

Errors in Training Computer Skills: On the Positive Function of Errors

Michael Frese, Felix Brodbeck,
Torsten Heinbokel, Christina Mooser,
Erik Schleiffenbaum, and Petra Thiemann
University of Munich

ABSTRACT

Traditionally, errors are avoided in training. In contrast to this approach, it is argued that errors can also have a positive function and that one has to learn to deal efficiently with errors on a strategic and an emotional level (error management). An experiment tested these assumptions. One group ($n = 9$) received guidance for error-free performance; another group ($n = 15$) received error training. In the latter group, errors were produced by assigning problems that were too difficult to deal with. The error-training group showed higher scores in the nonspeed performance tests. Error training seems to be positive for people with high scores on the cognitive failure questionnaire (Broadbent, Cooper, FitzGerald, & Parkes, 1982).

Authors' present addresses: Michael Frese, Department of Psychology, University of Giessen, Otto-Behaghel-Strasse 10, 6300 Giessen, Federal Republic of Germany; Felix Brodbeck, Torsten Heinbokel, Christina Mooser, Erik Schleiffenbaum, and Petra Thiemann, Department of Psychology, University of Munich, Leopoldstr. 13, D-8000 Munich 40, Federal Republic of Germany.

CONTENTS

1. THE FUNCTION OF ERRORS IN TRAINING
 2. METHODS
 - 2.1. Subjects
 - 2.2. Design of the Experiment
 - 2.3. Instruments
 3. RESULTS AND DISCUSSION
 - 3.1. Mean Differences Between Experimental Groups
 - 3.2. Intervening Variable of Cognitive Failures
 - 3.3. Mood Changes Across the Experiment
 4. OVERALL DISCUSSION AND CONCLUSIONS
-

1. THE FUNCTION OF ERRORS IN TRAINING

Errors are a persistent problem when working on a computer. Novices spend most of their time with error recovery. Even experts spend a lot of their time dealing with errors (Prümper, Zapf, Brodbeck, & Frese, in press). Errors lead to human costs because they are stressful and frustrating. Because errors are frustrating, people try to avoid them. Avoiding errors may sometimes lead to negative consequences, for example, when a person stops to explore the system.

Trainers have different strategies to deal with errors as well. Most trainers want to reduce the number of errors by preventing errors from happening (Brodbeck, Prümper, & Zapf, 1989). This is in line with the behavioristic tradition, in which errors were conceptualized as punishment and punishment was to be avoided (Skinner, 1953). Therefore, programmed learning, which grew out of this tradition, made strenuous attempts to eliminate errors. Curiously, humanistic traditions also suggested to reduce errors, because they are thought to be frustrating and anxiety provoking and would, therefore, disrupt good performance.

We suggest an alternative strategy: error management, a concept close to Brown and Newman's (1985) "management of trouble." There are four reasons for the positive role of errors in training.

First, the mental model of a system is enhanced when a person makes an error. An error in training indicates to the person that a certain area of the system may not be known well enough. New insights can develop as the result of having to deal with the error. Furthermore, errors can sometimes spur creative solutions of a problem. When an error occurs, the person may in fact be led to an unfamiliar part of the system and start exploring to find a better

way of doing things. The resulting better mental model can also enhance memory for the system parameters.

The notions developed here are similar to Winograd and Flores's (1986) reinterpretation of Heidegger's (1962) concept of breakdown. *Breakdown* implies that some habitual strategy is interrupted. Errors help to interpret a habit and thus provide a chance to reflect on the system or on one's own knowledge. In terms of action theory on which this article is based, this is interpreted to mean that errors give a chance to reintellectualize strategies that have been automatized (Frese & Sabini, 1985). Errors occur, of course, both in actions regulated by conscious goal setting and planning and in routinized operations (Zapf, Brodbeck, Frese, Peters, & Prümper, 1990). Only the latter ones stand in the foreground of the breakdown concept (Bødker, 1989; Winograd & Flores, 1986). Novices show more errors in higher level, consciously processed actions, and at least some type of experts show more errors in routinized actions (Prümper et al., in press). However, learning can take place from errors occurring in consciously regulated as well as in routinized actions. An interesting research question is: Where does more learning take place? (This question is not addressed in this article, because this study is based on novices who show few routinized actions.)

Second, mental models are better when they also encompass potential pitfalls and error-prone problem areas. Then the potential problem areas can be avoided, extra care (e.g., double-checking) can be taken when these areas are approached, errors are more quickly recognized, and explanations for errors are more readily available. All of this helps to either prevent errors in the work situation or to deal with them once they occur.

