The row of lights was mounted in a 6.5-cm by 8.5-cm by 152-cm black metal box, with the row of lights pointing toward the seated subject. The end of the box nearest the subject (48 cm away) was placed on the table surface, and the far end was elevated 45°. The apparent velocity of the light sequence was fixed at 134 cm/s toward the subject, with the duration of the light sequence being 1176 ms.

Above the table surface was an aluminum lever, attached at one end via a vertical axle in ball bearing supports so that it moved freely in the
horizontal plane over the table surface. A vertical handle was attached to the end opposite the axle; a pointer extended horizontally from the lever, and contained a small red lamp. The lever was positioned so that the path of the red lamp as the lever was moved appeared to intersect the line of the LED display at the level of the 30th LED (i.e., third from the subject’s end); this LED was marked with distinctive red tape. A small white lamp to the right of the LED display indicated a starting position for the subject’s movement. A potentiometer, whose output as well as signals from the LED display were sampled at 500 Hz by a DEC 11/23 laboratory computer, was attached to the lower end of the axle.

The task began as the seated subject faced the LED display, grasping the handle with the right hand and placing the right elbow over the axle. The LED display was aligned in the subject’s sagittal plane, aimed downward toward the subject’s chair. The subject located the lever at the white starting lamp situated at the 90° position (with 0° being parallel with the edge of the table surface), to the right of the light display. A tone from the apparatus began a trial, and was followed in 500 ms by the illumination of the first LED in the sequence. The subject’s task was to backswing leftward past the row of lights (toward 180°), reverse direction, and move to intercept the moving lights at the coincidence-point (the 30th light indicated above, situated at 110°), with a follow-through to the right. The overall task goal was analogous to providing as much impact as possible with the ‘ball’, so that it would travel as far as possible after it was struck. The subject could initiate the movement at any time after the first LED was illuminated, could backswing to any distance beyond the LED display up to 180°, and could initiate the forward swing at any time desired.

Subjects were instructed that this goal involved maximizing an overall score, analogous to the impact given to the ‘ball’. Described to the subjects in terms of this ball/bat analogy, maximum impact involved essentially two things. First, the impact was increased as the instantaneous velocity of the bat at the coincidence point increased. Second, the impact increased as the absolute spatial errors decreased. Very small spatial errors resulted in a ‘solid’ contact, and somewhat larger errors provided a ‘glancing’ blow to the ball; a spatial error larger than some absolute value (i.e., a ‘miss’) generated no impact at all. Both velocity and spatial error were used to compute an overall score, analogous to the distance that a ball was propelled, which we
used as the measure of overall performance; the subject’s goal was to maximize this score with practice.

Specifically, the computer measured (a) the instantaneous velocity of the lever as it crossed the coincidence-point, and (b) the absolute spatial error, defined as the distance (in degrees) between the lever’s position and the coincidence point at the moment that the LEDs arrived there. The subject’s score for any trial was the product of the instantaneous velocity and a weighting coefficient (which ranged from 0 to 1), which itself was a function of the spatial error, as shown in fig. 2. The weight was 1 when the spatial errors were between zero and 5° (a solid contact in the analogy), it decreased linearly as the spatial errors increased, and was zero for any spatial error equal to or greater than 20° (a miss). Thus, high-velocity movements were only effective if the subject was spatially accurate, with inaccuracy being penalized in a way defined by the weighting function in fig. 2. We designed the scoring procedure to reflect a trade-off among these competing goals. Research on movements of this type shows that high velocities are achieved only at the expense of large spatial errors (e.g., Schmidt et al. 1988), just as is the case in real-world activities which are modeled here.

Although the lever movement was constrained to one plane, this task had many separate dimensions that could be adjusted simultaneously, such as the time of the backswing initiation, its speed, the location of
the end of the backswing, the time of the forward swing's initiation and its speed, etc. Thus there was no one goal movement pattern, but rather the subject's problem was the discovery of some pattern of action that would maximize the task's overall score. Furthermore, this was a task in which the goal (the computed score) was distinct from the pattern of action necessary to produce it, just as it is in many real-world activities. Hence, KR about the movement outcome (here, the score) did not have a direct relationship with most of the particular features of the movement which produced it (e.g., backswing amplitude, time of peak acceleration in the direction of target), nor did KR tell the subject clearly what needed to be done to make the movement more effective on the next trial (see also Fowler and Turvey (1978) for more on this issue).

