The effects of concurrent verbalization on performance in a dynamic systems task

P. McGeorge* and A. M. Burton

Department of Psychology, University of Nottingham, Nottingham NG7 2RD, UK

Two experiments are described based on a dynamic systems control task previously shown to favour implicit learning. Subjects asked to verbalize their reasoning whilst performing the task showed significantly improved task performance. This result is in line with suggestions that concurrent verbalization may facilitate task performance. Presentation of graphical information representing the behaviour of the system also aids learning. Both concurrent verbalization and graphical representations probably have the effect of increasing the efficiency with which the available task information is assimilated. When both these aids are available to subjects, the effects of one may be masked by the effects of the other. This would account for the results of previous studies in which no effect of concurrent verbalization has been found. After relatively few exchanges with the system some subjects appear to be developing verbal task knowledge; this may represent the initial stages in the development of a mental model of the system.

In recent years a great deal of interest has been generated by questions of conscious (explicit) and unconscious (implicit) learning. Implicit learning has been reported in a great variety of experimental tasks, for example, artificial grammar learning (Reber, 1967; Reber, Kassin, Lewis & Cantor, 1980), dynamic control tasks (Berry & Broadbent, 1984, 1988; Broadbent, FitzGerald & Broadbent, 1986; Hayes & Broadbent, 1988; Stanley, Mathews, Buss & Kotler-Cope, in press), complex procedural tasks and stimulus covariations (Lewicki, 1986; Lewicki, Hill & Bizot, 1988; Lewicki, Hoffman & Czyzewski, 1987).

This work has important implications for theories of skill acquisition, offering an alternative to the previously widely held view that skill acquisition involves the transfer of control from declarative knowledge to proceduralized knowledge (Anderson, 1982, 1983; Shiffrin & Schneider, 1977). The results of experiments on implicit learning have suggested that under certain conditions the process may function in the reverse direction, that is, from procedural knowledge to declarative knowledge (Broadbent, 1987; Stanley et al., in press).

It has been suggested that the nature of a task will initially favour one of two modes of learning (Berry & Broadbent, 1988; Hayes & Broadbent, 1988). Tasks in which there is only a limited number of variables and in which the important relationships are particularly salient will favour a selective mode of learning. During selective mode learning hypotheses are explicitly generated and tested and the model of the system updated accordingly. The number of variables and associated

* Requests for reprints.
contingencies that can be considered is limited by the size of abstract working memory (Hayes & Broadbent, 1988).

In contrast, tasks having a large number of variables, and in which the important relationships are not particularly salient, favour a different mode of learning. Unselective mode learning is characterized as instance-based, leading to the formation of a number of condition-action pairs (Berry & Broadbent, 1988). An instance-based approach to learning entails the storage of both correct and incorrect exemplars. It has been suggested that this would make interpretation of such a store by articulatory processes difficult, if not impossible (Berry & Broadbent, 1988).

The traditional or ‘common-sense’ view (Broadbent et al., 1986) of human knowledge suggests that all mental processes have access to a shared store of information. Under such conditions this information should be available to articulatory processes (given the correct probes). Work on implicit learning suggests that there may be multiple stores, not all of which are open to, or suitable for, articulatory processes (Hayes & Broadbent, 1988). Stanley et al. (in press) have suggested that there may be two forms of information store—an instance-based store, such as that proposed by Berry & Broadbent (1988), and a mental model of the form proposed by Johnson-Laird (1983). They suggest that only when a mental model of the task has been formed are subjects able to demonstrate verbal knowledge equivalent to their task performance.

Several of the studies on implicit learning have attempted to obtain a measure of subjects’ verbal knowledge through the elicitation of concurrent protocols (Berry & Broadbent, 1984; Stanley et al., in press). It has previously been suggested that concurrent verbalization may affect the cognitive processes involved in task performance (Ericsson & Simon, 1980). However, the results of several experiments, involving dynamic systems control tasks, have suggested that concurrent verbalization may only have a significant effect when accompanied by pre-task verbal training (Berry & Broadbent, 1984), or after subjects have received sufficient exposure to the task to develop a mental model of the system (Stanley et al., in press). In both cases, following the suggestion by Berry & Broadbent, the effect may be accounted for through verbalization serving to maintain attention on salient features of the system. Such an explanation appears to fit well with the suggestion by Ericsson & Simon (1980) that concurrent verbalization leads to more deliberate planning. Such deliberate planning would require the important relationships to be carefully considered.

