Recognising Erroneous and Exploratory Interactions

Jonathan Back[1], Ann Blandford[1], and Paul Curzon[2]

[1] University College London Interaction Centre
[2] Queen Mary, University of London
j.back@ucl.ac.uk, a.blandford@ucl.ac.uk, pc@dcs.qmul.ac.uk

Abstract. A better understanding of "human error" is needed to help overcome problems of people assuming they are to blame for their inability to use poorly designed technology. In order to investigate people's ability to recognize, and reflect on the causes of, particular types of errors, a problem solving environment was designed that allowed participants to verbally self-report erroneous and exploratory interactions. It was found that the pervasiveness of errors was recognizable but underlying cognitive and attentional causes of errors were not. Participants found that providing a causal account of device-specific errors during interaction was especially difficult. A striking feature of device-specific errors is that they involve actions that do not move an individual towards a goal state, but remain critical to performing a task correctly. Successfully identifying why an error has occurred requires an understanding of environmental cues and salience. Findings imply that HCI practitioners need to develop techniques to adjust the visual salience of cues, making it is possible to recognize and recover from error.

Keywords: Human error, self-report, HCI.

1 Introduction

It is now recognized that many errors in routine interactive behaviour are not the product of some stochastic process, and that causal explanations of human error can be developed [1]. However, little is known about what factors influence an individual's ability to recognize errors. Recognition that an error has been made is a prerequisite for error recovery. The focus of this paper is on this recognition process rather than the error recovery process as a whole.

Errors are sensitive to external influences. For example, forgetting to collect the original document after making photocopies is more likely if an individual is thinking about 'other things' or is interrupted. Although forgetting your original document can be inconvenient, research has shown that similar underlying causal factors can result in catastrophic social consequences (e.g., aircraft crashes and nuclear power station failures) [2]. Much of the previous work on understanding human error has relied on participants generating retrospective self-reports, often some time after the event [3], or accounts of particular incidents – for example, accident investigation reports [4]. These forms of data collection have provided a high-level understanding of error, but lack the information about timing and context needed to develop a more detailed account of error phenomena and their cognitive causes.

C. Baranauskas et al. (Eds.): INTERACT 2007, LNCS 4663, Part II, pp. 127 – 140, 2007.

In an attempt to develop a cognitive account some work [1, 5, 6] has been based on behavioural traces, recording user activity and classifying actions as correct or erroneous according to experimenter-defined criteria. In the work reported here, we have investigated the use of real-time self-reports to further investigate human error, based on participants' own definitions of errors.

1.1 Background

One of the first attempts at demonstrating the non-stochastic nature of errors was suggested by Rasmussen and Jensen [7]. The idea that errors can be categorized as being skill-based, rule-based, or knowledge-based allows errors to be attributed to different cognitive factors. However, whether an error is classified as skill-based, rule-based, or knowledge-based may depend more on the level of analysis rather than on its ontogeny [8]. For example Gray [5] argued that the same behavior, e.g. "taking the wrong route during rush hour", can result from lack of knowledge (not knowing about a faster route) or misapplication of a rule (knowing that one route is the fastest during rush hour and the other is fastest on the off hours but applying the 'off hours' rule). In addition, this behavior could be caused by a slip (taking the more familiar route when the intention was to take the less familiar but faster one) or be intentionally wrong (cannot get into the correct lane due to traffic).

Investigating a situation where an error has occurred outside the 'laboratory environment' requires an individual to provide a self-report, or requires the use of an error analysis framework such as CREAM [9]. Such frameworks focus on the probabilities of error types occurring; 'fine grained' explanations are unlikely to be elicited. The use of self-report when investigating human error has traditionally been post hoc and incidental. Questionnaire studies can yield interesting data with regard to individual differences in error proneness, the relatedness of various error types, and the organization of the underlying control mechanisms [2]. Responses to questionnaire items asking about the incidence of a wide variety of minor cognitive failures remain consistent over several months [3]. Broadbent, Cooper, Fitzgerald and Parks [10] showed that stress-vulnerability (or a certain style of cognitive management) is associated with cognitive failure in everyday life.