Third, when error-free training is aspired for, the trainer will restrict the kind of strategies used by the trainees, because unrestricted strategies increase the chances of errors. There is growing evidence that, at least in adult learning of human-computer skills, trainees should be able to explore the system (Carroll & Mack, 1984; Frese et al., 1988; Greif & Keller, 1990; Wendel & Frese, 1987). However, exploratory learning presupposes the chance to make errors.

Fourth, errors not only appear in training but also in the actual work situation. As a matter of fact, we think that this is the most important reason why error management strategies should be taught. At work, one usually does not have the support that one receives in the training sessions. Therefore, learning how to deal with errors is an important prerequisite of being able to use a system in the work situation.

These positive aspects of errors in training are often overshadowed by the upsetting and frustrating effects of errors. Thus, the negative emotional feelings that accompany errors should be reduced. This can be done by teaching strategies of how to manage errors. Then the user knows what to do

in case of an error and is, therefore, less upset when an error appears. Another strategy is to become more tolerant of one's errors. Committing an error always implies additional frustration about oneself that one has "made such a stupid mistake." This additional emotional upset can be reduced by becoming more nonchalant about errors. This also decreases the tendency to give up in case of an error. Thus, training should encourage an active approach to learning from and dealing with errors rather than becoming emotionally upset because of errors.

We tried to investigate these issues in an experimental study. Because little is known about error training (i.e., a training that explicitly incorporates making errors), our experiment was an initial test of the usefulness. At this point, we did not yet tease apart the emotional side of error tolerance and the skill component of teaching error-management strategies. Rather, we developed a rough test. In order to reduce the negative emotions, certain heuristics were presented (more on this later). To present the subjects with error experience, difficult problems were given at a certain point in the training process.

Our hypothesis was that the group receiving error training would do better in a performance test than a group that was not allowed to make errors. This is true for the four positive functions of errors in the training process just stated. The result should be more pronounced with difficult tasks, because greater knowledge is necessary to solve them and errors (and, therefore, error handling) are more likely to occur. We did not hypothesize any differences in subjective satisfaction, because satisfaction is usually high in any type of training (Frese et al., 1988). Furthermore, we suggested that the negative emotional effects of error training should be reduced when people learn that errors help them understand the system. Conversely, people who are not allowed to make errors in training should show higher negative emotionality when they are confronted with unaided performance tasks.

One potential moderator variable, a general error tendencies (the concept of cognitive failures; cf. Broadbent et al., 1982), was also studied. We tested two opposing hypotheses. First, because people with a high score on the Cognitive Failure Questionnaire (CFQ) are more sensitive to stress (Broadbent et al., 1982), and because error training is more stressful, there could be negative correlations between the CFQ and performance variables in the error-training group. Similarly, one could argue that people with a high CFQ score could profit more from the error-avoidant-training group, because this group's tendency to make errors is reduced. Second, the opposite hypothesis is: People with high CFQ scores make more errors in general; therefore, they should profit more from error-management training, bringing them to the same level as people with lower CFQ scores. Because they would not learn error-management strategies in the error-avoidant-training group,

there should be negative correlations between CFQ scores and performance in this group.

2. METHODS

2.1. Subjects

For reasons of ecological validity, nonstudent subjects were recruited by running an advertisement on the radio for free training. The volunteers were asked a few questions. Computer novices and those who worked with a typewriter in their job were allowed to take the course ($N = 24$; 23 women and 1 man). Most of the subjects worked as secretaries, but there were also some educators and teachers; all of them were German. Typewriting knowledge was rather diverse, and ages ranged from 17 to 50 years (with a median of about 26 years).

2.2. Design of the Experiment

In many ways, the experiment simulated normal training conditions. The subjects received a 6-hr training program on a word processing system (WordStar). In addition, they were tested for 2 more hours after the training. The training took place on a weekend—4 hr on Saturday plus 2 hr of training and 2 hr of testing on Sunday. (There was a slight experimental “mortality” of people who did not show up on Sunday: Two subjects, one from each of the two experimental groups, were omitted.) The tests consisted of 1 hr of performance tests and 1 hr of filling in questionnaires. A few short questionnaires were also completed at the beginning of the training and after the sixth hour.

For practical reasons—access to the computer room at the specified times and our interest to include all qualified subjects—the subjects were trained and tested in three randomly assigned groups. Two of the smaller ones were given error training. To ensure that the two groups were not different from each other, 85 t tests were run on all possible variables; there were only three significant differences ($p < .05$) between these groups. Because 5% significances are to be expected on a random basis, there are no differences between the two subgroups, who received error training, for all practical purposes. Therefore, these two subgroups were treated as *one error-training group* described shortly.

At the beginning, subjects heard a 20-min lecture on the basics of computers, including how to turn the computer on, how to use a disk, how to

put the disk into the disk drive, how to start MS-DOS and WordStar. This introductory material was distributed to all subjects.

