Using a taxonomy of errors as a conceptual framework for differences in patterns of use for casual and novice users

Jocelyn R. Harper
University of Wollongong, jocelyn@uow.edu.au

Peter Caputi
University of Wollongong, pcaputi@uow.edu.au

Rohan Jayasuriya
University of Wollongong, ajayasur@uow.edu.au

Shae-Leigh C. Vella
University of Wollongong, vella@uow.edu.au

P. Hyland
University of Wollongong, phyland@uow.edu.au

Using a taxonomy of errors as a conceptual framework for differences in patterns of use for casual and novice users

Abstract

A taxonomy of errors was applied in a recent study of casual and novice users of a statistical analysis software. The taxonomy was found to be useful and several extensions to the taxonomy were proposed. The aim of this study is to confirm the theoretical validity of the proposed extensions and the usefulness of the taxonomy in describing the patterns of human-computer interaction and predicting changes in use patterns with learning.

Disciplines

Arts and Humanities | Life Sciences | Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Harper, J. R., Caputi, P., Jayasuriya, R., Hyland, P. N. & Vella, S. (2004). Using a taxonomy of errors as a conceptual framework for differences in patterns of use for casual and novice users. Australian Computer Human Interaction Conference (pp. 1-5). Online http://www.ozchi.org/proceedings/2004/index.html: OZCHI.

Using a taxonomy of errors as a conceptual framework for differences in patterns of use for casual and novice users.

Jocelyn Harper[a], Peter Caputi[a], Rohan Jayasuriya[b], Shae-Leigh Vella[a] and Peter Hyland[c].
Department of Psychology[a], Graduate School of Public Health[b], School of IT & Computer Science[c].
University of Wollongong. NSW, Australia

Abstract

A taxonomy of errors was applied in a recent study of casual and novice users of a statistical analysis software. The taxonomy was found to be useful and several extensions to the taxonomy were proposed. The aim of this study is to confirm the theoretical validity of the proposed extensions and the usefulness of the taxonomy in describing the patterns of human-computer interaction and predicting changes in use patterns with learning.

Keywords
Human errors, Human-computer interaction, Usability, Taxonomy of errors.

INTRODUCTION

Humans make errors because their tendency is for flexibility and creativity, and the ability to link meaning and concepts from inadequate data and knowledge. These skills in interpretation and extrapolation produce errors (Norman, 1990). For some individuals, environmental cues take precedence over their internal goals regardless of the person's will, perhaps accounting for some individual differences in errors (Larson, & Perry, 1999). Investigations of errors and accidents from health, aviation and engineering industries have sought to reduce the frequency and intensity of mishaps and disasters (Norman, 1990; Mulcahy & Rosenthal, 1999). Comparisons of errors between domains are difficult because of specific theoretical perspectives and methodologies such as those of Norman (1983), Reason (1990) and Rasmussen, Duncan, & Leplat (1987).

Understanding human error is critical in many domains. A number of taxonomies of human error have been developed to increase that understanding. Developing such an understanding is important not only because it is of interest to know why humans make errors, but because of the practical implications of humans making errors in the real world. Human-computer interaction errors can have economic, productivity and safety outcomes and therefore investigating their occurrence and prevention was seen as being important (Brodbeck, Zapf, Prumper & Frese, 1993). A taxonomy proposed by Zapf, Brodbeck, Frese, Peters and Prumper (1992) offers an approach to human-computer interaction errors that is centred on usability testing, rather than functionality.

Errors occur in response to intentional action (Reason, 1990). Failure to achieve this results in 1) failure for the action to follow through as intended (from well-rehearsed schema) which is called a slip and 2) the failure for the intended action to achieve the intended outcomes which is called mistake. Slips and mistakes stem from two separate cognitive mechanisms in human processing systems (Norman, 1983). Slips occur when an intention causes a number of schemata to become active but errors occur due to classification, faulty activation patterns or triggering. Schemas are automatic processes developed through increasing expertise, and slips are failures in dealing appropriately with these nodes of expertise. In comparison mistakes occur when the conscious application of rules or knowledge are incorrectly used (Norman, 1983; Leape, 1999).