Recently there is a move towards understanding how the task environment influences working memory, attention and other cognitive functions [5]. Questionnaires and interviews cannot be used to reveal failures in specific cognitive functions that may have caused an error. Moreover, the cognitive context in which error types occur, in relation to a specific cognitive activity, cannot be established. When investigating problem solving, one method of self-report that has been found useful is the collection of concurrent verbal protocols (a commentary provided by the participant throughout an experimental trial). The question is: can this type of reporting be used successfully in understanding the factors that influence the recognition of error?

1.2 Concurrent Verbal Protocols

Previous research on verbal protocols has argued that it is essential to ensure that self-report procedures do not interfere with task behavior. If it is found that procedures do interfere with task behavior then this invalidates the approach. Generally, concurrent verbalization has been shown not to change the nature or sequence of thought processes if two conditions are adhered to: a) participant should only be instructed to verbalize thoughts that are already the focus of attention (participants should not elaborate on past events) [11]; b) training in the use of the think-aloud procedure ensures that valid representations of participants' thought processes are obtained since causal explanations should be provided in 'real-time' [12]. It seems unlikely that these conditions can be adhered to when investigating human error.

The nature of recognizing a slip error (non-knowledge based error) always requires an individual's focus of attention to be shifted to the environment. Noticing that the system is in the wrong state requires a 'salient' signal from the device or a user will continue to attempt to execute task goals. For example when setting a wake-up alarm on a digital clock, a user is often required to switch to the alarm set mode before inputting the wake-up time. If the alarm set mode is not selected then inputting the wake-up time can reset the current time. Alarm setting is a trivial example of a mode error; however, mode errors can sometimes have catastrophic social consequences. They can cause automation surprises - a phenomenon which can trigger operator confusion about the status of an automatic flight control system [2].

Training individuals to report errors as they occur is impossible since cognitive limitations often prevent instant error recognition. For example, some errors are caused by a loss of activation - when the presumed mechanism associated with the 'activation' of a goal has decayed from working memory [13]. Omission errors (forgetting to do something) are often indicative of these limitations which may delay or prevent recognition, prohibiting 'real-time' causal explanations. Thus, the use of concurrent verbal protocols where participants should not elaborate on past events and always provide ‘real-time’ explanations is inappropriate for studying human error.

A self-report mechanism is needed that is able to represent the way in which individuals recognize they have made an error without interfering with thought processes. Error recognition may be initiated by unexpected changes in the device state. For example, slip errors can be recognized if device feedback alerts the user of a mismatch between intentions and performance. However, if feedback from a device is not 'salient enough' to be noticed, then an individual might remain unaware of an erroneous interaction. There are situations where it is not easy to recognize an error since many problem solving strategies are likely to be automated unconscious elements of cognition [14]. However in some situations, reviewing previous interactions may force individuals to assess the suitability of strategies and thereby facilitate the recognition of errors. What is explicit is likely to be those aspects of an interaction that are not routine. Capture errors, for example, occur when there is an overlap in the sequence required for the performance of two different actions, especially if one is done more frequently than the other [15]. Reviewing actions may enable the identification of these types of incidents. Accidentally switching to the 'automatic routine' for a different goal that begins with the same set of actions is a common slip (e.g., taking the familiar route home when the intention was to take the less familiar but faster one).

1.3 Proposed Mechanisms for the Self-report of Error

Errors are one measure of the quality of human performance. For example, Miller [16] identified an important property of working memory by discovering that individuals make errors when recalling more than 7 (+/-2) elements of information. However, the everyday concept of error presupposes a goal. This can make the classification of errors difficult if an individual is interacting in an exploratory way to satisfy a learning goal, especially when a user is adopting a trial-and-error approach. A full understanding of human error is only possible if there is a way of differentiating between errors and exploratory interactions (where errors or sub-optimal moves can be an expected or even a desired outcome). Humans are not always able to describe their goals or able to recognize the extent to which a goal has been addressed.

Research has shown that exploratory interactions are used extensively during problem solving [11]. Exploratory actions are typically attempts to address a learning goal. Problematically, situations are likely to arise where learning goals cannot or will not be explicitly described, making self-reporting difficult. For example, the goal could be integrated within an automated schema; an individual may not be aware that they are acting in an exploratory way; the goal could be recognizable but the additional cognitive effort required to report the goal might push levels of cognitive load beyond available mental resources; an individual might be unwilling to report it if they believe that the goal is trivial; or they simply might not want to expend additional effort that reduces the efficiency of the problem solving process.