Each experimenter supervised two subjects. All of them worked in a room with 10 computers in the department of psychology on tasks that were given to them by the experimenters (most of the task material was taken from Altmann, 1987, and Schulte-Göcking, 1987). Because the theoretical rationale of the study had to be clear to the experimenter (so that they could, e.g., make the right choices when confronted with unforeseen situations), the experimenters were not blind to the treatment.

The subjects were asked to pretend that they had just bought a computer and were now working with it at home with no help available except the material that we gave them. This implied that they were not supposed to ask the experimenters questions, but rather pretend they were not there.

One group (error-avoidant-training group; $n = 9$) received training designed to reduce the chances to make errors. In a way, this training mimics certain computer-driven tutorials (cf. Greif & Janikowski, 1987). For each task (the tasks were the same for all subjects), this group received written instructions that spelled out each step and commands that were to be used for solving the tasks. These instructions ensured that no errors could occur, but they did not give specific explanations for the steps and commands. Whenever the subjects in this group made an error (e.g., by mistyping), the experimenter would quickly intervene to correct it to the status before the error was made. No further explanation was given by the experimenter. The experimenters gave as a rationale that one should not make errors in training but should always learn the steps in the correct way. When subjects finished a certain task before the allotted time, they were asked to do the task again so that they could learn it better. They were discouraged from exploration, although in a few cases they insisted on doing it anyway.

The other group (the error-training group; $n = 15$) did not receive detailed instructions of how to proceed. In order to equalize treatment, they were given leaflets with all the commands needed to solve the tasks (actually an overhead projection of the same material was presented to the error-avoidant-training group for a brief period of time at the beginning of each session). The tasks increased in difficulty and were designed to emphasize certain parts of the word processing program. There were ample chances to make errors in the first 4 hr of the training. When errors were made, the subjects in the error-training group were encouraged to solve the problems themselves. This was supposed to enhance their error-management skills. If they could not find solutions within 3 min, the experimenter intervened and put them into a condition in which they could continue to work on their task. The experimenters did not give any further explanation or advice.

The error-training group was also given a set of four heuristics on paper, which were also presented as overhead during the training. These heuristics

were supposed to counter the emotional and frustrating quality of errors and to increase the problem-solving approach to error correction: "I have made an error. Great!," "There is a way to leave the error situation," "Look at the screen," and "I watch what is on the screen and what is changing."

The heuristics were introduced by stating that we wanted to show them that one could learn from errors. The first two heuristics refer to reducing the emotional impact of errors. The last two heuristics were to help error-management strategies, because we had found in previous studies that novices did not observe the screen very well. The heuristics were frequently recited by the experimenters. When subjects made an error, experimenters would often say something like "that is good, now you can learn from this error."

Although the error-training group could make errors all the time, we introduced a special "forced-error session" in the fifth and sixth hours. Here the subjects received tasks that were extremely difficult and that even WordStar experts often have difficulties with; for example, they were to write the Danish word *Øre*, or they had to create a double-column text and tables. (Of course, the error-avoidant group had to do the same tasks, although the solutions were again presented to them in the step-by-step, correct fashion.) During the 4 hr of training beforehand, the experimenters would help after the waiting period of 3 min; in this forced-error session, the subjects received no help and no additional written material. To reduce frustration, the subjects worked together as couples (actually, in some cases, they refused to do this; the experimenters did not insist on cooperative work but just encouraged them several times). However, in each case, they were asked to use two computers so that no division of labor would be introduced by one person suggesting how to proceed and the other person just typing it into the system.

2.3. Instruments

The following performance variables were ascertained in the testing phase at the end of the experiment.

Free recall: The subjects were asked to recall all the commands that they still knew and explain what they could be used for. Only correct answers were counted.

Competence: The subjects were presented with a flawed text on the screen. The required corrections were presented on a sheet of paper. The subjects had enough time to correct the flaws. Tasks varying from low to high difficulty were given. The experimenter wrote down the keystrokes used. Afterwards, these protocols were blindly rated on how good the tasks were solved on a 5-point scale anchored as follows: *correct and efficient solution* (5), *correct solution but inefficient procedure* (4), *problem not completely solved but direction correct* (3), *major*

flaws in solution (2), and *subject did not really try* (1). Ten subjects were judged by two raters. In only 5.4% of the judgments did a deviation of 1 point (none higher) appear. Thus, the rater agreement was 94.6%. The ratings were collapsed into three scales of low, moderate, and high difficulty (Cronbach's alphas shown in parenthesis): Competence/Easy (.84), Competence/Middle (.79), and Competence/Difficult (.81).