The effects of concurrent verbalization are not limited to planning. Concurrent verbalization is known to facilitate the interpretation of complex visual stimuli (Sokolov, 1972). The nature of the surface structure of the task appears to play an important role in deciding what effect, if any, concurrent verbalization will have.

In the study by Berry & Broadbent (1984) two versions of a dynamic systems task were used; both had the same underlying structure but each had a different surface structure. In one of these tasks subjects were asked to control the output from a simulated sugar factory, whilst in the other they were asked to control a simulated personal interaction. Though the tasks were constructed to be formally equivalent, in the factory task subjects were shown a graph representing the behaviour of the
system. In neither case was there an effect of concurrent verbalization without pre-
task verbal training. However, as mentioned above, the nature of the task plays an
important role in deciding what the effect of concurrent verbalization will be. It
seems likely that it is the surface structure of the task that is important in this respect.
Consequently, the two versions of the dynamic systems task used by Berry &
Broadbent (1984) need to be treated separately. This is supported by the work of
Stanley et al. (in press) who have demonstrated that the sugar production task is
significantly more difficult than the personal interaction task. They suggest that the
reason for this lies in context effects associated with the surface structure of the tasks.
The form of mental model activated by the personal interaction task more readily
assimilates task data than does that activated by the sugar production task. Under
these conditions the facilitative effects of a more congruent mental model may serve
to reduce any observable effect of concurrent verbalization. Therefore, in the sugar
production task it is the availability of a graph representing the systems behaviour
which may reduce the observed effects of concurrent verbalization, whilst in the
personal interaction task the same effect is achieved through the contextual effects of
the task.

In the study in which subjects were required to control the output of an imaginary
sugar factory (Berry & Broadbent, 1984), subjects had access to a numeric
representation of the last exchange (one exchange being a subject input followed by
a system output) and a graph representing the behaviour of the system for all
previous trials. Subjects in this study were therefore presented with a visual display
consisting of numeric values and a graph representing these values. The use of graphs
to aid the interpretation of numeric data is a widely accepted principle and it may be
that presenting subjects with such a representation facilitates their task performance.
As has been reported, one effect of concurrent verbalization may be to facilitate the
interpretation of complex visual stimuli (Sokolov, 1972). However, if the available
information is already being presented in a manner which allows maximum benefit
to be derived from it, then we might expect that concurrent verbalization would fail
to have any effect. Thus an alternative explanation of why no effect of concurrent
verbalization is found prior to the development of a mental model (without pre-task
verbal training) could be that the use of a graphic representation of the behaviour of
the system ensured that the available information was already presented in a
maximally efficient manner.

The experiments described below set out to investigate whether the availability of
a graph representing the behaviour of the system would alter the observed effects of
concurrent verbalization on performance of a dynamic systems task.

The task

The task used in the following experiments was taken from Berry & Broadbent
(1984). In this task subjects are told they have been placed in charge of a sugar
production factory and are required to reach and maintain target values for sugar
production through alterations in the size of the workforce. The relationships
between the variables are specified by the following equation:
\[ P = [(W \times 2) - P'] + R, \]

where

\[ \begin{align*}
P &= \text{new output} \\
P' &= \text{previous output} \\
W &= \text{size of workforce} \\
R &= \text{random value \((1, 0, -1)\)}
\end{align*} \]

Each new output production figure is thus based not just on the size of the workforce selected by the subject but also on the size of the previous output. This fact ensures that there is not a one-to-one correspondence between workforce and output production. The addition of a random value in the range 1 to \(-1\) further complicates the task by ensuring that the system tends to wander from any value on which it settles. The use of this random element means that a scoring criterion must be adopted in which output production falling within one step either side of the target is counted as a success.

For each set of exchanges subjects were given a set of initial conditions, consisting of a target production level and values for the current workforce and production. They were then asked to select values for the workforce (a value ranging from 100 to 1200 in integral hundreds), after which the system responded with the new production level. Each workforce selection and subsequent production output counted as one exchange. The system was set with a minimum output of 1000 tons and a maximum of 12000 tons. The aim of the task was to reach and maintain the target production levels.

**Experiment 1**

As discussed above, subjects in the Berry & Broadbent (1984) study had access to a graph representing the system’s behaviour. It is suggested that this graph could explain why no effect of concurrent verbalization was detected without pre-task verbal training. To investigate this possibility subjects in the experiment described below only had access to a numeric representation of their last 10 exchanges with the system. Under these conditions it was predicted that concurrent verbalization would affect task performance.