Procedures

Subjects were given 90 trials in the acquisition phase and two 30-trial no-KR retention tests — after 10 min and 2 days, respectively. Subjects received one of four summary-KR acquisition conditions (1, 5, 10, or 15 trials), where summaries consisted of a graph of the score against the appropriate number of trials. That is, in the 1-trial summary condition, subjects received a graph with a single data point representing that trial, whereas subjects in the other groups saw graphs with 5, 10, and 15 trials, respectively. For all groups, trials 5, 10, 15, ..., 90 had intertrial intervals of 15 s to allow viewing of the longer summaries (if given). All other trials had intertrial intervals of 10 s. KR delay, the interval from the end of the action to the presentation of the summary graph, was always 6 s. No verbal KR was given. Information about velocity or spatial errors was never provided, although subjects could probably detect these values for themselves to some extent (see Young and Schmidt 1990: fig. 3, no-KR group). Subjects were instructed carefully about the singular task goal, were shown the actual weighting function (i.e., fig. 2), and were apprised of the methods for computation of the overall score. Several inquiries of the subjects at the end of the explanation satisfied us that they had understood the instructions on scoring procedures.

One hypothesis to explain the enhanced learning of summary KR in the earlier experiments involves the notion that the long series of no-KR trials between summaries engenders the development of stronger error-detection capabilities (Salmoni et al. 1984). In this experiment, we
sought evidence about this hypothesis by giving an additional transfer test after the final retention test on day 2. In this test, which was not announced until just before it began, subjects were asked to perform 30 more no-KR trials, but in addition they were to estimate their score after each trial. We derived measures of the subjects' accuracy in subjective estimation in two ways. First, within-subject correlations (across all 30 trials, converted to $Z'$) between objective (actual) and subjective (estimated) errors were computed; the $Z'$ should increase with subjects' increasing sensitivity to their own errors (e.g., Schmidt and White 1972). Second, we computed the mean absolute difference between objective and subjective error across the 30 trials (e.g., Newell 1974); this score should decrease as subjects become more accurate in their estimations. Changes in one or both of these measures should allow inferences about the sources of any improvement in learning as a function of the summary-KR conditions.

Results

In the following sections, we present data on the overall score for the coincident-timing task, which was the global response goal for the subjects. In addition, we analyzed the individual components of this score – the velocity and spatial errors \(^1\) at the target point – to give a more complete understanding of the pattern of effects, but these analyses added little and are not included formally here. Conclusions about effects of summary KR on learning are based solely on relative performance levels in the retention tests, where the temporary (e.g., guidance and/or motivating) effects of the KR manipulations are theoretically equated.

Acquisition phase

Fig. 3 contains mean between-subject outcome scores for six blocks of 15 trials in the acquisition phase. In addition to the general increase

\(^1\) Generally, the velocity paralleled the score measure in both acquisition and retention tests, with the 5-trial summary condition having highest velocity in both retention tests; the 5-trial condition had a 30% higher velocity than the 1-trial condition, $p < 0.05$. For absolute spatial errors, on the other hand, the summary-KR conditions were ordered irregularly in acquisition; but in retention tests, errors generally increased directly (but nonsignificantly) with the summary length in acquisition. Thus, the 5-trial condition's large advantage in terms of velocity dominated the 1 trial condition's small advantage in spatial errors, leading to the 5-trial condition having the highest composite score as seen in fig. 3.
in performance across blocks, $F(5,260) = 3.3$, $p < 0.05$, there was a strong tendency for higher scores as the length of the summary decreased, with the groups being ordered inversely with summary-KR length. Compared to the 1-trial summary condition, the 5-trial summary condition had only slightly lower scores, especially in view of the relatively low scores for the 10-trial and (especially) the 15-trial conditions. This effect of conditions was significant, with $F(3,52) = 5.0$, $p < 0.05$. These differences between groups were present across the acquisition phase, and appeared to increase slightly with later blocks, especially in the 15-trial condition, but there was no significant Groups $\times$ Blocks interaction here, $F(15,260) < 1$. However, the apparently similar slopes of the performance functions (based on 15-trial averages) in acquisition disguised the trial-to-trial changes during the first block. Considering the average trial 1 score for all groups and subjects (before any differential effects of conditions could be manifested) of 56 units, it is evident that the rate of performance improvement – at least in early practice – was slowed as the summary-KR length increased. In any event, longer summaries clearly depressed performance.

