Journal of Occupational and Organizational Psychology (1993), 66, 303–317 *Printed in Great Britain*
© 1993 The British Psychological Society

3

Error handling in office work with computers: A field study

Felix C Brodbeck*, Dieter Zapf, Jochen Prumper and **Michael Frese**

Fachbereich Psychologie, Universitat Giessen, Otto-Behagel Straße 10/F, 6300 Giessen, Germany

An observational field study gives an account of error types, error handling time and use of support in case of an error situation when working with computers in the office Subjects were 198 clerical employees from 11 companies and seven small firms in Germany The analyses are based on 1155 observed errors which were concordantly classified into an error taxonomy by two independent re-raters Clerical employees spent approximately 10 per cent of their computer working time handling errors Error handling time is also positively related to indicators of emotional strain Although the tasks performed were largely routine, more than 11 per cent of all errors required the use of supports such as advisory services, co-workers, on-line help and menus or user manuals Different error classes showed variations in the amount of support used and in error handling time On the basis of the results, we discuss how the error taxonomy and measures of the human error handling process can be of practical use for evaluation in software ergonomics and for improving human error handling while interacting with computers

Errors inevitably occur every day at work and they incur economic costs A detailed analysis of a single error was done by Smelcer (1989) From his experiments, he estimated the average time spent correcting errors related to the 'joint statement' command of Structured Query Language This he converted into an estimate of $58 million lost in the US per year due to this single error situation Card, Moran & Newell (1983) found in one of their experiments that 26 per cent of the total time for text editing was spent dealing with errors However, errors give rise not only to economic costs, but also to psychological stress and the perception that one's job is unpleasant (Frese, 1987, Johansson & Aronsson, 1984)

There is a large body of literature on human error and underlying processes (see Reason, 1990 for a current overview, and Fitts & Jones, 1961 for a classic paper) Due to the fact that errors traditionally have been investigated in close relation to accidents in high risk settings, general emphasis is put on the genesis of errors and on the prevention of their occurrence The job is done when error rates are minimized Another reason for the common preference for error prevention may be that most traditional man–machine systems were designed to be unforgiving to errors (Perrow, 1984)

If we differentiate between the occurrence of an error and its consequences it becomes clear that often human action aims at avoiding negative consequences of errors and not

* Requests for reprints

necessarily at avoiding errors themselves When a person stumbles an error occurs, but the person succeeds in coping with it The negative consequence of falling down could be avoided It would be a rather cumbersome endeavour to walk carefully enough in order to prevent oneself from stumbling at all Thus for a specific array of human actions it seems to be more beneficial to cope adequately with negative error consequences than to put limitations on the performance of error prone actions This strategy could be adopted for human–computer interaction in certain work settings

For this it is important to know more about different types of errors, about what is done after an error has happened, and about how to support the human error handling process during human–computer interaction

The question may arise why an observational field study is needed to analyse human error handling processes The limited number of studies that systematically look at what people do after they have committed an error were conducted in laboratory settings (Zapf, Brodbeck, Frese, Peters & Prumper, 1992) For example, laboratory studies do not tell us what people do when they have committed an error in real work situations and how much time is spent handling different types of errors in real work situations The results of laboratory studies cannot be easily generalized to normal working conditions There is always a threat to their ecological validity, i e the generalizability to everyday situations (Neisser, 1976)

Thus, the present field study was designed to give an empirical descriptive account of error handling during human–computer interaction in the natural environment at work It aims at a clearer understanding of what happens after an error has occurred How long does it take to handle errors while working with computers? What supports are employed for handling errors at work with computers? And how do supports and error handling time relate to potentially different types of errors?

Theoretical concept of error

What is an error? It is a useful assumption that human behaviour is a result of goal-oriented action (Frese & Sabini, 1985, Hacker, 1986, Miller, Galanter & Pribram, 1960, Norman, 1981) Within an action-oriented framework the following assumptions about errors can be made (Zapf *et al* , 1992) First, errors only appear in the pursuit of a goal If a person presses keys at random, an error cannot be committed Second, an error implies non-attainment of a specific goal or a higher order goal A specific goal may not be attained, for example, because one reaches a 'dead end', as when a file is deleted and cannot be recovered A higher order goal may not be attained when plans are inadequate or wrongly arranged The latter is the case when a set of correct specific goals is attained but the sequence of subgoals is set up wrongly and the goal of the whole plan cannot be attained (Reason, 1990) Efficiency of work performance can also be a higher order goal Thus taking an action detour can be seen as an error as well Finally, the non-attainment of a goal should be potentially avoidable If something is not avoidable it will not be considered an error (e g when sabotage or a blizzard cause loss of data)

An advance in error research has been the *mismatch concept* (Rasmussen, 1985) It attributes the cause of an error to the interaction between an individual (e g a computer user) and his or her environment (e g a computer system) and not to either one alone In nearly every case one can argue that an error could have been prevented had either part of a system been different (e g a more knowledgeable user or a better designed computer) Following the mismatch concept, one does not attribute an error to any cause *per se* However, attribution to a system component is important for practical purposes The decision to change only one part of a man–machine system should then be taken on the basis of empirical data For example, one may change the software if many people make the same error, as in the case of the joint-command in the Structured Query Language (Smelcer, 1989)