In the original Berry & Broadbent study (1984) each subject received two sets of exchanges, resulting in a total exposure of 40–60 exchanges. In all these studies fixed initial conditions were used (target output: 9000; start: output 6000; workforce 600). The results of this study demonstrated that, whilst subjects improved in their ability to control the system, there was no equivalent improvement in their verbal knowledge. The characterization of unselective mode learning proposed by Berry & Broadbent (1988) suggests that this effect should be found if the initial conditions are not fixed but free to vary between sets of exchanges. This experiment set out to replicate the previous work of Berry & Broadbent (1984) using randomly selected initial conditions for each set of exchanges. These conditions consisted of a target level randomly selected from the range 4000–9000 tons and an initial production level selected from the same range. An added restriction was placed on initial production to ensure that it was not equivalent to the target value. The associated initial workforce was simply the initial production level divided by 10.
Method

Subjects. Thirty-five undergraduates acted as subjects during the experiment and all were new to the computer task.

Design. A mixed design was used, each subject being allocated either to a control group \((n = 15)\) or to the experimental group \((n = 20)\). The initial conditions were randomly chosen for each set of exchanges and each subject received three sets of 30 exchanges. During the third set, subjects in the experimental group were asked to verbalize and describe any heuristics they were following and the reasoning behind each of their selected workforce sizes. The dependent variable was the number of times the output fell within \(\pm 1000\) of the target.

Procedure. Subjects were tested individually, each being given written instructions concerning the nature of the task. These instructions included the permissible input values and the possible range of outputs (including the minimum and maximum values). Subjects were not aware of the scoring criterion and were told only that they were expected to achieve and maintain the target values. Once these instructions had been given, subjects interacted with the system for three sets of 30 exchanges, one exchange corresponding to a workforce input and a subsequent output. The duration of each set of exchanges was controlled by the subject. Whilst the response of the system to a workforce selection was immediate, no time constraints were placed on subjects for making their selections. Subjects were thus allowed ample time to consider how to proceed. Following the second set of exchanges, subjects in the experimental group were told that they would have to interact with the system for one further set but that this time they should attempt to describe what they were doing and any rules or heuristics they might be employing. These protocols were taken concurrently with the task. The control group were simply told that they would have to interact with the system for one further set of exchanges. After the third set of exchanges the number of criterion outputs \((\pm 1000\) target\) were calculated for each set.

Results

The mean number of output productions falling within the criterion of \(\pm 1000\) tons of the target, for each of the three sets of exchanges, are represented by the graph in Fig. 1.

Non-parametric statistics were used throughout the analysis because of non-homogeneity of the group variances. Analysis of the data for set 1 and set 2 in the experimental group (Wilcoxon test) showed no significant difference. Analysis of sets 2 and 3 showed a significant difference \((T = 25, n = 20, P < 0.05)\). Analysis of the data in the control group showed no significant performance improvement across sets. Comparison between the control and experimental groups (Mann–Whitney U test) revealed no significant difference except between the final sets in each group \((U = 81, n = 15, 20, P < 0.05)\).

The protocols from subjects contained too little information for any formal analysis. They consisted mainly of descriptions of the current state of the system, along with statements suggesting that an input was a test of the behaviour of the system with no other justification offered. This is consistent with experiences reported by Berry & Broadbent (1984). Statements about the apparent randomness of the system were also quite common.

Discussion

The results of this experiment indicate that no significant performance improvement occurred across the three learning sets in the control group. This contrasts with the findings of Berry & Broadbent (1984), where significant improvements in
performance were obtained over two sets of exchanges. However, from these results alone we cannot conclude that no learning occurs, either selective or unselective. The combination of random initial conditions along with the absence of a graphic representation of the system makes the task more complex. It has been suggested (Hayes & Broadbent, 1988) that unselective mode learning proceeds by logging the number of successful actions and employing the one most frequently successful. It is possible that while using random conditions does not decrease the number of successful actions it may lead to their being categorized in a different manner. For example, consider two successful exchanges. If in the fixed initial condition set an input of 900 achieves an output of 9000 in both exchanges this would reinforce the 900→9000 action. However, consider two learning sets in the random initial condition, one with a target of 9000 and one with a target of 6000. At one level an input of 900→9000 and an input of 600→6000 could reinforce the action: input the same number of workers as the required production. These inputs could also be interpreted as two separate actions. If we imagine implementing such a system in an instance-based architecture, then the random initial condition would take far more examples in order to appear to be following the underlying rule. So whilst we can say we have detected no learning across the three sets, this may be due to insufficient exchanges between the subjects and the system.