**Immediate retention**

The average scores for the immediate (10-min) retention test are shown to the right in fig. 3. There was no significant Blocks effect, $F(5,260) = 1.4$, $p > 0.05$, nor a Blocks $\times$ Groups interaction, $F(15,260) < 1$, so the data were collapsed across all six blocks for presentation.
The main result was that subjects in the 5-trial summary condition performed slightly more effectively than those in the 1-trial condition, both of which far outperformed the 10- and 15-trial conditions. This overall effect of conditions was significant, $F(3,52) = 13.5$, $p < 0.05$. A Tukey's HSD post hoc test revealed that the 5-trial condition had larger scores than the 10- and 15-trial conditions, but that the 1- and 5-trial conditions were not statistically different.

Delayed retention

The mean overall scores for the four conditions on the delayed retention test are shown to the far right in fig. 3. As in the immediate retention test, the Blocks effect, $F(5,260) = 1.2$, $p > 0.05$, and the Blocks $\times$ Groups interaction effect, $F(15,260) < 1$, were not significant, and data were pooled across all six blocks for presentation. The groups were ordered in the same general way as in the immediate test, with the 5-trial group having higher scores than the 1-, 10-, and 15-trial conditions, respectively. However, the difference between the 5- and 1-trial conditions was now much larger than on the immediate retention test, the 5-trial group showing a mean score increase from immediate retention, the 1-trial group showing slightly decreased scores. The Groups effect was significant in the ANOVA with $F(3,52) = 23.5$, $p < 0.05$. A Tukey's HSD post hoc test showed that the 10- and 15-trial conditions were not statistically different, but the 5- and 1-trial conditions were, both of which differed from the 10- and 15-trial conditions.

Recognition test (subjective estimation)

After the retention test of day 2, subjects were asked to perform an additional 30 no-KR trials, in which they estimated their own performances score after each trial. As background for the analysis of the estimation data, the subjects' overall performances as measured by the scores on these 30 trials are shown in table 1. The overall performances on these trials were similar to those reported in the preceding retention test. The 5-trial condition continued to have the highest overall scores (270 score units), although the 1-trial condition was considerably closer than it was earlier (258 score units), with the other groups being considerably lower. This effect was significant, with $F(3,52) = 7.4$, $p < 0.05$. However, post hoc analyses of score data failed to show that the 1- and 5-trial groups were significantly different. It was perhaps not surprising that the difference between these two conditions was slightly
Table 1
Mean performance and subjective-estimation data on the recognition test for the four summary KR conditions.

<table>
<thead>
<tr>
<th>Summary-KR condition in acquisition</th>
<th>1-trial</th>
<th>5-trial</th>
<th>10-trial</th>
<th>15-trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(velocity × weight)</td>
<td>258</td>
<td>270</td>
<td>193</td>
<td>156</td>
</tr>
<tr>
<td>Objective/subjective correlations</td>
<td>0.34</td>
<td>0.43</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td>Objective/subjective absolute differences</td>
<td>182</td>
<td>159</td>
<td>236</td>
<td>279</td>
</tr>
<tr>
<td>(score units)</td>
<td>(83)</td>
<td>(36)</td>
<td>(68)</td>
<td>(53)</td>
</tr>
<tr>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(0.13)</td>
<td>(0.14)</td>
<td></td>
</tr>
<tr>
<td>Note: Within-groups SDs are in parentheses.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

smaller than in the delayed retention test. Between initial acquisition and this recognition test, subjects had received two retention tests (immediate and delayed) under identical conditions. Because subjects can continue to learn under these no-KR conditions in this task (Schmidt et al. 1987: exp. 3; Young and Schmidt 1990), and because the 1-trial condition began these tests at considerably lower levels of proficiency as compared to the 5-trial condition, differential rates of improvement over those trials would be expected. Thus, we do not consider this slight shift in relative performance as contradictory, or as a failure to ‘replicate’ the data from the delayed retention test.