When an individual is able and willing to self-report an exploratory interaction, an explanation requires the lack of knowledge to be coherently described. If the task environment explicitly indicates that such knowledge is required, then an individual can simply report the feature or object that needs to be better understood. For example if an individual encounters an unknown feature or object that they believe might be pertinent to a task, they might decide to perform an exploratory interaction. Once the interaction is performed, it might then be possible to provide a self-report about whether they were able to discover an unknown property. Likewise it should be possible for an individual to self-report instances where an interaction is easily recognized as the wrong approach for the situation, prompting subsequent exploratory interactions. In this case, an individual can provide a self-report by justifying an exploratory interaction on the basis that a prior approach was erroneous.

For this work it was decided that when providing a self-report a participant should be able to differentiate exploratory interactions from erroneous ones. It was hoped that by identifying situations where differentiating errors from exploratory moves was difficult, and comparing them to situations where differentiating was easy, this would elucidate factors associated with the self-recognition of error. As previously discussed, the use of a concurrent verbal protocol is inappropriate. Instead, a participant must be given the opportunity to report an error when it is discovered. The simplest way of allowing this is to encourage participants to provide an 'Elective Report' at any time during interaction. The time that a participant chooses to make a report may reveal some contextual information that can be used to provide an insight into the process of recognizing errors. There is a strong possibility, however, that some errors may remain undetected due to the level of cognitive load imposed by the problem solving environment or due to an attentional failure. It is for this reason that a 'Debrief Reporting Mechanism' [17], which requires a participant to review a trace of their own behaviour after a task is completed, was also implemented. This type of self-report mechanism prompts an individual to review previous interactions in an attempt to identify those that can now be seen as erroneous. By comparing elective reports with debrief reports, factors that make the recognition of errors during a task possible may be better understood.

One major difference between historical and non-historical judgment is that the historical judge typically knows how things turned out. This may influence the way in which self-reports are provided. The utility of such self-reports may be compromised by hindsight bias. In an attempt to explore the influence of hindsight bias, Fischhoff [18] presented a series of descriptions of clinical or historical events for which four possible outcomes were provided. Participants were asked to rate the likelihood of occurrence. It was found that participants who were told that one of the events was true were more likely to increase its perceived probability of occurrence. Critically, Fischhoff found that participants who had outcome knowledge were largely unaware of the effect that outcome knowledge had on their perceptions. As a result, they overestimated what they would have known without outcome knowledge. It was concluded that this lack of awareness may restrict the ability to learn from the past. Within the medical domain, Flach [19] argued that error elimination strategies rely on hindsight as they involve systematically reducing causes of error so that the system is made increasingly safer. Given outcome knowledge, it has been suggested by Woods et al. [20] that reviewers will tend to simplify the problem solving situation. Woods et al. claimed that outcome knowledge blinds individuals to causal factors by biasing the uncertainties, tradeoffs, and the attentional demands. The very notion of causality may indeed be a symptom of hindsight bias. Flach suggested this is why safety management strategists prefer to focus on constraints rather than causes.

In this work we focus on better understanding the error recognition process. Although comparing elective self-reports with debrief self-reports is not a reliable method for the development of error elimination strategies, it may provide an insight into the factors that make errors unrecognizable during interaction.

2 Method

An adventure game was designed that allowed participants to interact freely within a problem solving environment. This game was developed using the 'Quest Development Tool Kit' (available at http://www.axeuk.com/quest/). The game specified a series of locations (rooms) and placed objects within rooms or within the player's inventory (possessions). The key element of the game involved problem solving. Objects such as a locked door were not designed as permanent obstacles, but merely as problems to be tackled. Solving problems frequently involved finding objects (adding them to the inventory) and then using them in the appropriate way. The game required a high level of attention to be maintained by participants as solutions to similar looking problems were not learnable. An aim of this cognitively demanding environment was to provoke a high error rate.