The analysis indicates that there is a significant difference in the means obtained in the third set of the experimental group and those obtained in both set 1 and set 2 of that group and the final set in the control group. It appears that the concurrent verbalization required in the third set significantly affected subjects' performance.
This demonstrates that concurrent verbalization does have an effect on the performance of a dynamic systems task when a graphical representation of the system is absent.

In the original study by Berry & Broadbent (1984) the surface structure of the two versions of the dynamic systems tasks (sugar production, personal interaction) differed both in the visual display with which subjects were presented and the context of the task. The experimental group in the present study received a visual display that was similar in format to that received by the personal interaction group in Berry & Broadbent's study. The main difference between the two tasks was thus contextual. The fact that a significant effect of concurrent verbalization was found in the experimental group of this study, and not in the personal interaction group in the study by Berry & Broadbent (1984), appears to add support to the suggestion by Stanley et al. (in press) that the contextual effects created by the surface structures of the tasks play a significant role in subjects' performance.

The subject protocols mainly concerned the values being displayed and statements as to the apparent randomness of the system. The main reasons given for workforce selections consisted of statements that a value was a test of the behaviour of the system or a complete guess. This lack of substantial verbal knowledge is in line with the findings reported by Berry & Broadbent (1984) and is suggestive of a dissociation between task performance and verbal knowledge. It also serves to indicate that the effect of concurrent verbalization cannot be attributed to the existence of a mental model of the system.

**Experiment 2**

The previous experiment investigated the effects of concurrent verbalization on task performance when subjects did not have access to a graph representing the behaviour of the system. The subsequent experiment set out to investigate the effect that the presence of a graph representing the behaviour of the system would have on task performance in the absence of concurrent verbalization. It was predicted that when such a representation was available learning would be facilitated and this would be indicated by improvements in task performance. In the absence of concurrent protocols an attempt was made to elicit any verbal information that subjects might have using an open question administered after the three sets of exchanges.

In the previous experiment no learning was detected in the control group. The reason suggested for this was that the use of randomly selected initial conditions made the task more complex, a situation further exaggerated by the absence of a graph to represent the system's responses. If, as has been suggested, the presence of a graph plays such an important role in learning then the use of a graph should facilitate learning when random initial conditions are used.

In addition to these performance issues, Expt 2 was designed to examine further the issue of verbal knowledge. The absence of task-specific information in subjects' protocols (Expt 1) may be related to subjects' expectations concerning the nature of the experiment. It is possible that when asked to give the reasoning behind their inputs subjects were aiming at a more detailed level of explanation than was
necessarily required or available to them. To put this another way, what subjects counted as being a reasonable explanation of their behaviour did not include the limited information they possessed concerning how they selected workforce values. A possible way of avoiding this problem is to remove the requirement to explain the reasoning to the experimenter and replace it with a request to provide any information that might help the next subject coming to the task.

A further advantage of this method of assessment is that the resultant information can be readily translated into computational procedures. These procedures can then be used to evaluate the correspondence between verbal knowledge and observed performance. This can be done by comparing the results obtained from simulations using these procedures to the subjects’ observed scores. This is the method used to evaluate levels of verbal knowledge in the present experiment.

In order to err on the side of caution, the following assumptions were adopted in the analysis of procedures:

1. Only those procedures which produce scores equal to or greater than the observed score were taken as representing verbal knowledge equivalent to performance knowledge.
2. Workforce values were chosen at random whenever the reported heuristic was imprecise about the action to be taken in a given situation.

Method

Subjects. Twenty-eight volunteers acted as subjects during the experiment; ages ranged from 16 to 19. Each subject received a small payment for their participation and all were new to the computer task.

Design and procedure. A mixed design was used with subjects randomly allocated to one of two groups: group A, random initial conditions; group B, fixed initial conditions.

The task used was the same as that described previously, except that as well as numeric representation of the previous 10 exchanges subjects were also shown a graph representing the behaviour of the system. The procedure was the same as in Expt 1. All subjects interacted with the system for three sets of targets with 30 exchanges per set (one exchange being a workforce value followed by an output production). No protocols were elicited during the performance of the task. Immediately after the third set of exchanges subjects completed an open question in which they were asked to offer any advice they could to the next potential manager.

Results

The mean number of output productions falling within the criterion of \( \pm 1000 \) tons of the target for each set of exchanges are represented by the graph in Fig. 2. Analysis of variance revealed a significant effect of learning set \((F = 6.34, \text{ d.f.} = 2, 52, P < 0.05)\) and a non-significant effect of initial conditions (fixed or random). There was no interaction.