The averaged within-subject correlations between objective and subjective errors, and the average absolute difference between objective and subjective errors, are shown in table 1 for the four summary-KR conditions. The highest correlation between objective and subjective errors was for the 5-trial summary condition (0.43), followed by the 1-, 10-, and 15-trial conditions in that order (0.34, 0.24, and 0.21, respectively), indicating that subjects in the 5-trial condition tended to be somewhat more sensitive to their own errors than those in the other groups. These effects were significant, with F(3,52) = 5.6, p < 0.05. The post hoc test showed that the 1- and 5-trial conditions were not reliably different, but that these conditions had higher correlations than the 10- and 15-trial conditions. Even though there were small differences between groups, the correlations were all relatively low, indicating that none of the groups had developed particularly strong error-detection capacities. The companion measure of the absolute
error in estimation also showed the 5-trial condition to be most accurate (159 score units), followed again by the 1-, 10-, and 15-trial conditions (182, 236, and 279 score units), but these effects were not statistically reliable, \(F(3,52) < 1\).

**Discussion**

The results from experiment 1 are reasonably clear in demonstrating the performance (during acquisition) and learning (during retention) effects of summary KR in this coincident-timing task. First, all of the summary-KR groups improved performance across practice. However, in relation to 1-trial summary KR, increasing the summary-KR lengths caused systematically lower overall performance in acquisition. As in a previous study (Schmidt et al. 1989), the groups with longer summary lengths improved their performance at a slower rate, which was somewhat obscured by the fact that 15 trials are combined to form the blocks plotted in fig. 3. Therefore, by two criteria, namely score values across acquisition as well as rate of score increase, the performances during the acquisition phase were clearly degraded as the summary-KR lengths increased.

On both no-KR retention tests, though, we show an inverted-U effect of the summary-KR length, with the 5-trial condition being the most effective relative to the other conditions. This pattern was considerably different from that shown with the ballistic-timing task (see fig. 1 in Schmidt et al. 1989), where the errors in delayed retention varied inversely with summary length. Also, as in our previous experiment, the pattern of differences among groups showed up most strongly on the delayed retention test. In contrast to the effects reported by Schmidt et al. (1989), the same general pattern was shown on both the immediate and delayed retention tests, although the sizes of the effect were somewhat smaller in immediate retention. These effects on the overall scores were paralleled closely by the mean velocities at the coincidence point, in that the 5-trial summary condition generated faster movements than did the 1-, 10-, or 15-trial summary conditions, with only minimal increases in the absolute spatial errors relative to the 1-trial summary condition (see footnote 1). These findings suggest that the subjects in the 5-trial summary condition were able to operate at high velocities without suffering a decrement in accuracy, implying improvements in movement control relative to the 1-trial group. Given the
nature of this task, with several kinematic dimensions to be controlled, the 10- and 15-trial summary-KR conditions in acquisition probably did not provide sufficient guidance for the learners to approximate the proper movement patterning, leaving them with a deficient basis for performance on the retention tests.