Usability errors describe human action and come from a theory of human action regulation (Zapf et al, 1992). The human action regulation theory describes a nested pyramid of goals and subgoals that control the sequence of human actions and plans (Zapf et al, 1992; Brodbeck et al, 1993). The theory has three elements for the definition of an error: 1) there must be a goal; 2) not achieving a goal constitutes an error, and 3) an error should be avoidable (Zapf et al, 1992). In addition, work efficiency is a higher order goal for a worker or organisation, and any additional actions superfluous to the required task constitute an "action detour" and can be defined as error (Brodbeck et al, 1993). This theory posits a knowledge base for regulation and a three-tiered action regulation level. Two orthogonal dimensions underlie these taxa: levels of action regulation (representing increasingly more automaticity), and the steps in the action process. Knowledge base for regulation deals with the stored facts and understandings used to develop specific goals and plans (Brodbeck et al, 1993). The knowledge base (SKB) for regulation refers to the user knowledge of commands, keys and rules used in the system, so when incomplete, errors occur because the user fails to successfully apply commands or concepts. This encompasses knowledge of facts, procedures, and understandings and has no equivalent in Rasmussen's taxonomy (Zapf et al, 1992). The tiered action regulation level refers to three action levels which are: 1) intellectual level of regulation (IR); 2) flexible action patterns (FAP) and 3) sensorimotor regulation (SM). Each level of human action represents different levels of conscious activity. Intellectual level of regulation (IR) is defined as "conscious development and activation of goals and plans … for working on complex tasks that have not often been performed" (Brodbeck et al, 1993, p. 306) and is the equivalent of Rasmussen's knowledge-base level (Zapf et al, 1992). An example of an error at this level of regulation would be incorrectly defining a width of columns for a spreadsheet to fit the page-width. Three subtypes of errors occur in this level: Thought error: plans poorly developed although user knows system features; Memory error: part of plan is forgotten; Judgement error: user cannot interpret computer feedback. Flexible action patterns (FAP) are "well practiced actions regulated by plans stored in memory which have to be adjusted to particular situations" (Brodbeck et al, 1993, p. 306). Similar to schemata by Norman (1983) and rule-based level from Rasmussen they require less attention, especially for well-routinized plans, actions and evaluating outcomes (Zapf et al, 1992). An example of an error at this level of regulation is when a user continues to use keystrokes that pertain to the previously used and well-known software. Three subtypes of errors occur at this level: Habit error: Well known sub-plans are not executed or feedback is ignored; Omission error: user does not execute well-known sub-plan (eg due to interruption); Recognition error: user fails to notice / recognise well-known message. Sensorimotor level of action (SM) describes "highly routinized actions that are regulated without conscious attention" (Brodbeck et al, 1993, p. 306). According to Zapf et al, (1992) this is equivalent to Rasmussen's skill-based level. Keyboard typographical and mouse errors are an example of this level of error.

Zapf et al's (1992) taxonomy offers a comprehensive and parsimonious description of categories of errors that occur in IS use (Hyland & Brown, 2004). Because it allows comparisons between different computer systems, it was used and extended in a recent study by Hyland and Brown (2004) and provided greater usability especially with novice and casual users. They found it allowed an effective system of classification of the observed errors and a determination of their frequency. However, the study found that three additional errors should be added to the taxonomy: Logic Errors (where the user is required to use complex logic), Transparent Feedback Errors (where the user fails to interpret new but transparent feedback) and Domain Knowledge Errors (DKB: where the user does not know a concept or rule used in the domain). These errors occur at Intellectual Regulation, Flexible action pattern, and Knowledge base levels respectively. However, Hyland and Brown (2004) questioned whether these new error types were generalizable, had sufficient utility and were consistent with the underlying theories.

The Study Aims

The study sought to validate the utilization of the taxonomy by Zapf et al's (1992) and extension proposed by Hyland and Brown (2004), especially as it relates to novice use of a computer system. This taxonomy was used as a conceptual framework as part of a broader investigation of individual differences in error behaviour and prediction in the human-computer interaction. This paper looks primarily at the decisions made by novice users in their use of a statistical analysis package (SPSS) in a laboratory setting. In this context the taxonomy could present a framework which explored users learning methodologies in their interaction to a IS program which was novel to them.

Method

Participants were 77second year psychology students, drawn from a larger course sample of students whose mean age of was 22.60 years, range 18 to 48 years. They were given course credit points for participation. Their mean computer experience was 2.63, (expertise category range from 1- 4) and 16-22 hours of SPSS tuition.