When commencing a new task, a participant was prompted to click on the ‘objectives button’ (see Figure 1). Pressing the button activated a scrolling ticker. The ticker was only capable of displaying one line of text (approximately 14 words) at any time. Task objectives were always six lines long. Ticker speed was set at 0.25 seconds per word. It took approximately 20 seconds for the ticker to display all of the task objectives, after which it was deactivated. Instructions presented to participants at the beginning suggested that they need not attempt to remember the task objectives as they can press the ‘objectives button’ at any time during the task. The ‘objectives button’ and associated ticker perform two important functions. Firstly, both the level of cognitive load imposed on an individual (remembering objectives) and amount of attentional load (monitoring the status of the ticker) is high. The status of the ticker is likely to require attention as it will not always be displaying required information. Deciding to perform another interaction while the ticker is scrolling requires a participant to predict how long they have until they have to shift their attention back to the ticker. This increased load on the individual was designed to provoke high rates of erroneous behaviour. Another reason for encouraging the regular use of the 'objectives' button is to draw attention to that particular interface panel. Maintaining the salience of this panel is important as the interface features used for the elicitation of self-reports were located there.

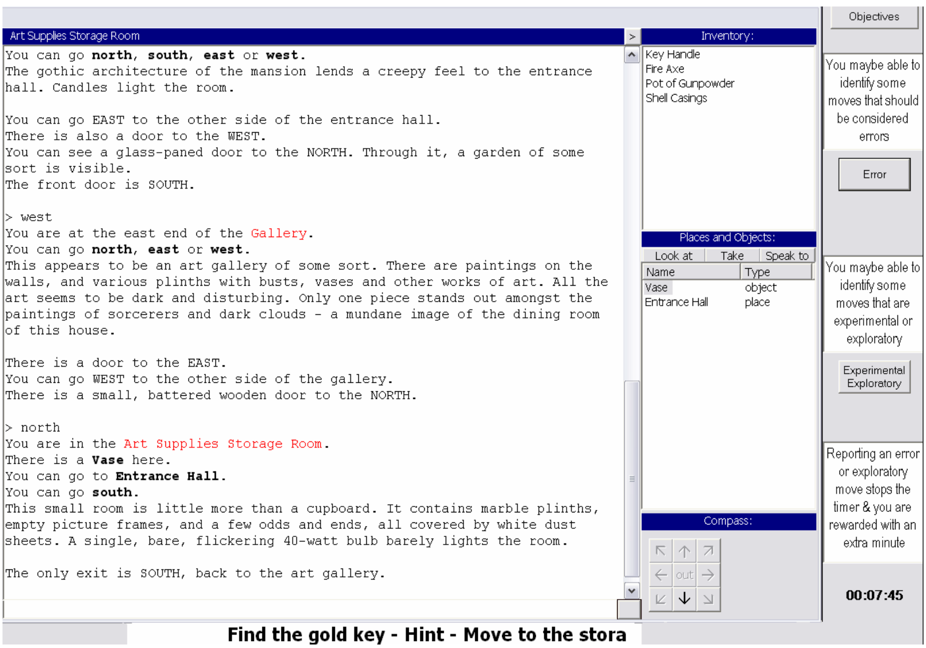

Fig. 1. The user interface, including reporting buttons (right) and ticker tape display (bottom)

Navigating around the game environment required the use of a paper-based map. Before commencing experimentation, a participant was shown their starting position relative to other rooms and floors in the house. Participants were free to refer to the map at any time. An example of the type of problem a participant was required to solve is presented below:

Task Aim: Dissolve concrete to get a golden key.
Task Objectives: Find concrete block and some acid, then make use of the sink.
Hints: Concrete block is hidden in a bedroom, you will need to find a key in the store room. Acid is hidden in dining room. Sink is in the workshop.

Possible Solution: Move to store room: the only object is a vase. Examine vase to discover that something rattles inside. Drop vase to break it and take key 001. Move to bedroom 1 (since key is labeled 001); use key to unlock door. Find concrete block in ensuite bathroom. Move to dining room. Examine drinks cabinet to activate pop-up dialog "Do you want to open cabinet doors?". Click yes and look at drinks cabinet (since doors are now open). Take bottle and examine bottle (i.e. open it) to identify that it contains acid. Move to workshop. Examine sink to activate pop-up dialog "Do you want to plug sink?". Click yes and use bottle on sink. Use concrete on sink, concrete dissolves to reveal golden key. Examine sink to activate pop-up dialog "Do you want to unplug sink?". Click yes and take golden key.