Error classification

There are two kinds of mismatches that are of interest for errors in human computer interaction the mismatch between the work task and the computer, and the mismatch between the user and the computer The former gives rise to functionality problems, the latter to so-called usability problems (Zapf, Brodbeck & Prumper, 1989)

Functionality problems Functionality problems are identified when the computer system makes it more difficult for the user than is technically feasible, or it does not even allow reaching a goal that is implied by a specific work task Functionality problems imply that the computer system and the task are not well adjusted to each other (e g when a spelling program is not able to deal with plural words) To a certain extent functionality problems can be identified on the basis of task descriptions and the computer application, independently of observing computer users performing their tasks However, in order to find out how functionality problems are handled, the consequences they impose on the user's actions have to be observed They impose on the user different types of action consequences (a) action blockade users are forced to give up task-specific goals when a certain task-relevant action cannot be performed with the software being used, (b) action repetitions parts of the user's work are lost and actions have to be performed again, (c) action interruption ongoing work is interrupted for an inappropriate amount of time, (d) action detour specific functional deficits of the computer system have to be worked around by the users Note that the diagnostic reference point of functionality problems is a mismatch between a set of goals that is implied by the description of a working task and the functionality of an application software

Usability problems A usability problem is rated when users do not attain their individual goals and a functionality problem cannot be assumed This implies that users' actions cannot be inferred solely on the basis of descriptions of their work tasks Erroneous actions have to be observed The diagnostic reference point of usability problems is human action and its goals This does not mean that human action is the only cause of a usability problem—the latter follows from the mismatch concept

Since usability is concerned with people's actions, usability problems are further subdivided according to several parameters of a theory of human action regulation (Zapf *et al* , 1992) For our purposes it is sufficient to explain only one dimension of the error taxonomy Within the German tradition of action theory (Frese & Sabini, 1985, Hacker, 1973, 1986, Volpert, 1982) three levels of action regulation are differentiated intellectual level of regulation, level of flexible action patterns, and level of sensorimotor regulation[1] The three levels of regulation portray the dimension of controlled vs automatic processes (Shiffrin & Schneider, 1977) Furthermore, information for regulating actions is based on the knowledge base for regulation

Intellectual level of regulation This is characterized by a conscious development and activation of goals and plans This is necessary for working on complex tasks or tasks that have not often been performed For example, a spread-sheet table was supposed to fit on one page After working a while the user realized that the columns were too wide to fit the page Therefore, she had to change the column widths

Level of flexible action patterns On this level well-practised actions are regulated by plans stored in memory which have to be adjusted to particular situations The concept of flexible action pattern is comparable to the concept of schemata described by Norman (1981) Less conscious attention is required on this level, especially for triggering execution of a routinized plan and for evaluating the outcome The plans are of higher specificity and routinization than the plans on the level of intellectual regulation An error that belongs to the level of flexible action patterns happens when well-practised rules are triggered in wrong situations (For that type of error Norman 1981 coined the term 'mode error') Such an error happens, for example, when a person switches from using one word processing system to another and uses function keys that were correct in the former situation but are incorrect in the latter

Sensorimotor level On this level highly routinized actions are regulated without conscious attention Sensorimotor errors include such acts as wrong mouse movements or touching the wrong function key located on a position that is nearby or similarly exposed as the intended one[2]

Knowledge base for regulation In the knowledge base for regulation facts and understanding are stored to develop specific goals and plans Knowledge errors exist when a user is unable to attain a goal because s/he doesn't know certain command names, function keys or concepts

It is not the purpose here to discuss the theory behind the classification in detail (this was done in Zapf *et al* , 1992) It may be sufficient to note the similarities of these levels of regulation with Rasmussen's (1983) and Reason's (1990) distinction of knowledge-based errors, rule-based errors and skill-based errors However, in contrast to Rasmussen and Reason, we distinguish further errors that relate to either the intellectual level of regulation or the knowledge base for regulation

Error handling

Error handling includes everything that is done by the user after an error has occurred The measurement of error handling time is used to give an account of the overall

[1] The term regulation is used in the same sense as Miller *et al* (1960) use control, i e the order of instructions for actions are given from a central processing unit in accordance with the person s goals, the plans to achieve the goal and the feedback from the environment The term regulation is less confusing because the term control is already adopted by other concepts in psychology (e g in the sense of internal vs external control)

[2] Typos during text editing are definitely sensorimotor errors However, they are not considered further in our investigation because errors that occur while writing continuous text were excluded from observation

3P

percentage of working time used for the recovery of errors This provides an estimated margin for how much costs can be reduced One would also suppose that error handling time is associated with different error classes, errors on the intellectual level of regulation and the knowledge base for regulation showing the highest error handling times Frequency of support use can help to answer questions about how to improve users' error handling Relations between different error classes and use of support help to identify specific user needs that support facilities should be suited to Here one would expect that more support is needed for errors due to lack of knowledge because additional information is needed The evaluation of the relation between emotional stress and error handling time can shed some light on the question of whether reduction of error handling time reduces economic as well as human costs Stress research suggests that errors requiring a higher effort, i e a longer handling time, are seen to be more stressful (Schonpflug, 1985)

For a clearer understanding of human error handling in a natural work setting, this article reports empirical data on the aspects described above of the error handling process In addition, we reconsidered whether a theoretically derived error taxonomy and measures of the error handling process can be of use in field settings Practical implications for improvement of error management facilities are discussed in a final section.