Participants were interviewed individually. Each participant completed a series of questionnaires including items regarding their computer experience and demographic information. At the completion of these measures, a specified SPSS task was undertaken by each participant while their performance was captured by a screen capture camera (LotusScreenCam). This task included data entry, the definition and labelling of variables and performance of a statistical procedure, namely a one-sample t-test analysis.

Results

A total of 881 errors were recorded in this study. Error descriptions and their designated categories were generated by matching descriptions of errors by Zapf et al and Hyland and Brown with errors that occurred for these participants in their pursuit of the completion of the task. Comparisons of inter-rater judgements were satisfactory (84% agreement). Of the total number of errors 626 (71.06%) errors were SKB errors (in comparison to 9.5% in Zapf et al, 1992; 27.4% in Hyland & Brown, 2004) reflecting the novice level of user of the recent studies. Next most frequent with 157 (17.82 %) were SM errors (16.6%: Zapf et al, 1992; 10.0%: Hyland & Brown, 2004). These categories were easy to classify, in keeping with Hyland and Brown's (2004) findings. None of the users demonstrated normal FAP errors. This finding was expected because no participants were expert SPSS users, in comparison the 21.9% found in the Zapf et al study.

However, there were ten errors (1.14%) in which the correct t-test analysis was performed with the correct outcome, which were not recognised as such and re-ran. They could be categorized as IR judgement errors, where the user cannot interpret computer feedback (17.0%: Zapf et al, 1992) or as FAP errors transparent feedback errors, in which any user should be able to understand the system feedback, and hence recognise it as the correct analysis (Hyland & Brown, 2004). Whether this was truly IR error or a transparent feedback form of the FAP error could be debated, the movements captured on the screen camera in this instance do not provide evidence for the user's reasoning.

Fifty-two (5.9%) errors were difficult to classify as lack of experience with SPSS, and were thus determined to be DKB errors, this rate was half that of Hyland & Brown results (11.0%). Other errors, difficult to distinguish, were 24 (2.72%) errors that stemmed from either a system or domain lack of knowledge (S/DKB). In this error, the user selected the incorrect analysis, but withdrew from continuing the analysis, and cancelled the procedure before obtaining an output. That is, these errors occurred because the user did not have knowledge of the system (either a command, key or rule) or of the domain (a concept or rule) for this task. Again it was very difficult to determine the motives for the withdrawal from the task, and whether the recognition of the decision path being incorrect stemmed from visual cues of the procedure's sub-actions and system graphical interface, or domain knowledge of the decision paths to be expected or met.

Furthermore, 12 (1.36%) errors (so classed because they were unproductive in the rapid production of the desired output) were also difficult to classify in terms of the taxonomy. These were those termed as "inefficient behaviour" errors by Zapf et al, 1992). However, these errors appeared purposeful, and the users were often persistent in their pursuit of additional utilities. Such errors resulted in a more descriptive output, streamlined data or the use of the system 'Help' for verification of the analysis and the sub-actions. Other users manipulated the data and output windows to view both simultaneously, so as to assess (eyeball) the output results in terms of the data. Other users ran two forms of the analysis, and reviewed both. The low percentage of these errors however, does not reflect the amount of time that users spent on these activities; often participants expanded and spent time in these decision paths. These building knowledge base (BKB) errors were not found in the previous studies.

Discussion

The results of this study demonstrated that the inclusion of novice users increased the percentage of knowledge base errors and decreased the percentage of FAP errors in comparison to Zapf et al's study. This pattern of results was also found by Hyland and Brown (2004). Most of the errors found in this study could be classified according to the Zapf et al – Hyland and Brown taxonomy. Eighty nine percent were easily classified into the Zapf et al taxonomy, while the majority of the remaining 9.76% of errors could be located in Hyland and Brown's extensions of the taxonomy. However, some forms of errors (1.36 %) could not be classified according to either taxonomy, and were seen from a different perspective. Although these 'errors' were not immediately recognised as such, they were classed as errors according to the taxonomy and when the criterion is productivity and time constraints are viewed as critical. These 'inefficient behaviour errors' or 'action detours' such as running two forms of the analysis may be seen as errors because it occurred as a follow up action from the correct analysis, and thus constitutes an unnecessary action. However, this may be seen as a knowledge building response, either re-running a duplicate of the correct analysis or a version of the correct analysis, may be a confirmatory and validating response. Although no one ran the correct analysis, re-ran a similar but incorrect analysis and then deleted the incorrect one, some participants left the two analyses. Other responses that were not strictly required were more detailed investigation of variable options. The use of the Help menu, and manipulation of data and output windows shows a sophisticated use of the system clearly assists both visual cues for recognition of the correct analysis i.e. a system knowledge base, and also the domain knowledge base. This expanded effort appeared to involve deliberation and higher cognitive functioning from some participants. This is an error according to the definition of taxonomy, and during systems' use because they delay the work process, but during knowledge building, are investigations that later may yield shortcuts and non-direct expertise. In addition, in the re-conceptualisation of the work process, and ownership of the learning process and building an integrated understanding of the system's architecture, these errors may prove to be very valuable.