There was some suggestion that subjects in the 5-trial condition had developed a stronger error-detection capability than did subjects in the other groups, revealed both by slightly higher objective–subjective correlations and a tendency toward lower objective–subjective differences (table 1) during the estimation phase of the experiment. Thus, increased sensitivity to movement-related errors in no-KR retention may have been one factor leading to the superior performance of the 5-trial summary condition, as these errors could be subsequently corrected, providing a basis for maintaining accuracy across the 30-trial retention test. Whatever capability for error detection was present, though, it was not particularly strong, indicated by the relatively low correlations and large errors in estimation. (The analogous correlations in the ‘simpler’ ballistic-timing task were as high as 0.85 in Schmidt and White’s (1972) study.) However, estimating performance in the present task seems considerably more complicated than in the unidimensional tasks studied earlier, as it involves many sources of subjective error in arriving at an estimated score. This possibility, coupled with the fact that subjects had received no practice at estimating – or even a warning that estimation would be required prior to the actual recognition test – could explain why subjects were not able to estimate their scores very effectively here. In any case, the hypothesis that increased error-detection capabilities underlie the more effective performances in retention was not contradicted by these data.

**Experiment 2**

Experiment 2 was motivated by our desire to address the processes underlying the learning effects of summary KR shown in experiment 1. The finding that, in the delayed retention test, the 5-trial summary KR condition produced higher retention scores relative to the 1-, 10-, and 15-trial manipulations can be interpreted in terms of the guidance hypothesis (Salmoni et al. 1984; Schmidt et al. 1989), which predicts an optimum summary length. However, an alternative explanation for the
results obtained by Lavery (1962), as well as the more recent ones of Schmidt et al. (1989) and those of experiment 1 (5-trial group versus 1-trial group), is that increased summary length was more effective due to the particular no-KR retention test used. According to this view, groups receiving longer KR summaries in acquisition practiced under a condition relatively more similar to the retention condition (a no-KR test), and thus they would be expected to perform more effectively in a retention test that was most similar to acquisition conditions. This specificity of learning hypothesis (Henry 1968; see also the encoding specificity hypothesis of Tulving and Thomson (1973)) states that, for a given set of conditions in retention, those acquisition conditions that are most similar will maximize learning as measured by performance on the retention test.

To test whether a specificity view can account for the effects of summary KR as observed previously under no-KR retention conditions, a second experiment using our coincident-timing apparatus was conducted. Subjects in two groups obtained either KR after every acquisition trial (1-trial summary), or were given 5-trial summaries, which was our estimate of the optimal summary length for this task determined by experiment 1. Subjects received a 2-day delayed retention test with KR – that is, with KR after every trial during retention. A specificity view predicts that the 1-trial summary group would now have higher mean scores on a delayed KR-retention test because the external conditions present during acquisition and retention more closely matched for this group than for the 5-trial group.

**Method**

**Subjects**

Right-handed (20 female and 19 male) undergraduate students from the University of California, Los Angeles, participated for extra class credit. All were naive to the purposes of the experiment, and none had received previous exposure to the task.

**Apparatus and task**

The apparatus was the same coincident-timing task as used in experiment 1. Task and scoring procedures were identical as well.

**Procedures**

Subjects were assigned to either a 1-trial summary group (10 females,
10 males) or a 5-trial summary group (10 females, 9 males), where summary KR was given in the same form as in experiment 1. Acquisition consisted of 90 trials under the respective feedback conditions. The interval from the end of the movement to KR (KR delay) was constant at 8 s for both groups. The interval from KR to the warning tone (post-KR delay) was 4 s for subjects in the 1-trial group, who viewed summary graphs with only a single trial. This interval was lengthened to 7 s for the 5-trial group to allow more viewing time for the longer summaries. This confounds post-KR delay length with summary length. However, this confound was not considered serious, as Salmoni et al. (1984), in their review of the KR literature, found post-KR delay not to be a learning variable (at least in the range from 4 s to 7 s). All subjects participated in a two-day delayed retention test consisting of 30 trials, KR being displayed after every trial in graphic form, exactly as it had been in the acquisition phase for the 1-trial condition.

Results

For this experiment, as for experiment 1, we performed analyses on the overall score for the coincident-timing task, as well as on the score components (target velocity and absolute spatial error). Even though the nature of the retention test here differed from that of experiment 1 (experiment 1 had a no-KR retention test) we again drew conclusions about effects of summary KR on learning from relative performance levels during retention.