Transfer competence: The subjects had to solve three tasks that were not taught during training. Performance was rated using the procedure just noted. Rater agreement was 96.7%.

Speed test variables: Similar to Roberts and Moran's (1983) benchmark test, it was ascertained how many errors the subjects made when copy-typing a text under speed conditions (called *error count*). Each error was scored; if the subject did not recognize an error, the experimenter pointed it out. Average error correction time (in seconds) was taken with a stop watch to measure how long the subject needed to correct each error. The number of words written was counted as well.

In addition, the following variables were used.

Overall performance rating: The experimenters gave an overall rating on how well the subjects were able to use the commands taught in the course on a 5-point scale. This rating was done twice: once after the first day and again after the second day. However, only the second rating was analyzed here. In contrast to the other variables just described, this rating may have been biased by experimenter knowledge of the hypotheses. However, the experimenters' ratings correlated well with objective performance measures (e.g., $r = .66$ with Competency/Easy, $r = .63$ with Competency/Middle, $r = .75$ with Competency/Difficult, and $r = .52$ with Free Recall; all of these correlations were significant).

Satisfaction: A nine-item scale was presented (Frese et al., 1988) on how satisfied the subjects were with the training ($\alpha = .84$).

CFQ: The item content relates to general "scattiness." The questionnaire was translated from the English version (Broadbent et al., 1982) and was tested for reliability in a pilot study with 74 students. Cronbach's alpha in that study was .87. In our sample, Cronbach's alpha was .82 and the test-retest correlation was .94.

Frustration: This 6-item scale was constructed for this study to measure how much the subjects were frustrated by having to deal with errors during their work on the tasks. Two sample items are: "Work with the computer is now frustrating" and "I would like to stop working with the computer now." Frustration was measured twice: once after the forced-error training (or its equivalent in the error-avoidant-training group) and the second time directly after testing performance at the end of the experiment. Cronbach's alphas were .65 for the first measurement and .75 for the second measurement. The correlation between the two measurement points was .49 ($p < .05$). We

expected a low correlation across time because the wording of the items related to the mood state at a specific point in time. Changes are, therefore, expected.

Present State Report: This is a state measure of mood and strain (known as *Eigenzustandsskala*, which was developed by Nitsch, 1976, and modified by Apenburg, 1986). We analyzed only the following scales: Lack of Tension, Self-Assurance, Positive Mood, and Readiness to Work Hard. The scales relate to reports on mood and tension, on how one feels about oneself, on positive or negative affectivity, or on whether one wants to exert energy in work, respectively. Cronbach's alphas ranged from .80 to .90. The scales were used three times: before the experiment (t_1), after the forced-error training (sixth hour; t_2), and directly after the testing phase (t_3).

Typewriting skills: Because the subjects differed in their typewriting skills, we wanted to have at least a rough estimate of their ability. Therefore, the experimenters rated their skills on a 5-point scale. This variable was used as a covariate in the analyses of speed test variables (error count, number of words written, and average error-correction time).

3. RESULTS AND DISCUSSION

3.1. Mean Differences Between Experimental Groups

Figure 1 presents the mean differences between the treatment groups. The error-training group was significantly better in free recall. This may have been the result of a more adequate mental model of the error-training group, of higher motivation to learn, or of a deeper level of processing (in the sense of Craik & Lockhart, 1972) in the error-training group. The error-training group was also higher in competence for dealing with difficult tasks and (marginally significant) in the overall performance rating than the error-avoidant-training group. In accordance with our hypothesis, the difference in competence was more pronounced with difficult tasks than with easy tasks. In contrast to our hypothesis, the error-training group was not better in transfer competence. We are not quite sure why this is so.

There are additional interesting findings. The correlation between the performance rating in the transfer task and the number of keystrokes was higher in the error-training group ($r = .61$, $n = 15$, $p < .01$) than in the error-avoidant-training group ($r = .30$, $n = 8$, ns). Apparently, the error-training group learned to explore, and those who explored were also better; this was not so in the other group. This is an interesting finding for methodological reasons as well. A keystroke approach, which follows from Roberts and Moran (1983) and which was used by our work group in previous studies (e.g., Frese et al., 1988; Wendel & Frese, 1987), would suggest that

Figure 1. Mean differences between the two treatment groups.