2.1 Procedure

Twenty participants were recruited for experimentation. Participants were pooled from: research staff at University College London Interaction Centre; research staff at the Department of Computer Science at Queen Mary, University of London (CSQMUL); MSci students at CSQMUL. Each participant completed three tasks in total: one training task and two test tasks. The objective of the training task was to allow participants to familiarise themselves with the interface. Participants were allowed to refer to instructions that explained interface features during training. After the training task was successfully completed these instructions were removed. The training trial did not provide opportunities to self-report.

All tasks were isomorphic, in that the same number of problem solving moves were needed to solve task objectives. Although participants were set task objectives, the nature of the game ensured participants adopted an exploratory approach. Task objectives were presented on the ticker display, as described above, but executing these objectives required exploratory iterations. Four objectives were set for each task. In order to complete an objective, at least one set of actions had to be performed that were not explicitly specified in the instructions. This requirement could only be discovered during interaction. A correct way of solving problems was not specified. Participants were able to solve problems in any order, although all problems had to be solved to complete a task.

After the training task was completed, counterbalancing was performed as shown in Table 1. Participants performed either the Elective or Debrief self-report for each task. The 'Elective Report' required participants to click either the 'error button' or the 'exploratory button' at any time during problem solving (*see* Figure 1). When either button was clicked, the participant was asked to give a brief oral description of the error or exploratory move. The 'Debrief Report' required participants to review the moves they had performed in order to identify any errors and exploratory moves. After a task was completed, a trace of the moves they had made was presented on screen.

Table 1. Allocation of participants to conditions

Task Order	Report Mechanism Order	
	Elective then Debrief	Debrief then Elective
Task 1 then Task 2	S1...S5	S11…S15
Task 2 then Task 1	S6…S10	S16…S20

Unlike the training trial, the task trials were time limited. Participants were given fifteen minutes to complete each task trial. In a pilot study, four participants took an average of nine minutes thirteen seconds to complete each task trial. The imposed time limit was designed to further increase levels of cognitive load in an attempt to increase the number of errors. Participants were informed that when they provided a self-report the clock would stop and they would be given an extra minute as an incentive to self-report. They were informed that providing a self-report that did not relate to an error or exploratory move was unacceptable. It was decided that a participant should determine themselves whether a particular move should be described as an error or exploratory (or not at all) and, if so, how it should be described. It was anticipated that analysing different error reporting styles might provide an insight into the type of error or exploratory move. Before undertaking task trials participants were presented with a second set of paper-based instructions. Instructions included an illustration of where the self-report buttons were located on the interface, and included a guideline that clarified the difference between reporting erroneous and exploratory moves. Although a guideline was provided, participants were informed that these were only suggestions. Participants were encouraged to develop their own distinctions between what should be considered erroneous or exploratory.

For the debrief report, participants were asked to use the paper based map to trace back through the navigational decisions they made. Decisions to manipulate objects could be reviewed by identifying the associated 'action label' in the move description panel (*see* Figure 1). Participants did not have access to task objectives. It was hoped that this would encourage individuals to cognitively reconstruct the original context in which they performed an action. If an erroneous or exploratory move could be identified then participants were required to click the 'Error Button' or the 'Exploratory Button' and provide a brief verbal description. If a participant was unable to recall the context of their actions, they were prompted by the experimenter. The experimenter identified the relevant task objectives that were the focus of activities and reminded the participant. If this was not possible or the participant was still unable to understand the context, the experimenter moved the reviewing processes forward through to a point in the trace that a participant was able to identify with.

3 Results and Analysis

The overall aim was to discover whether self-reports provide useful information about the recognition of error. By performing a quantitative analysis (*see* Section 3.2) on the categories that emerged from a qualitative analysis (*see* Section 3.1), the following questions were addressed: What types of errors and exploratory moves can be self-reported? When comparing elective reports to debrief reports; can factors that allow or prevent the recognition of error during interaction be identified?