The study realised its aim. It validated the utilization and extension of the taxonomy, especially as it relates to novice use of a computer system. In addition, two further error categories were found. The taxonomy also allowed for a conceptual framework for error behaviour and prediction to be utilised as suggested by both Birdi et al, (1997) and Hyland and Brown (2004). The taxonomy therefore again demonstrated its generalisability. Novice use in computer interactions increased the frequency of knowledge based errors. The form of these errors could stem from a lack of knowledge in either the software or statistics, or a mixture of both, as seen in the results of this study. This result is in agreement with Brodbeck et al (1992) who found increased handling time occurred for Intellectual Regulation and Knowledge Base errors, and with Prumper et al (1992, as cited in Birdi et al, 1997) who found that error handling time for novices was greater than for experts. However, the positive nature of these forms of errors, that is, recognition that knowledge investigation derives benefits as well as costs, were not recorded in any of the previous studies. The study also had a pragmatic basis to review the praxis of user interactions with SPSS. Tutorial classes are seen to provide theoretical and experiential knowledge base from which student users are expected to gain sufficient SPSS experience along with laboratory access for the course requirements. Some participants made very few errors, and obtained the results with minimum effort and concern, while others supplied embellishments, and others focused on detail and accuracy rather than speed. This finding implied that there were individual differences in how participants approached a standardised simple SPSS statistical analysis task.

Future research could use this taxonomy to determine individual differences in learning approaches to a computer task, especially of novice response sets. It could be a useful means to explore how error production is integrated in the iterative process and praxis when building a knowledge base for system software. This study would be strengthened by a post task interview with the user to clarify whether these errors are system or domain knowledge base. These findings could assist in training purposes, to clarify whether domain or system requires additional conceptual or experiential understanding. These errors occurred less frequently over the course of the study, which supports that these are due to lack of expertise. Using this taxonomy as a means of differentiating the cause of an error could assist trainers and users to separate domain from system knowledge weaknesses in learning contexts.

REFERENCES

Birdi, K., Pennington, J, and Zapf, D. (1997) Aging and errors in computer-based work: An observational field study. Journal of Occupational and Organisational Psychology, 70, 35-47.

Brodbeck, F.C., Zapf, D., Prumper, J. and Frese, M. (1993). Error handling in office work: A field setting. Journal of Occupational and Organisational Psychology, 66, 303-317.

Hyland, P. and Brown, R. (2004) Applying a taxonomy of error in usability testing. Paper submitted to OZCHI2004.

Larson, G.E.& Perry, Z.A. (1999) Visual capture and human error. Applied Cognitive Psychology, 13(3) 227-236.

Leape, L. (1999). Error in medicine. Medical Mishaps: Pieces of the Puzzle. Chapter 2. Open University Press. Buckingham. Great Britain.

Mulcahy and Rosenthal, (1999). Beyond blaming and perfection. Medical Mishaps: Pieces of the Puzzle. Chapter 1. Open University Press. Buckingham. Great Britain.

Norman, D. (1983). Design rules based on analyses of human error. Communications of the ACM, 26, 254-258.

Norman, D. (1990). Human error and the design of computer systems. Editorial. Communications of the ACM, 33, 4-7.

Rasmussen, J., Duncan, K. and Leplat, J. (1987). New Technology and Human Error. John Wiley & Sons. Great Britain.

Reason, J. (1990). Human Error. New York, NY: Cambridge University Press.

Zapf, D., Brodbeck, F.C., Frese, M., Peters, H., and 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 (4) 311-339.

ACKNOWLEDGEMENTS

This article is based on an ongoing project supported by an Australian Research Council Grant from the Australian Federal Government. The chief investigators are Dr Rohan Jayasuriya and Peter Caputi.

COPYRIGHT