Method

Subjects

Fifteen departments (in 11 different companies) and seven small firms in Germany participated in the study The size of the units ranged from a department of a large public administration service with 260 employees to small trading firms with two to three employees The companies and firms provided a wide range of administrative tasks in branches like automobile production, electronics and computer software, banking, insurance and public administration

A total of 259 office workers (secretaries, specialists, lower level managers) were included in the study Participants were interviewed and observed at their work place Afterwards they filled out a standardized questionnaire For organizational reasons, not all subjects could be both observed and questioned In this report all analyses are based on the 198 subjects that participated in the observational investigations

Most of the participants worked for the company for at least three years and reported an average computer experience with the observed software application of at least one year A differential analysis of the degree of subjects' expertise is given by Prumper, Zapf, Brodbeck & Frese (1992) The mean age of subjects was 31 1 years, range 16 to 60 years, 73 per cent were females and 27 per cent were male

An essential component of each subject s job was computer work The average percentage of daily computer work time was in the category of 50 to 60 per cent The array of computer tasks observed included data entry, text editing, graphic design, database management, spread-sheet use and electronic mail Only in very few cases were workers being observed while adjusting their application software to their own needs by use of simple language programming tools Mainly, menu-driven dialogue interfaces in combination with command dialogue techniques were employed for task performance Direct manipulation (especially for graphic design) and simple mask dialogue (especially for data entry, and partly for database management) were also used Altogether, the subjects tasks were not very complex (a more detailed description of task complexity and control at work is given by Zapf, 1991)

Procedure and measures

In a first step the observers tried to become familiar with the user's job This was done by applying Part A of the German job analysis instrument VERA, developed by Volpert, Oesterreich, Gablenz-Kolakovic, Krogoll & Resch (1983) In addition, a rating instrument for job complexity and control at work (Semmer, 1984) had to be filled in by the observers after the observation period The quality of different support facilities for error handling was assessed by use of the standardized questionnaire

During a two-hour observational period the tasks performed and error descriptions were recorded on paper by the observers Simple typing errors were excluded from the analysis, since it was assumed that their extremely high frequency would substantially distort error distributions and error handling time The observers sat next to or behind the subjects in order to see both the screen and the keyboard In addition, descriptions of disruptive events (e g telephone), ratings of error detection, error handling time, use of supports and negative emotional reaction during error handling were recorded on a structured protocol sheet Observers were students and researchers who already knew the underlying theory and the error taxonomy All observers were trained to use the error taxonomy, the protocol sheet and appropriate interview techniques in a three-day hands-on course

Error handling time In the field observational study the error handling time was estimated from error detection to the end of error correction It was not advised to use a stop-watch in the firms The use of a stop-watch was not allowed by shop stewards, and fear of being timed for performance would inhibit the cooperation of the subjects Especially when recording errors, a visible time measurement can cause the subjects to select tasks for the observation period that they can perform easily In order not to endanger ecological validity, error handling time was gauged by an observer estimate on a five-point scale from immediately, up to two minutes, up to five minutes, up to 10 minutes, 10 minutes and longer For that wrist-watches and other existing clocks (on the office desk or on the wall) were used in a way to make sure that the subjects did not become suspicious Since these data are ordinal, statistical analyses comparing error handling time were calculated with non-parametric methods For explanatory purposes ordinal scale data are transformed into interval scale data on the basis of the arithmetical mean of the category boundaries Additional studies were performed in order to justify this procedure[3] In a different study, data from 23 subjects were gathered from two independent observers using the same procedure (see Herbrich, Frese & Prumper, 1991) The inter-rater reliability for observers' time estimates showed an ordinal data Kendall's tau of $\tau = 69$

Use of support facilities The use of supports was rated with the following categories (a) use of manuals, which most often were shortened versions either self-made or provided by the companies, (b) help system and menu information on the screen, (c) asking the advisory service, and (d) asking a co-worker According to our reliability study with a subset of 23 subjects, the use of supports was concordantly rated by two observers for 96 per cent of all error events This is equivalent to a Cohen s kappa of $\kappa = 81$, $N = 123$ (Cohen, 1960)

Error classes Originally, 15 different error classes were recorded by the observers Four of the 15 error classes were excluded because measuring error handling time would be rather difficult (This makes the overall time estimate more conservative (Zapf *et al* , 1992)) The rest are subclasses of the five previously described error categories For the purpose of this article, a more precise account of the error classes is not needed (a complete description is given by Zapf *et al* , 1992) A functionality problem was identified when the computer system was determined to be insufficient for the tasks involved and the users were forced into an action blockade, an