Acquisition phase

Fig. 4 shows the outcome scores in acquisition, where data were organized into six blocks of 15 trials. The data appeared very similar to those obtained for the 1-trial and 5-trial groups in experiment 1. Scores increased significantly across blocks, \( F(5,185) = 9.82, p < 0.05 \). There was no significant Groups effect, \( F(1,37) < 1 \), and no significant

\[ \text{\footnotemark}[2] \] The 5-trial summary condition produced higher velocities than the 1-trial condition in both acquisition and retention phases, though these differences were not statistically significant. The 5-trial condition exhibited significantly greater absolute spatial errors in acquisition, \( p < 0.05 \). During retention, the 1-trial summary group did not appreciably change its level of absolute spatial error, in contrast to the 5-trial group, which reduced its absolute spatial errors across retention blocks. This Groups \( \times \) Blocks interaction paralleled the interaction effect observed on overall outcome scores and was significant, \( p < 0.05 \). Thus it appeared that the increase in performance for subjects in the 5-trial condition during the retention test was due to their progressive lowering of spatial error values, with no corresponding change in the 1-trial group.
Groups × Blocks interaction, $F(5,185) < 1$. The 1-trial group immediately increased its performance more rapidly than the 5-trial group, as evidenced by higher scores on block 1, and seemed to have a slight advantage over the 5-trial group on most blocks. Although the Groups effect failed to reach statistical significance, the mean score, averaged across blocks, of the 1-trial group was 6% larger than that of the 5-trial group. Thus, the ordering of the groups relative to summary lengths in acquisition, though not pronounced, followed that of experiment 1.

Retention phase

The overall outcome scores, combined into 10-trial blocks, for the two-day retention test are shown in the right portion of fig. 4. The effect of Blocks was significant, $F(2,74) = 4.81$, $p < 0.05$. Though there was no significant difference between groups, $F(1,37) < 1$, the Groups × Blocks interaction did reach significance, $F(2,74) = 4.59$, $p < 0.05$. Scores for the 1-trial summary group were slightly larger on the first retention block than those of the 5-trial group. The 1-trial group appeared to maintain its performance level across blocks, with no substantial change in scores. However, the 5-trial group, which initially had slightly lower scores than the 1-trial group, exhibited a slight increase in scores on the second block, and a marked increase on the third retention test block. By the end of the retention session, it had surpassed the 1-trial group in overall outcome score. Thus the 1-trial summary group did not improve during retention, while the 5-trial
group improved its scores from 196 to 267 score units across blocks, which was a 36% increase. An analysis of simple main effects (one-way ANOVA) revealed that the two groups did not differ statistically on the first two blocks, $F(1,37) < 1$. On the third block, group differences were statistically significant, $F(1,37) = 4.86$, $p < 0.05$, with subjects in the five-trial group exhibiting 24% higher scores than those in the one-trial group.

Discussion

For the acquisition phase, the results of experiment 2 were essentially similar to those of experiment 1, with the 1-trial group having nonsignificantly higher scores than the 5-trial group. During the KR retention phase, however, subjects from the 5-trial group increased their performance across blocks while subjects from the 1-trial group maintained their scores. This effect was also reflected in the absolute spatial errors, which decreased in the 5-trial group while remaining essentially constant for the 1-trial group (see footnote 2). Subjects in the 5-trial condition seemed to derive their enhanced performance mainly from a decrease in mean AEs across blocks, while maintaining constant velocity.

The specificity hypothesis predicts that, in order to ensure the most effective performance on a given retention test, acquisition conditions should match as much as possible with those in retention. Thus, the 1-trial group should have higher retention test scores relative to the 5-trial group. Two lines of evidence contradict this view. First of all, at the beginning of the retention test there was no significant difference between groups, even though, according to the specificity hypothesis, the retention conditions should have favored the 1-trial group, whose feedback conditions during the retention test were identical to those in acquisition. Second, in the last retention block the 5-trial group exhibited significantly higher performance scores relative to the 1-trial group, a finding which is also difficult for the specificity hypothesis to explain.