Dependent Variable	Error Training ^a		Error Avoidant Training ^b		<i>t</i> or <i>F</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Free recall	13.20	5.99	7.66	2.74	<i>t</i> = 3.08**
Competence					
Easy	3.07	.98	2.63	.87	<i>t</i> = 1.12
Middle	2.95	.79	2.42	.82	<i>t</i> = 1.56
Difficult	2.47	.71	1.81	.50	<i>t</i> = 2.57**
Transfer competence	6.20	2.14	5.50	2.14	<i>t</i> = .75
Overall performance rating	2.93	.88	2.44	.53	<i>t</i> = 1.70 [†]
Speed test variables:					
Error count	14.66	5.28	21.00	7.95	<i>F</i> = 6.40*
Average error correction time	14.90	10.76	7.05	3.77	<i>F</i> = 4.12 [†]
Number of words written	138.40	51.30	171.80	36.90	<i>F</i> = 2.20
Satisfaction	1.19	.23	1.16	.26	ns

Note. *t* tests for independent samples were used, except in the three speed test variables error count and average error correction time and number of words written where an analysis of covariance with typewriting skill was used as the covariate.

^a*n* = 15. ^b*n* = 9.

p* < .05. *p* < .01. [†]*p* = .056. [‡]*p* = .052.

there should be a negative correlation between performance and keystrokes. Apparently, the raters were able to differentiate among those who tried blindly (with many keystrokes), those who tried systematically (with many keystrokes), those who did not try much at all (with few keystrokes), and those who could solve the problem very efficiently (again with few keystrokes).

The speed test variables did not provide evidence for a superiority of the error-training group. Although the subjects in the error-training group made significantly fewer errors in the speed test, this does not mean that the error-training group was better. There was a tendency for this group to write less text under speed conditions (this difference was not significant, however). Moreover, the error-training group needed more time to correct those errors they made (this difference is only marginally significant). The well-known speed-accuracy tradeoff may have been operative here. The error-training group was less time pressured by writing a lot and, therefore, made fewer errors. Actually, both groups were told to work as quickly as possible. However, the error-training subjects may have become more interested in learning from errors as a result of the experimental procedure. Therefore, they took longer to correct an error and did not write as many words as the error-avoidant group.

Figure 2. Correlations of CFQ with performance in the two treatment groups.

	Error Training	Error Avoidant Training	Difference z
Competence/Difficult	.02	-.61*	1.38 [§]
Free recall	-.01	-.66*	1.56 [§]
Transfer competence	.34	-.72*	2.52**

Note. Difference between correlations was calculated after r -to- z transformation.

* $p < .05$. ** $p < .01$. [§] $p < .10$.

The two groups were equally satisfied with their training. This result goes along with the observation that satisfaction measures are bad indicators of the effectiveness of training (Frese et al., 1988).

3.2. Intervening Variable of Cognitive Failures

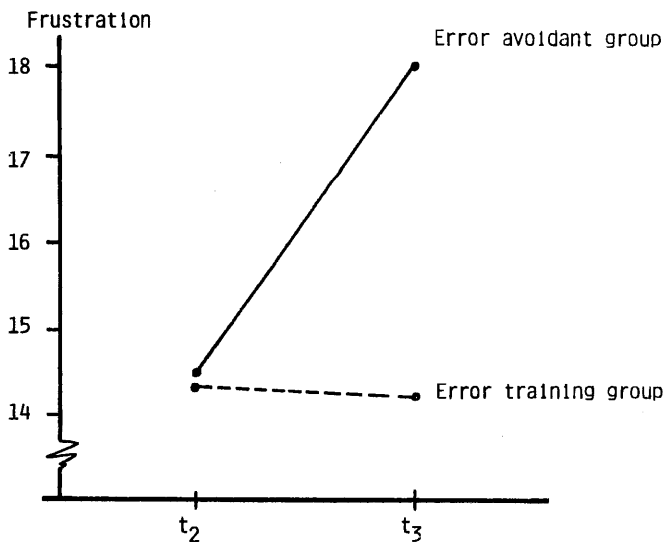
Cognitive failures measured with the CFQ is conceptualized to be a relatively stable personality variable; therefore, the mean of the two measures of CFQ was used. Figure 2 shows that there were significant negative correlations between CFQ and performance variables only in the error-avoidant group. The differences between the correlations was either significant or marginally significant. The results do not support our first hypothesis but rather the second one. In general, one would expect negative correlations of the CFQ with performance in a task area like computer work. However, this picture was only true of the error-avoidant group. Thus, "scatty" people profited from error training but did not profit from the error-avoidant training. Error training may have taught them to deal with their errors effectively.

3.3. Mood Changes Across the Experiment

In terms of the Present State Report and frustration, we assumed that negative mood and frustration would first rise in the error-training group and then fall, whereas the opposite pattern should appear in the error-avoidant-training group. The latter group would first have an easy task of just following the written instructions in the training process. But they should be frustrated in the testing period when they find their working with the computer to be not as easy as they might have assumed in the training sessions.

Figure 3 shows the means for frustration. There were no changes in the error-training group. However, the error-avoidant-training group became more frustrated in the second measurement period, $t(9) = -2.13, p < .05$,

Figure 3. Changes in frustration for the two treatment groups.