[3] For reasons of ecological validity, error handling time had to be evaluated on ordinal scale in this field study Since the latent variable of error handling time is interval scale it seems reasonable to try some justification for using arithmetical means of the ordinal scale indicators A comparison between interval scale estimators of error handling time and the transformed data presented here has been undertaken For every error type of the field study the arithmetic means of the observers ordinal scale time estimates were substituted by the mean values of a computer-driven time measure that was also recorded to observers time estimates in a separate study by Zapf, Lang & Wittmann (1991) This study was conducted with conditions similar to the field setting Text editing was observed with a commercial word processing application The Spearman rho between observers time estimates and the computer driven measure was $r = 75$ ($N = 118$, $p < 01$) The arithmetical means from the field study and the computer-driven means are compared for each interval of the time estimation scale Results show that for the time intervals immediately up to two minutes and up to five minutes the computer-driven time measures are within the interval boundaries For these three intervals the arithmetical means are acceptable estimates of the computer measured values However observers do overestimate the two higher time intervals of up to 10 and over 10 minutes Only 3 per cent of all error events fall into these latter categories Thus only a few errors have been overestimated by the observers Given these data, we conclude that the observational method would not fare too badly and can be used especially in field settings when other means like stop-watches or computer-driven timing are not possible

action repetition or an action interruption To support the classification process for different usability problems, eight guiding questions had to be answered by the observers These include 'Rate the degree of routinization of the action underlying the error event' and 'Are all action steps known to the user? In those cases where it was not possible to classify the errors, the observers had to ask the subjects for the missing information after the error was corrected In addition, a short description of each error was recorded

The inter-rater reliability of two observers for these five error classes is characterized by a concordance rate of 77 per cent and a Cohen's kappa of $\kappa = 70$ ($N = 123$) The main observational study did not use two observers A proxy to inter-rater agreement was developed (Prumper, 1991) Based on the written error descriptions, two re-raters rated the errors into the 15 uncollapsed error classes referred to above The inter-rater concordance rate based on the five collapsed error classes was 82 per cent (Cohen's kappa, $\kappa = 77$, $N = 1415$) To improve the quality of the data only those errors concordantly rated were included in the analyses, as was done by Allwood (1984), Bagnara, Stablum, Rizzo, Fontana & Ruo (1987) and Rizzo, Bagnara & Visciola (1987) as well This leads to a pool of 1155 concordantly rated error events

Negative emotional reactions Negative emotional reactions as an indicator of stress were rated on a five-point scale (very strong, strong, intermediate, slight, no reaction) Angry verbal outbursts, verbal expression of frustration, or nervous tension displayed through the intonation of verbal comments were taken as indicators of negative emotional reactions The distribution of the observers' ratings was strongly skewed due to the high proportion of errors that produced no observable emotional reactions The observers could discriminate best between no or slight emotional reactions on the one hand and intermediate, strong or very strong emotional reactions on the other Accordingly, the ratings were divided into two classes The concordance rate for two observers was 96 7 per cent and a Kendall's tau of $\tau = 64$ ($N = 120$) indicates moderate to high reliability

Results

On average subjects were observed for a period of 109 minutes During this time they worked on average for 85 minutes with the computer The average number of errors that occurred during the observed time period is 8 8 per person The average error handling time per person across all error classes using the arithmetical means is 8 5 minutes (minimum 15 s, maximum 85 25 min, $N = 1306$)[4] That is, 10 per cent of the computer working time is spent handling errors We consider this percentage to be high for several reasons (a) our time estimation method underestimates error handling time, (b) the subjects were very familiar with their jobs, (c) subjects' tasks were not very complex, (d) subjects' computer expertise is relatively high—most of them have worked for at least one year with their respective computer software (Prumper *et al* , 1992)

Use of support facilities

After an error was detected, subjects usually tried to solve the problem themselves and succeeded in doing so in 85 2 per cent of error events Table 1 shows that 3 6 per cent of all errors were not successfully corrected, that is, in these cases handling attempts did not lead to a correct solution or the subject did not even try to handle the error Only one case showed an unsuccessfully handled error after the use of support sources (0 1 per cent) This should not be interpreted as meaning that support is nearly always successful Often the users already knew that support would not help them to recover from a particular

[4] The overall average error handling time can be calculated independently from particular ratings of error classes Therefore error events that were not concordantly rated are included Thus in this particular case the analysis is based on $N = 1306$ error events

Table 1 Distribution of error handling strategies

Error handling	N^a	Per cent
Successful		
Without support	940	85 2
With support	123	11 2
Unsuccessful		
Without support	39	3 5
With support	1	0 1
Total	1103	100 0

[a] Number of errors per error handling strategy

error, and they therefore did not even try to get help Of those errors which were successfully handled, 11 2 per cent required some kind of support, but the majority of error situations were handled without any support (85 2 per cent) Against the background of most tasks being well known and the considerable computer experience of the users, it is surprising that as many as 11 2 per cent of the errors could only be recovered after using support Thus support is clearly necessary

An example may demonstrate unsuccessful error handling A user transformed an array of numbers on a spread-sheet into a graphical representation on the screen The headers were supposed to be formatted in a different font However, the user could not find the option for doing this She finally gave up the original intention and the headlines were printed in the default format