One interpretation of these findings is that, relative to the 5-trial group, subjects in the 1-trial group experienced a temporary advantage in acquisition due to receiving KR after every trial. Immediate KR may have served to guide them to the proper response, holding errors at a minimum, and thus increasing their scores. However, subjects in the
5-trial summary group, who did not have the benefit of trial-to-trial feedback, may have engaged in some additional (or different) information processing during acquisition which proved beneficial in other ways, perhaps by enabling a more efficient error detection and correction process, as argued in experiment 1. Thus, due to their already increased sensitivity to the relationship between kinesthetic feedback and movement outcome, subjects in the 5-trial group were then better able to improve once they were presented with feedback in retention.

A second possibility is that 5-trial summaries resulted in greater learning, but that subjects in the 1-trial group, due to receiving immediate KR, temporarily performed more effectively during acquisition relative to those in the 5-trial group. In addition, the 5-trial group's greater learning was obscured at the beginning of retention, because this group was switched to a novel KR condition (KR immediately following every trial), whereas the 1-trial condition performed under identical conditions in acquisition and retention. Then, with continued practice during retention, subjects in the 5-trial condition became accustomed to the new feedback conditions, and improved performance, more closely approximating their actual capabilities.

Whatever the explanation, the findings from experiment 2 suggest that earlier results (Lavery 1962; Schmidt et al. 1989), as well as the results of experiment 1, which demonstrate the superiority of a particular summary manipulation over immediate KR, cannot be explained only in terms of a specificity of learning effect. Instead, summary KR may possibly force subjects to engage in a type of information processing during practice which produces better performance on retention tests, regardless of whether or not outcome feedback is received during this time.

**General discussion**

Schmidt et al. (1989) did not find an optimum summary length as predicted by the guidance hypothesis, whereas the present data indicate an optimal summary length at or near 5 trials, suggesting that the type of task interacts with summary length for optimizing learning. The coincident-timing task used in this study and the ballistic timing task employed by Schmidt et al. (1989) differ in several respects, though, such as movement goal, dependent measures, and number of kinematic
dimensions. Of course, any formal contrast of these tasks here is not possible, as the comparison is done across experiments which allows many alternative interpretations. Nevertheless, it is tempting to suggest that the optimum summary-KR length for learning was found to be relatively small in experiment 1 because the coincident-timing task, here, relative to the ballistic-timing task (in Schmidt et al. 1989), was associated with an increased number of kinematic dimensions. The coincidence-timing task involved numerous potential patterns, and the subject had to discover the trade-offs inherent among them in order to come to an effective solution for the motor problem. Accordingly, as compared to the 'simpler' tasks, systematically more guidance from KR may have been needed in order to arrive at an appropriate movement pattern for the skill. Particularly in early practice, errors associated with ineffective movement styles would have been signalled clearly, allowing the subject to eliminate them and to substitute more effective styles.

The results for experiments 1 and 2 support the general learning-performance distinction (Salmoni et al. 1984) for summary KR, in that the performance levels seen in practice with KR are not necessarily indicative of the acquired performance capabilities of the subjects. Temporary phenomena associated with the presence of KR enhance performance while KR is present, but performance may be altered markedly when these temporary factors are removed (in delayed retention tests), shown by some reordering of the treatment conditions there relative to those in acquisition. Apparently, these data show that the most effective performance conditions for learning – at least as it is measured on 2-day delayed retention tests – are not necessarily those which produce the most effective performance during practice.