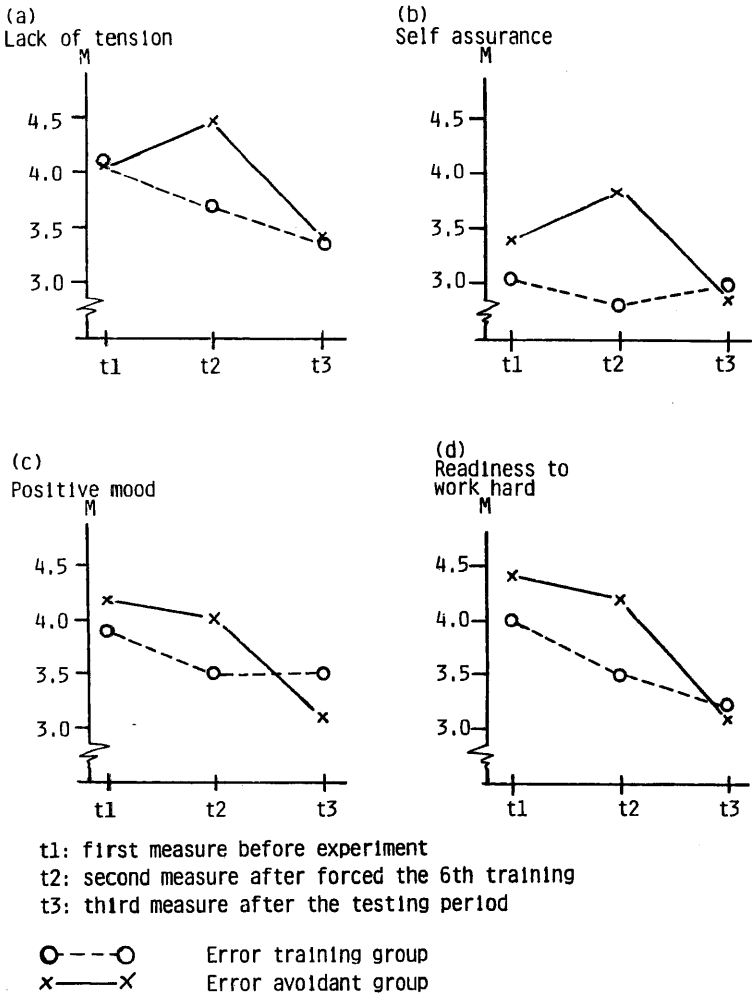
t_2 : Measurement after forced error training

t_3 : Measurement after the performance test

t test for dependent sample. The first measurement was taken directly after the forced-error training. The tasks during this period were made easy for the error-avoidant-training group because they just followed instructions on how to solve these problems. The second measurement occurred when the error-avoidant-training group learned in the testing phase (in which no subject was allowed to use the materials previously handed out) that they had not mastered the system yet. Of course, the error-training group also had not completely mastered the system, but they had at least learned to deal with their errors.

The results of the Present State Report presented in Figure 4 corroborate this interpretation. There was an additional measurement point before the training started. Again, the error-avoidant-training group showed a steep decline in mood from the easy training (t_2) to the hard test performance phase (t_3). This decline was significant for all mood variables, tested with *t* tests for dependent samples, $t(9) = 3.98$, $p < .01$ for tension; $t(9) = 2.47$, $p < .05$ for self-assurance; $t(9) = 2.23$, $p < .05$ for positive mood; $t(9) = 4.40$, $p < .01$ for readiness to work hard. In contrast, the error-training group did not exhibit any changes in mood during the experiment (except one significant

Figure 4. Changes in emotional processing in the two treatment groups.



drop from the beginning of the experiment [t1] to t3 in lack of tension).

Thus, there is evidence for a differential change of subjective feelings depending on the treatment group. The error-avoidant-training group showed sharp declines in mood from t2 to t3. In contrast, the error-training group was not much affected by their "ordeal" of going through the enforced-error training. They may have learned strategies to manage their errors, and the heuristics probably helped to make them less frustrated when errors appeared.

4. OVERALL DISCUSSION AND CONCLUSIONS

The most important finding of this study is the predicted superiority of the error-training group in solving difficult tasks under nonspeed conditions. This difference is corroborated by the overall performance evaluation of the experimenter but not by the results on transfer competence. The latter pointed in the right direction but was far away from significance. Furthermore, the better performance of the error-training group in free recall suggests a well-organized mental model. Error training seems to have helped those people who otherwise make many mistakes: There was no negative correlation (usually expected and true of the error-avoidant group) between the CFQ and performance in this group. However, the error-training group was not better in the speed test. The subjects of this group may have been more motivated to learn from their errors. Therefore, they wrote less words and took more time to correct their errors than the error-avoidant group.