Table 2 presents a more detailed picture of the successful error handling events with supports used (N = 123) Users sought support most frequently in the form of advice from co-workers, on-line help and menu facilities User manuals and computer advisory services were of secondary importance In some cases (N = 18) more than one type of assistance was required to handle a single error—in 16 cases this involved asking a co-worker and in 16 cases it involved using on-line help and menus Thus asking co-workers and using help screens and menus actually occurred more frequently than is indicated by the individual percentages in Table 2 Co-workers provided the most frequently used support for handling errors People also preferred to ask co-workers (Brodbeck, 1991, Frese,

Table 2 Distribution of support used to handle errors successfully

Support used	N^a	Per cent
Co-worker	45	36 6
Help screens and menus	31	25 2
Advisory services	15	12 2
User manuals	14	11 4
Two or three combined	18	14 6
Total	123	100 0

[a] Number of errors per support facility used

Brodbeck, Zapf & Prumper, 1990) However, to concentrate solely on co-worker support would be misleading The data show that a variety of external supports were used

Support use and error classes

Table 3 shows a distribution of supports for each error class More than half (52 6 per cent) of the errors located on the knowledge base for regulation required the use of support This error class was significantly more closely associated with the use of support than all the other usability error classes combined ($\chi^2(1, 1063) = 232$ 7, after Yates' correction, $p <$ 01) Errors on the knowledge base for regulation require use of support more frequently than errors on the intellectual level of regulation ($\chi^2(1, 336) = 58$ 5, after Yates' correction, $p <$ 01) Subjects most frequently asked co-workers for help (note on 16 of the 18 occasions when more than one source of support was sought, both co-worker and help screens or menus were consulted) Manuals were used almost exclusively for handling knowledge errors The manuals used were mainly self-made scripts that contain information about function keys, menu structures and lists of code numbers

Table 3 Error classes and support used for error handling

		Percentage of errors		
Error classes	N^a	Support used ($N = 123$)	Support not used ($N = 940$)	No error correction ($N = 40$)
Knowledge base for regulation	118	52 6	41 5	5 9
Intellectual level	238	14 2	80 3	5 5
Level of flexible action pattern	285	3 0	94 2	2 8
Sensorimotor level	216	0	98 6	1 4
Functionality problems	201	8 5	87 5	4 0

[a]Number of errors per error class Total $N = 1103$

Supports were required relatively often for errors located on the intellectual level of regulation (14 2 per cent of the successfully handled errors), with significantly greater use of supports than for errors at the two other levels of regulation ($\chi^2(1, 759) = 48$ 4, after Yates' correction, $p <$ 01) There were errors on the intellectual level (5 5 per cent) and the knowledge base (5 9 per cent) which could not be corrected Unsuccessful error recovery at these two levels occurred significantly more often than at the level of flexible action patterns and at the sensorimotor level ($\chi^2(1, 902) = 6$ 4, after Yates' correction, $p <$ 05) The errors associated with these latter two levels of regulation could usually be corrected without support Support was required for as few as 3 per cent of errors associated with flexible action patterns and none was needed on the sensorimotor level

Error handling time and error classes

A Kruskal–Wallis test was performed showing that the higher the level of regulation the longer the error handling time ($\chi^2(5, 894) = 181\ 52, p < 01$) Error handling time for errors on the knowledge base and on the intellectual level of regulation is greater than for the other usability errors Functionality problems show a longer handling time than errors on the level of flexible action patterns and sensorimotor errors, but they do not differ significantly from knowledge errors and errors on the intellectual level of regulation

Error handling time and emotional reactions

Table 4 presents the data on the relationship between error handling time and negative emotional reactions which were recorded by the observers (signs of anger, frustration and tension) Of the errors that required handling for more than 10 minutes, 57 per cent produced a high degree of emotional upset, this was in contrast to only 7 6 per cent for errors which could be handled immediately This translates into a correlation of $r = 38$ ($p < 01$) Thus emotional strain is positively related to error handling time

Table 4 Percentage of errors with negative emotional reactions (anger, frustration, tension) and error handling time

Error handling time	N^a	Percentage of errors with negative emotional reactions
Immediately	608	7 6
Up to 2 minutes	330	15 5
Up to 5 minutes	127	33 9
Up to 10 minutes	11	36 4
>10 minutes	28	57 1

Note χ^2 (4, 1104) = 107 57, $p < 01$
[a]Number of errors corrected per time period
Total $N = 1104$

Discussion

The results provide a relatively cohesive picture About 10 per cent of computer working time is spent handling errors About 11 per cent of successfully handled errors required use of support, which is accompanied by a prolonged error handling time Given that the users knew their task and their computer systems quite well and that mainly routine tasks were being observed, both values are seen to be rather high

As was demonstrated, a reduction in error handling time goes along with reduced emotional strain Presumably, a reduction in error handling time would also go along with increased productivity For the last assumption some indirect support is given by Prumper *et al* (1992) They demonstrated that error handling time, rather than error frequency, is reduced as a function of the users' computer experience Experts commit about as many errors as novices do but they need less time to correct them

With these results in mind one could argue that for office automation a reduction in error handling time can be of benefit Apart from this, if one keeps in mind that many