3.1 Qualitative Analysis

From the 262 reports elicited 27 (10%) provided insufficient information to allow a useful insight into an individual's thought processes because they did not comprehensibly describe the nature of the error or exploratory move; these reports are omitted. 48 (18%). A qualitative analysis identified two main self-report categories: 1) reports about goal-specific actions; 2) reports about device-specific actions. Goal-specific actions move an individual towards a goal. These types of actions can be considered to be 'salient' since they are always associated with performing recognized task goals. In contrast, device-specific actions do not move an individual towards a goal state though still remain critical to performing a task correctly. The requirement to perform device-specific actions often varies from device to device (e.g., some ATMs do not require users to press enter after entering a PIN). Examples of device-specific actions include: initializing a device before using it (e.g., on a Web page form: clicking on a text entry field before entering text); switching modes so that inputs are interpreted correctly (e.g., switching to non-predictive text mode on a mobile/cellular phone); and performing an additional step after completing a task that cannot be done prior to task completion (e.g., collecting the original document after making photocopies). Both the two main categories were assigned two sub-categories: a) learning; b) attending and remembering. Reports about learning were focused on the development of problem solving strategies. Reports on attending and remembering were about executing intentions.

Learning goal-specific actions: 48 reports (18%) – 29 Erroneous, 19 Exploratory. Clicking on the task objectives button activated the ticker. Each objective required a participant to engage in problem solving. It became apparent to participants that these objectives could not be addressed without exploration of the game environment. Participants discovered that in order to address task objectives, new aspects associated with task objectives had to be learnt. These aspects were not explicitly specified by the ticker. It was found that participants were able to describe learning requirements in terms of the relationships between game objects and locations. For example when participants recognised that achieving Objective 1 using Object 1 was impossible without finding Object 2 first, they were willing to report this inferred requirement. Learning took place either by undertaking exploratory interactions, or recognising why an interaction was erroneous and thereby reducing the problem state space i.e. by adopting a trial and error approach.

Remembering and attending to goal-specific actions: 39 reports (15%) – 28 Erroneous, 11 Exploratory. Participants reported situations where they thought that they might have forgotten to perform a required objective – for example, forgetting to retrieve Object 1 from Location 1 before going to Location 2. Participants also provided self-reports that suggested that a lack of attention was to blame for forgetting – for example, when participants knew what they should be doing but were somehow caught up in a different activity.

Learning device-specific actions: 68 reports (26%) – 26 Erroneous, 42 Exploratory. Participants were able to identify situations involving a mismatch between a reportable intention and perceived performance. This mismatch was attributed to a lack of knowledge associated with how to interact with objects when a participant believed they knew what they had to do – for example, when a participant knew they had to play a video tape but did not know how to operate the VCR. Participants used exploratory interactions and drew conclusions from errors in order to develop interaction strategies. Participants were willing to disclose when these strategies, learnt from previously successful interactions, did not work.

Remembering and attending to device-specific actions: 32 reports (13%) – 24 Erroneous, 8 Exploratory. Participants reported situations where they believed that they had forgotten to perform a required interaction – for example, forgetting to examine Object 1 (to change the object state) before using it on Object 2. Reports also suggested that errors or exploratory behaviour could have been avoided if the participant had paid more attention to feedback from the system. For example, when using the paper-based map, participants 'overshot' locations when navigating around the game due to a failure in attending to feedback about where they were.

Critically, the qualitative analysis also looked at the proportion of reports that were considered to be 'reasoned'. If the objective of the interaction was clearly stated, and the description identified one or more causal factors, then a report was classified as being reasoned (*see* Table 2). Example of a reasoned report (from Participant 7, Timed Task A, Self-report 5): ‘After finding where a key is hidden I must remember to take the key, add it to my inventory, and not leave it in the room. Now that I have recognized this error hopefully I will not make it again’. Example of an unreasoned report (from Participant 11, Timed Task B, Self-report 6): ‘I seem to have missed out a crucial step, maybe I'm not using the right command, but I'm not sure what happened’.