The results of this experiment need to be cross-validated with larger samples and with observations of the processes that operate during the learning process. The experiment was a first attempt to look into error training and is unable to answer certain questions. For example, the role of speed tests has to be examined.

Furthermore, we are not sure about the processes of how the superiority of the error-training group in the nonspeed performance test came about. There are four possible mechanisms: (a) The error-training group might have learned more effective strategies of dealing with their errors. (b) It may have learned emotional strategies that made it less stressful to deal with errors. This, in turn, led to higher performance. (c) The error-training group may have been more motivated (more on this later). (d) Error training may have provided a deeper level of processing, leading to better memory.

We did not measure the work strategies of the subjects directly. As a next empirical step, strategies of people confronted with errors have to be researched.

These strategies are possibly affected by mood changes. The results of the mood and frustration scales suggest that the error-training group was in a better mood and had less frustration at the end of the experiment than the error-avoidant group (Greif, 1986, even likened his type of error training to a stress inoculation training). From our qualitative observations, we think that the positive effects of errors are partly due to presenting and repeating the heuristics. Errors are stressful, and they may be even more stressful in the German culture where perfectionism is highly valued. We observed many stress responses—blushing, not wanting the experimenters to see the errors, or becoming jittery—in the trainees when they made an error.

On the other hand, qualitative observations also showed us that the error-training group was more motivated. Although the participants in the

error-avoidant group left after the training without waiting much longer, it was difficult to persuade the trainees of the error-training groups to leave the computer. They "bombardeed" us with further questions and seemed much more interested. This may be partly due to a sort of *Zeigarnik* effect (Ovsiankina, 1928)— the difficult problems had not been completely solved, thus, they were resumed again. This phenomenon also reminds us of the literature on the achievement motive, which showed that problems of moderate difficulty lead to higher motivation to solve them (Heckhausen, 1980). The higher motivation may also have contributed to more rehearsal of the commands, leading to their higher recall.

Finally, level of processing theory (Craik & Lockhart, 1972) may account for the results as well. According to this theory, memory traces are stronger and more stable when the processing of the information is done at deeper levels. For example, if the meaning of certain commands is contemplated, this would constitute a deeper processing of the information. Thus, although the problems given to groups were the same, being able to make errors may have forced the error-training group to a deeper processing. This produced higher learning and retention. The fact that the error-training group showed better free recall speaks for such an explanation.

Probably all of these four potential mechanisms have had an influence in this experiment. They may be complementary rather than competing explanations for the superiority of the error-training group. However, future research may want to tease apart the workings of these mechanisms.

If our results can be cross-validated, they may have theoretical and practical implications. In terms of theory, the "bad name" of errors in various approaches to training may have to be reversed. In terms of practical issues, errors could be used constructively in the training process. Error training right at the beginning of the training probably is not optimal. At this point, trainees are still in the process of mastering the first steps in dealing with the computer. Rather, error training should be reserved for the middle part of the training when a certain level of confidence has been achieved. Some approaches have suggested that not giving any feedback (and thus no error feedback) may lead to good training results (Carroll & Carrithers, 1984; Carroll & Kay, 1985). We suggest that these approaches are probably most useful at the very beginning of novice training. Afterwards, error training may be more adequate and lead to better results.

Obviously, there are many research questions not answered by our study. The processes by which people deal with and learn from errors need to be better known. These processes may be different for different types of errors (Brodbeck, Zapf, Prümper, & Frese, 1989; Norman, 1981). We do not know whether strategy, memory, motivational, or emotional effects are important or whether all of them have an influence. The question of sequencing— which training method is best at which point in time during the training process—

has not been researched in the area of computer training. Furthermore, the answer to the question — what kind of impact have heuristics had on strategies and experiencing of the person? — is still not clear. Nevertheless, we think that it is useful to rethink the role of errors in the training process.