9,C

errors inevitably occur a strategy of supporting human error handling seems to be a suitable response Another point against the potential of error prevention in software design is its limitation with respect to the detection of errors and the restriction of error-prone actions The strategy of error prevention implies that one is able to anticipate specific errors that will appear in a certain program environment in a given task context Zapf, Maier, Rappensperger & Irmer (in press) present evidence that the extent to which errors can be anticipated by the computer system is limited because about half of the errors can only be detected if the higher order goals of the users are known This seems to be impossible for computer systems Thus a strategy of supporting human error handling—we call it error management—through systems design seems to be a good supplement to the strategy of error prevention Computer support in case of an erroneous action that is interpreted by the computer system as being a legitimate input sequence is a rather different problem from automatic error detection or the design of forcing functions (Norman, 1986), which are commonly used strategies for error prevention

If one is willing to accept the notion that error management in human–computer interaction is important, questions about how to support error handling can be raised They are now discussed in the light of our results

When errors were located on the intellectual level and on the knowledge base for regulation, error handling was more complicated, more time consuming, more upsetting and required more help These error classes pose relatively complex problems that require the user's full attention, an extra amount of time, differential support and strategies for coping with emotional strain

In contrast, there was a low average error handling time and less use of support for sensorimotor errors and errors on the level of flexible action patterns They seem to be adequately covered by the software design at present (see Shneiderman, 1987, Smith & Mosier, 1986) and by the subjects' computer experience However, there are still some errors (about 3 per cent) on the level of flexible action patterns where error handling is not a trivial task for the users For these errors, where only a few action steps are performed, the undo-function is one example of an error management strategy that can be of specific help error handling time It allows the user to reverse an operation and to achieve the *status quo ante* by typing one key (Yang, 1987)

However, the usefulness of an undo-function may not be the same for errors on the intellectual level of regulation and for knowledge errors These errors require more complex actions and the user would have to go back many more operations History functions or freezing points, provided in some systems, seem to be more appropriate for handling these errors They allow the user to go back a whole action path or to go back to a significant point in the action history (Paul, 1992, Paul & Foks, 1991) Empirical investigations of error handling time for different error classes may help to solve the debate on what kind of undo-function matches the users' needs while performing tasks of different complexity (the debate was described by Lenman & Robert, 1992)

While in the field setting all kinds of support were used—asking a co-worker, consulting a manual, help and menus and advisory services—co-workers were clearly consulted most often This is something that is not taken into consideration when computer applications are evaluated in laboratory settings Our results on use of support point towards a kind of error management that extends the view beyond pure software ergonomics With respect to the use of co-workers' advice our results are in line with Scharer's (1983) argument that co-worker support is most often sought This favours the concept of local experts or decentralized advisory services In Frese *et al* (1990) additional evidence is provided that a local advisory service is generally more often consulted and also preferred compared with a centralized advisory service In general, support is considered to be beneficial if it is quickly available and reduces error handling time Moreover, training of local advisory boards can improve support (Zapf & Scherubel, 1991)

Another error management strategy is to teach people to deal with errors effectively Several experimental studies provide evidence that hands-on performance is better and that emotional frustration is reduced by computer training that explicitly confronts the trainees with error-prone situations—called error management training (Frese, Brodbeck, Heinbokel, Mooser, Schleiffenbaum & Thieman, 1991, Greif & Janikowski, 1987, Irmer, Pfeffer & Frese, 1991)

While these results give some credit to the notion that human–computer interaction in the office could profit from a better understanding of the error handling process, there are clearly some methodological problems in our observational field study The reliability of error classification should be improved in future studies Note, however, that we mostly took into account those errors that were consistently reclassified from the written error descriptions by two raters Furthermore, the timing of error handling was not exact because we could not use a stop-watch However, there is some credence to our procedures as they lead to results that are similar to those of computer-driven time taking The emphasis on ecological validity also poses the well-known trade-off problem between internal validity and external validity, of which ecological validity is a subtype (Cook & Campbell, 1979)

While such problems do exist in a field study, the advantages should not be overlooked For example, Card *et al* (1983) argued that about 26 per cent of the time spent performing their text editing tasks was 'error time' Similarly, Shneiderman (1987) quotes an error rate of up to 53 per cent for the use of commands This appears to be highly exaggerated, given normal work tasks and relatively skilled users Our results can be generalized to a wider range of mainly routine tasks and subjects than the results of the experimental studies

Furthermore, a similar pattern of frequency distribution of error types and of error handling time per error class to the one reported here is shown in experimental studies by Zapf *et al* (1991) Thus the error taxonomy seems to be applicable in laboratory settings as well as in field settings

Finally, results also show that error handling time can be perceived as a good indicator of the difficulties users have in dealing with errors once they have occurred Error handling time is sensitive to several aspects of mismatch situations types of errors, use of support facilities and emotional strain

Realization of the practical importance of the error handling process may lead to a reconsideration of how to treat the problem of errors in office work with computers We believe that our theoretical and methodological approach can be of use for investigating the computer user's difficulties while handling errors—in the field and in laboratory settings The concept of error management (Frese & Altmann, 1989, Frese, 1991) advocates that error handling strategies should be aided by different approaches software design, training and advisory support systems In this area there is considerable scope for improvement