3.2 Quantitative Analysis

The aim of the quantitative analysis was to determine whether differences between the elective reports and the debrief reports existed. The numbers of reports in each category for each reporting mechanism are shown in Table 2. This table also shows the proportion of each report type that was classed as reasoned (expressed as a percentage).

Results that suggest elective reports are better than the debrief reports: Significantly more exploratory reports were made using 'elective report' than 'debrief report' (Wilcoxon signed-rank test, $Z = 2.160$, $p<0.05$). This suggests that participants were better able to recognize exploratory moves during interaction. Furthermore, elective reporting enabled a significantly greater proportion of reasoned exploratory reports to be made (Wilcoxon signed-rank test, $Z = 2.040$, $p<0.05$).

Results that suggest the debrief report is better than the elective report: The number of error reports made were evenly distributed between the elective and debrief mechanisms (Wilcoxon signed-rank test, $Z = 0.360$, $p>0.05$). However, a trend towards more reasoned error reports during debrief is observable. On closer inspection it is clear that this trend is due to the difficulty participants have when reasoning about device-specific errors during interaction. The proportion of reasoned device-specific self-reports associated with the debrief mechanism is significantly higher than the elective mechanism (Wilcoxon signed-rank test, $Z = 1.940$, $p<0.05$).

Table 2. Self-report categories and frequencies

	Number of **elective** reports	Percentage of reasoned **elective** reports	Number of **debrief** reports	Percentage of reasoned **debrief** reports
Exploratory (all)	55	51%	25	31%
Erroneous (all)	53	36%	54	68%
Goal-specific errors	28	50%	29	66%
Device-specific errors	25	20%	25	72%

4 General Discussion

It was found that reporting a wide range of both erroneous and exploratory interactions was possible. A qualitative analysis revealed two main report categories: 'goal-specific' and 'device-specific'. Two sub-categories: 'learning' and 'remembering / attending' were also identified. In order to provide a 'reasoned' self-report an individual must attribute their erroneous or exploratory behavior to causal factors. Identifying these factors is dependent on how well an individual can elucidate the context in which the erroneous or exploratory move occurred. A cognitive context defines the set of all possible interactions that an individual is capable of performing at that moment. The context in which an individual makes an error is likely to be different from the context in which an individual self-reports. Therefore during interaction, an individual must be able to reconstruct the series of actions that led to an error or exploratory move. This was not an easy process: only 51% of exploratory and 36% of erroneous reports provided a causal explanation.

When comparing the elective mechanism with the debrief mechanism no significant differences were associated with the frequency of erroneous reports. However, exploratory interactions were significantly more frequently reported using the elective self-report mechanism. Woods et al. [20] argued that self-reports can be biased by hindsight which prevents them from being a useful tool for understanding interaction. Our analyses showed that the elective mechanism was able to elicit a significantly wider range of exploratory move types than the debrief mechanism. This supports the notion that outcome knowledge (knowing how things turned out) biases self-reporting processes, especially when reporting exploratory moves.

Comparing the qualities of reports elicited using both elective and debrief mechanisms provided an insight into how individuals detect erroneous interactions. Interestingly analysis revealed that during interaction, the pervasiveness of device-specific errors was recognizable but underlying cognitive and attentional causes were not. The frequency of reporting these error types was identical: 50 elective reports, 50 debrief reports. However, only 20% of elective error reports associated with device-specific actions were reasoned accounts of error. During debrief reporting, participants were more able to provide a reasoned account of device-specific error. 72% of these reports were reasoned.

Based on these findings we argue that the error recognition process is dependent on cognitive context and the availability of environmental cues. Reporting device-specific errors during interaction is harder than when performing a debrief report because different environmental cues are 'salient'. During the debriefing session participants were required to debug their task performance. Critically, participants were not reminded of task objectives. Therefore, the only way of detecting erroneous moves was to recall intentions based on the availability of environmental cues. When performing a debrief report immediately after interaction, participants were able to remember intentions and were actively looking for environmental cues that could be used to execute those intentions. In the following example, a participant remembered that they had the intention to open a door but then made a device-specific error. An example of a debrief report (from Participant 16, Timed Task A, Self-report 2) is: 'Now I can see why I tried to use the wrong key to open the door (intention). I forgot that I had to examine the key first to check that it was the right key (device-specific error). When it did not work, I wrongly assumed that the door could not be opened using any key (erroneous intention formulated)'. For this error to be reportable a participant must recognize that they formulated an erroneous intention. This is not easily recognized during interaction since attention is allocated to achieving task goals and not on the retrospective analysis of previous actions that were unsuccessful.