REFERENCES

- Altmann, A. (1987). *Lernprozesse an Textverarbeitungssystemen: Direkte Manipulation vs. konventionelle Interaktion* [Learning processes with text processing systems: Direct manipulation vs. conventional interaction]. Unpublished master's thesis, University of Munich, Department of Psychology.
- Apenburg, E. (1986). Befindlichkeitsbeschreibung als Methode der Beanspruchungsmessung [Feeling states as a method of measuring strain]. *Zeitschrift für Arbeits- und Organisationspsychologie*, 30, 3-14.
- Bødker, S. (1989). A human activity approach to user interfaces. *Human-Computer Interaction*, 4, 171-195.
- Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, 21, 1-16.
- Brodbeck, F. C., Prümper, J., & Zapf, D. (1987). Was denken Experten über Fehler in der Mensch-Computer Interaktion? [What do experts think of errors in human-computer interaction?]. Unpublished manuscript.
- Brodbeck, F. C., Zapf, D., Prümper, J., & Frese, M. (1989). *Error handling in office work with computers: A field study*. Unpublished manuscript, University of Munich, Department of Psychology.
- Brown, J. S., & Newman, S. E. (1985). Issues in cognitive and social ergonomics: From our house to Bauhaus. *Human-Computer Interaction*, 1, 359-391.
- Carroll, J. M., & Carrithers, C. (1984). Training wheels in a user interface. *Communications of the ACM*, 27, 800-806.
- Carroll, J. M., & Kay, D. S. (1985). Prompting, feedback and error correction in the design of a scenario machine. *Proceedings of the CHI '85 Conference on Human Factors in Computing Systems*, 149-154. San Francisco: ACM.
- Carroll, J. M., & Mack, R. L. (1984). Actively learning to use a word processor: By doing, by thinking, and by knowing. In J. C. Thomas & M. L. Schneider (Eds.), *Human factors in computer systems* (pp. 13-51). Norwood, NJ: Ablex.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Frese, M., Albrecht, K., Altmann, A., Lang, J., Papstein, P. V., Peyerl, R., Prümper, J., Schulte-Göcking, H., Wankmüller, I., & Wendel, R. (1988). The effects of an active development of the mental model in the training process: Experimental results on a word processing system. *Behaviour and Information Technology*, 7(3), 295-304.
- Frese, M., & Sabini, J. (Eds.). (1985). *Goal-directed behavior: The concept of action in psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Greif, S. (1986). Neue Kommunikationstechnologien — Entlastung oder mehr Streß? Beschreibung eines Computer-Trainings zur "Streß-Immunsierung" [New communication technologies — Ease of burdens or more stress? Description of a computer

- training for "stress immunization"]. In K. H. Paulig, U. Schäkel, & J. Scholz (Eds.), *Stress*. Hamburg: Windmühle.
- Greif, S., & Janikowski, A. (1987). Aktives Lernen durch systematische Fehlerexploration oder programmiertes Lernen durch Tutorials [Active learning through error exploration or programmed learning by tutorials]. *Zeitschrift für Arbeits- und Organisationspsychologie*, 31, 94-99.
- Greif, S., & Keller, H. (1990). Innovation and the design of work and learning environments: The concept of explanation of exploration in human-computer interaction. In M. A. West & J. L. Farr (Eds.), *Innovation and creativity at work* (pp. 231-249). New York: Wiley.
- Heckhausen, H. (1980). *Motivation und Handeln* [Motivation and action]. Berlin: Springer.
- Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). New York: Harper & Row.
- Nitsch, I. R. (1976). Die Eigenzustandsskala (EZ-Skala): Ein Verfahren zur hierarchisch-mehrdimensionalen Befindlichkeitsskalierung [The Present State Scale: A method for hierarchical and multidimensional scaling of feelings]. In I. R. Nitsch & I. Udris (Eds.), *Beanspruchung im Sport*. Bad Homburg, Germany: Limpert.
- Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, 88, 1-15.
- Ovsiankina, M. (1928). Die Wiederaufnahme unterbrochener Handlungen [Taking up interrupted actions]. *Psychologische Forschung*, 11, 302-379.
- Prümper, J., Zapf, D., Brodbeck, F. C., & Frese, M. (in press). Errors of novices and experts: Some surprising differences in computerized office work. *Behaviour and Information Technology*.
- Roberts, T. L., & Moran, T. P. (1983). The evaluation of text editors: Methodology and empirical results. *Communications of the ACM*, 26, 265-283.
- Schulte-Göcking, H. (1987). *Die Erfassung von unterschiedlichen Lernstilen* [Measuring different learning styles]. Unpublished master's thesis, University of Munich, Department of Psychology.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.
- Wendel, R., & Frese, M. (1987). Developing exploratory strategies in training: The general approach and a specific example for manual use. In B. Shackel & H. J. Bullinger (Eds.), *Human-computer interaction, Proceedings, Interact '87* (pp. 943-948). Amsterdam: North-Holland.
- Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Norwood, NJ: Ablex.
- Zapf, D., Brodbeck, F. C., Frese, M., Peters, H., & Prümper, J. (1990). *Errors in working with office computers. A first validation of a taxonomy for observed errors in a field setting*. Unpublished manuscript, University of Munich, Department of Psychology.

HCI Editorial Record. First manuscript received March 6, 1989. Revisions received January 8, 1990, and June 29, 1990. Accepted by Donald Norman. Final manuscript received September 3, 1990. — *Editor*