Acknowledgements

This article is based on the project FAUST (a German acronym for 'error analysis for the investigation of software and training) The project was supported by a grant from the Work & Technique (Humanization of Work) Fund of the Ministry of Research and Technology of the Federal Republic of Germany (01 HK 806 7) (F Brodbeck, M Frese (principal investigator), C Irmer, H Peters (Technischer Uberwachungsverein), J Prumper, D Zapf Special thanks are due to Linda Kershaw and Dave McKellin for their useful comments on an earlier draft of this article

References

Allwood, C M (1984) Error detection processes in statistical problem solving *Cognitive Science* 8, 413–437

Brodbeck, F C (1991) Fehlerbewaltigungsdauer und die Nutzung von Unterstutzungsmoglichkeiten In M Frese & D Zapf (Eds), *Fehler bei der Arbeit mit dem Computer Ergebnisse von Beobachtungen und Befragungen im Burobereich*, pp 80–94 Bern Huber

Bagnara, S , Stablum, F, Rizzo, A , Fontana, A & Ruo, M (1987) Error Detection and Correction A Study on HCI in a Hot Strip Mill Production Planning and Control System Paper presented at the First European Meeting on Cognitive Science Approaches to Process Control, Marcoussis, France, 19–20 October 1987

Card, S K , Moran, T P & Newell, A (1983) *The Psychology of Human-computer Interaction* Hillsdale, NJ Erlbaum

Cohen, J (1960) A coefficient of agreement for nominal scales *Educational and Psychological Measurement* 20, 37–46

Cook, T D & Campbell, D T (1979) *Quasi-experimentation Design and Analysis Issues for Field Settings* Chicago, IL Rand McNally

Fitts, P M & Jones, R E (1961) Analysis of factors contributing to 460 pilot-error experiences in operating aircraft controls In W H Sinaiko (Ed), *Selected Papers on Human Factors in the Design and Use of Control Systems*, pp 332–358 New York Dover

Frese, M (1987) The industrial and organizational psychology of human–computer interaction in the office In C L Cooper & I T Robertson (Eds), *International Review of Industrial and Organizational Psychology*, pp 117–166 Chichester Wiley

Frese, M (1991) Error management or error prevention Two strategies to deal with errors in software design In H J Bullinger (Ed), *Human Aspects in Computing Design and Use of Interactive Systems and Work with Terminals*, pp 776–782 Amsterdam Elsevier

Frese, M & Altmann, A (1989) The treatment of errors in learning and training In L Bainbridge & A Ruiz Quintanilla (Eds), *Developing Skills with Information Technology*, pp 65–86 Chichester Wiley

Frese, M , Brodbeck, F C , Heinbokel, T , Mooser, C , Schleiffenbaum, E & Thieman, P (1991) Errors in training computer skills On the positive function of errors *Human Computer Interaction,* 6(1), 77–93

Frese, M , Brodbeck, F C , Zapf, D & Prumper, J (1990) The effects of task structure and social support on users' errors and error handling In D Diaper (Ed), *Human-Computer Interaction—INTERACT '90*, pp 35–41 Amsterdam Elsevier Science

Frese, M , Irmer, C & Prumper, J (1991) Das Konzept Fehlermanagement Eine Strategie des Umgangs mit Handlungsfehlern in der Mensch-Computer Interaktion In M Frese, Chr Kasten, C Skarpelis & B Zang-Scheucher (Eds), *Software fur die Arbeit von morgen Bilanz und Perspektiven Anwendungsorientierter Forschung*, pp 241–251 Heidelberg Springer

Frese, M & Sabini, J (1985) (Eds), *Goal Directed Behavior The Concept of Action in Psychology* Hillsdale, NJ Erlbaum

Greif, S & Janikowski, A (1987) Aktives Lernen durch systematische Fehlerexploration oder programmiertes Lernen durch Tutorials? *Zeitschrift fur Arbeits- und Organisationspsychologie*, 31, 94–99

Hacker, W (1973) *Allgemeine Arbeits—und Ingenieurspsychologie* Berlin VEB Verlag

Hacker, W (1986) *Arbeitspsychologie* Bern Huber

Herbrich, M , Frese, M & Prumper, J (1991) Beobachterubereinstimmung bei der Analyse von Benutzerfehlern Manuscript, University of Munich

Irmer, C , Pfeffer, S & Frese, M (1991) Konsequenzen von Fehleranalysen fur das Training Das Fehlertraining In M Frese & D Zapf (Eds), *Fehler bei der Arbeit mit dem Computer Ergebnisse von Beobachtungen und Befragungen im Burobereich*, pp 158–174 Bern Huber

Johansson, G & Aronsson, G (1984) Stress reactions in computerized administrative work *Journal of Occupational Behaviour*, 5, 159–181

Lenman, S & Robert, J -M (1992) What commands should be undoable and what should the undo granularity be? In H Luczak, A E Cakir & G Cakir (Eds), *Proceedings of the Third International Scientific Conference on Work with Display Units*, pp E 7–E 8 Berlin Papyrus Druck

Miller, G A , Galanter, E & Pribram, K H (1960) *Plans and the Structure of Behavior* London Holt

Neisser, U (1976) *Cognition and Reality* San Francisco, CA Freeman

Norman, D A (1981) Categorization of action slips *Psychological Review*, 88, 1–15

Norman, D A (1986) Cognitive engineering In D A Norman & S W Draper (Eds), *User Centered System Design*, pp 31–61 Hillsdale, NJ Erlbaum