Evidence against a specificity of learning hypothesis

The specificity viewpoint (Henry 1968; Tulving and Thomson 1973) predicts that increased summary length in acquisition is positively related to more efficient performance on a no-KR retention test because of the maximum similarity of acquisition conditions to the retention conditions. Previous data (Schmidt et al. 1989) may have appeared to lend support to this view, but our current findings violate the predictions of the specificity hypothesis. In experiment 1, the specificity view would have predicted optimal retention performance
for the 15-trial summary group, because the strings of no-KR trials in acquisition were longer, thus resembling the no-KR retention test more completely than the 5-trial summary did. However, retention results demonstrated an inverted-U relationship between summary length and performance, with the 5-trial condition being approximately optimal. Furthermore, the results of experiment 2 do not support a specificity view, which would have predicted more effective performance on a KR-retention test for the 1-trial summary group. Instead, we observed essentially no group differences at the start of retention, plus an improvement in performance in the course of retention for the 5-trial subjects. By the end of retention, the latter were producing scores 24% higher than the 1-trial group, opposite to the predictions. Finally, the specificity of learning hypothesis predicts optimal retention performance when acquisition and transfer conditions are completely congruent, and it therefore predicts that a 0-trial summary group (i.e. no KR at all in acquisition) will do best on a no-KR transfer. This prediction has been contradicted several times (e.g., Bilodeau et al. 1959; Trowbridge and Cason 1932). There may be some sense in which the similarity, per se, of the practice and retention conditions aids retention performance, but the data from experiments 1 and 2, as well as recent findings obtained by Weinstein and Schmidt (1988) and Young (1988) suggest that this explanation is not satisfactory, and that other alternatives need to be considered.

Support for the guidance hypothesis

The guidance hypothesis (Salmoni et al. 1984; Schmidt et al. 1989) postulates that KR has a powerful informational function, especially in early acquisition, where it guides the learner to the correct response. However, the learner may become dependent upon the information KR provides and may come to rely on it to the exclusion of engaging in other information-processing activities conducive to successful retention performance. This two-factor view, which postulates beneficial guidance-like processes operating together with detrimental dependency-generating processes, is supported by the inverted-U effects on retention in experiment 1. These factors are presumably adjusted to allow the 5-trial condition to be nearly optimal for retention. We also find some support for the guidance hypothesis from the experiment 2 retention data. Presumably, the 5-trial summary group received less
guidance in acquisition relative to the 1-trial group, which resulted in its learning the task at least as well as the 1-trial group, as suggested by essentially equal scores on the first retention block. Alternatively, the lack of excessive guidance during the practice phase may have enabled subjects in the 5-trial summary KR group to make better use of KR to increase their performance once they began obtaining feedback in the retention phase, as seen in their progressive increase in scores. By contrast, the performance of subjects in the 1-trial summary group did not increase, even though they were performing under the same KR conditions as in acquisition.

Factors operating to allow summary KR to aid retention have not been definitively identified. One possibility is that presenting KR after each trial may encourage subjects to constantly change their movement pattern (Bilodeau 1966), with these continual adjustments being counterproductive. Perhaps summary KR — as well as other KR manipulations in which feedback is not accessible after every trial — fosters consistent behavior, discouraging subjects from making ‘maladaptive short-term corrections’ which would prevent them from acquiring a more successful and stable response. There is evidence that subjects are more consistent under a ‘bandwidth’ KR condition, in which responses falling within a designated band of correctness are not followed by KR (Sherwood 1988). This is somewhat similar to summary KR, in that both manipulations involve strings of trials with no overt KR. Perhaps, also, increased consistency in responding enables better use of the information contained in the KR-summary (e.g., mean score and trend information as well as variability information).

Another idea is that no-KR trials as given in summary KR force subjects to attend to intrinsic feedback information in lieu of KR, and to develop stronger error-detection abilities (Schmidt and White 1972; Schmidt et al. 1989). Our data from the delayed recognition test in experiment 1 offer further support for this view. Subjects who experience trials without KR may be more apt to engage in subjective estimation of their performance than do those to whom performance outcome is easily available via KR. This subjective estimation may result in the development of more sophisticated recognition processes, and ensuing benefits for retention.

One important implication of the guidance hypothesis is that the most effective summary-KR conditions for no-KR retention will also be the most effective for essentially any retention test — even one
involving KR on each trial. Therefore, an important general experimental question for the future will involve the possible interaction between acquisition and test (retention) conditions, much as Bransford et al. (1979) have done in the verbal domain, and as Lee (1988) has suggested for the motor domain. A lack of crossover interaction between practice and test conditions would argue against a specificity view, and provide evidence toward some version of a guidance notion of KR. And, if there continues to be support for a guidance hypothesis, searches for the differential kinds of acquisition processes (e.g., the recognition analyses done here) will be an important research goal to come to an understanding of how various forms of KR function to facilitate learning in skills.

References