As described above, analysis of the open questions was carried out using computer simulations of the reported heuristics. Comparisons were made between simulated performance and observed performance. These indicated that five out of 14 subjects in the fixed initial state group and six out of 14 subjects in the random initial state group reported heuristics that resulted in simulated performances that were either equivalent to, or better than, observed performance.
Figure 2. Group means across sets of exchanges by initial condition. [square] [square], random; [circle] [circle], fixed.

Discussion

These results show that a significant performance improvement occurred over the three learning sets for both groups and this appears to replicate and extend the findings of Berry & Broadbent (1984). The significant performance improvement in the group receiving random initial conditions is counter to the results observed in Expt 1. This would appear to suggest that the presence of a graph to represent the behaviour of the system does significantly affect task performance.

Unlike in Expt 1, where protocols were elicited, in this experiment subjects' verbal knowledge was investigated using an open question administered after the completion of the three sets of exchanges. Examination of the responses revealed that, at least in some instances, subjects did report heuristics that they felt might be of use to the next potential manager of the factory. For example:

If production levels fall, take on more workforce. Once you have the level keep around the same number of workforce (fixed initial conditions).

Use around 1000 and 1100 people in his workforce (fixed initial conditions).

Begin with experimenting around 600-strong workforce and study relationship of graph lines. Once a workforce has been found that produces little change on the graph, then alter the workforce 100 either side to remain near or on the target (random initial conditions).

Don’t vary the workforce by too much each time. Try figures close to the target level (random initial conditions).

The use of computer simulation allows us to investigate the effects of different strategies on task performance. These studies can be used not just to implement
strategies obtained from experimental reports but also as an aid in classifying such reports. The numeric nature of the sugar control task makes it an ideal task on which to conduct such simulations.

Several of the strategies suggested in the open questions, such as those outlined above, when implemented resulted in simulated performance scores that were either equivalent to or better than the observed performance levels for that subject. It does appear that at least in some cases subjects do have verbal access to the knowledge around which they are basing their decisions.

General discussion

The results of these experiments raise several interesting points, chiefly concerning the effects of concurrent verbalization on task performance. In Expt 1 a strong effect of verbalization was detected, subjects in the experimental group showing a marked improvement on their previous performance and on the performance of the control group. This result is apparently contrary to that expected on the basis of previous work by Berry & Broadbent (1984) or Stanley et al. (in press). Subjects had had neither pre-task verbal training nor sufficient exchanges, judging from the poor quality of the obtained protocols, to develop a mental model of the system. The results of Expt 2 indicated that given a graphic representation of the behaviour of the system, subjects in the random initial condition group did indeed show a significant performance improvement within 90 exchanges. This suggests that a graphic representation plays a very important role in subjects' interpretation of task data. It seems likely that the facilitative effects of a graph would be sufficient to mask any effect of concurrent verbalization. The absence of any effect of concurrent verbalization found by Berry & Broadbent (1984) cannot therefore be taken as evidence against the proposed effects advanced by Ericsson & Simon (1980). Indeed, the effects Berry & Broadbent observed both with and without pre-task verbal training appear in keeping with those proposed by Ericsson & Simon.

The analysis of the responses to the open question obtained in Expt 2 also raises some interesting points. It does appear that at least in some instances subjects have access to a form of verbal knowledge that can account for their task behaviour. Stanley et al. (in press) have suggested that verbal ability only attains levels comparable with actual performance after subjects have developed a mental model of the system. Prior to the development of this model only an instance-based store is available to which verbal processes either do not have any access, or cannot interpret due to the large amount of irrelevant information such a store would contain (Berry & Broadbent, 1988). The availability of verbal knowledge to some subjects may represent the initial stages in development of a mental model. This mental model appears to be developing after only 90 exchanges between the subject and the system.

The development of a mental model, and the apparent narrowing in the dissociation between task performance and associated verbal ability, do not necessarily reduce the importance of unselective learning in the acquisition of expertise. Even after the development of a mental model there may still be a dissociation between task performance and verbal performance. This may not be in observable levels of performance but in the knowledge stores on which performance is based.
In an implicitly learned task verbal performance will be based on the mental model, while task performance may be based on either the currently active mental model (in the case of rare instances) or on the instance-based store (in the case of common instances). It follows that only in the case of rare instances need both verbal and task performance be based on the same knowledge store. The dissociation that may occur in the case of common instances will not be important provided that the mental model is an accurate and complete representation of the knowledge in the instance-based store.

References


Received 18 August 1988; revised version received 20 February 1989