Paul, H J (1992) Explorative acting in interactive systems In M Mattila & W Karwowsky (Eds), *Computer Application in Ergonomics Occupational Safety and Health*, pp 89–96 Amsterdam Elsevier North Holland

Paul, H J & Foks, T (1991) *Exploratives Agieren in Interaktiven Systemen Project Report EXPLORE* (IAT PT-02) (Available from Institut fuer Arbeit und Technik, Floranstr 9, 4650, Gelsenkirchen, Germany)

Perrow, C (1984) *Normal Accidents Living With High-risk Technologies* New York Basic Books

Prümper J (1991) Die Inter-Rater-Reliabilitat von Fehlerbeobachtungen im Feld In M Frese & D Zapf (Eds), *Fehler bei der Arbeit mit dem Computer Ergebnisse von Beobachtungen und Befragungen im Burobereich*, pp 44–57 Bern Huber

Prumper, J , Zapf, D , Brodbeck, F C & Frese, M (1992) Errors of novices and experts Some surprising differences in computerized office work *Behaviour and Information Technology* (in press)

Rasmussen, J (1985) *Human Error Data Facts or Fiction* Roskilde, Denmark Riso National Laboratories

Rasmussen, J (1987) The definition of human error and a taxonomy for technical system design In J Rasmussen, K Duncan & J Leplat (Eds), *New Technology and Human Error* Chichester Wiley

Reason, J (1990) *Human Error* New York Cambridge University Press

Rizzo, A , Bagnara, S & Visciola, M (1987) Human error detection processes *International Journal of Man Machine Studies*, 27, 555–570

Scharer, L I (1983) User training Less is more *Datamation* 29, 175–182

Schonpflug, W (1985) Goal-directed behaviour as a source of stress Psychological origins and consequences of inefficiency In M Frese & J Sabini (Eds) *Goal-directed Behaviour The Concept of Action Theory in Psychology* Hillsdale, NJ Erlbaum

Semmer, N (1984) *Stressbezogene Tatigkeitsanalyse Psychologische Untersuchungen zur Analyze von Stress am Arbeitsplatz* Weinheim Beltz

Shneiderman, B (1987) *Designing the User Interface Strategies for Effective Human–Computer Interaction* Reading, MA Addison–Wesley

Shiffrin R M & Schneider, W (1977) Controlled and automatic human information processing II Perceptual learning, automatic attending, and a general theory *Psychological Review*, 84, 127–190

Smelcer, D B (1989) Understanding user errors in database query Unpublished dissertation thesis University of Michigan

Smith, S L & Mosier, J N (1986) *Guidelines for Designing User Interface Software* Bedford, MA US Department of Commerce, National Technical Information Service

Volpert, W (1982) The model of the hierarchical-sequential organization of action In W Hacker, W Volpert & M Cranach (Eds), *Cognitive and Motivational Aspects of Action*, pp 35–51 Berlin Deutscher Verlag der Wissenschaften

Volpert, W (1983) *Handlungsstrukturanalyse als Beitrag zur Qualifikationsforschung* Koln Pahl-Rugenstein

Volpert, W , Oesterreich, R , Gablenz-Kolakovic, S , Krogoll, T & Resch, M (1983) *Verfahren zur Ermittlung von Regulationserfordernissen in der Arbeitstatigkeit (VERA)* Koln Verlag TUV-Rheinland

Yang, Y (1987) *Undo Support Models* Report from Scottish HCI Centre, Heriot-Watt University Edinburgh

Zapf, D (1991) Stressbezogene Arbeitsanalyse bei der Arbeit mit unterschiedlichen Burosoftwaresystemen *Zeitschrift fur Arbeits- und Organisationspsychologie* 35, 2–11

Zapf, D , Brodbeck, F C , Frese, M , Peters, H & Prumper, J (1992) Errors in working with office computers A first validation of a taxonomy for observed errors in a field setting *International Journal of Human–Computer Interaction*, 4, 311–339

Zapf, D , Brodbeck, F C & Prumper, J (1989) Handlungsorıentıerte Fehlertaxonomıe ın der Mensch-Computer Interaktıon *Zeıtschrıft fur Arbeıts- und Organısatıonspsychologıe*, 33, 178–187

Zapf, D , Lang, T & Wıttmann, A (1991) Der Fehlerbewaltıgungsprozeß In M Frese & D Zapt (Eds), *Fehler beı der Arbeıt mıt dem Computer Ergebnısse von Beobachtungen und Befragungen ım Burobereıch*, pp 58–80 Bern Huber

Zapf, D , Maıer, G W , Rappensperger, G & Irmer, C Error detectıon, task characterıstıcs and some consequences for software desıgn *Applıed Psychology An Internatıonal Revıew* (ın press)

Zapf, D & Scherubl, K (1991) Praktısche Konsequenzen von Fehleranalysen fur dıe Softwareberatung In M Frese & D Zapf (Eds), *Fehler beı der Arbeıt mıt dem Computer Ergebnısse von Beobachtungen und Befragungen ım Burobereıch*, pp 175–184 Bern Huber

Receıved 25 October 1991, revısed versıon receıved 6 November 1992