During interaction, participants were better able to report causal accounts of goal-specific errors (50%) compared to when reporting device-specific errors (20%) (statistically significantly better). When engaged in problem solving participants reported switching between goals. In contrast to device-specific actions, goal-specific actions are salient. When occupied by multiple goal-specific activities it is easy to forget to make an attentional check. This type of error can be reported during interaction. Example (from Participant 10, Timed Task A, Self-report 5): ‘I didn’t realize that the Shells were already in my inventory. I wasn’t really thinking about that objective at the time. I realize that I have made an error in not checking my inventory, I did not need to go and find any more Shells.’

In summary, participants found it easier to reason about goal-specific errors during interaction when compared to device-specific errors. When activities are seen to be essential to enabling progress towards a goal state, errors associated with these activities can be better explained. This is because they are more salient than activities that are required to perform an intention but do not themselves enable progress towards a goal (*see* [21] for an extended discussion on salience).

Participants used for experimentation were HCI experts. They are likely to be practiced in evaluating their own behaviour. Therefore, their ability to recognize and reason about error is representative of the 'best case scenario'. Novice users are likely to find it even harder to discover causes of error. A better understanding of "human error" is needed to help overcome problems of people assuming they are to blame for their inability to use poorly designed technology.

5 Conclusion

While "human error" may be the immediate and direct cause of failure, other factors such as system design are instrumental in facilitating or provoking error. Laboratory research suggests device-specific actions are the hardest type of actions to perform correctly [5]. Work reported in this paper has shown that although these types of errors are recognizable, they are the hardest types of actions to provide reasoned reports about during interaction. An inability to provide reasoned reports suggests that recovering from these types of errors is unlikely since their causes cannot be easily identified. Unfortunately, designing systems where device-specific actions are not required may not always be possible. Not all device-specific actions can be made goal-specific. For example: some devices need to be initialized before running a process; some devices require modes of operation; some require a post-completion step. Therefore, error recognition and recovery needs to be better supported by system designers. During debrief, participants were able to identify device-specific errors because they remembered previously formulated intentions and found environmental cues that allowed these intentions to be executed. During interaction these cues are not salient because participants are allocating their attention to the task, and do not develop retrospective accounts of actions that do not move them towards a goal state. Therefore, designers should attempt to modify visual salience. For example: start buttons should be 'grayed out' making them less visually salient if initialization steps (before pressing "start") may be required; modes should have visually salient indicators; post-completion steps should incorporate just-in-time cues [21]. Although the cognitive salience of an action for a particular individual may not always be captured by visual salience, when an error is recognized then the visual salience of these cues may help individuals to develop a more reasoned retrospective causal account.

Acknowledgments. This work is funded by EPSRC grants GR/S67494 and GR/S67500. We are grateful to all participants in our studies.

References

1. Byrne, M., Bovair, S.: A Working Memory Model of a Common Procedural Error. Cognitive Science 21(1), 31–69 (1997)
2. Reason, J.T.: Human Error. Cambridge University Press, Cambridge, UK (1990)
3. Reason, J.T., Lucas, D.A.: Using cognitive diaries to investigate naturally occurring memory blocks. In: Harris, J., Morris, P. (eds.) Everyday Memory, Actions and Absent-Mindedness, Academic Press, London (1984)
4. Johnson, C.W.: Representing the Impact of Time on Human Error and Systems Failure. Interacting with Computers 11, 53–86 (1998)
5. Gray, W.D.: The nature and processing of errors in interactive behavior. Cognitive Science 24(2), 205–248 (2000)
6. Back, J., Cheng, W.L., Dann, R., Curzon, P., Blandford, A.: Does being motivated to avoid procedural errors influence their systematicity? People and Computers XX - Engage. In: Proceedings of HCI 2006, London, UK, vol. 1 (2006)
7. Rasmussen, J., Jensen, A.: Mental procedures in real-life tasks: A case of electronic troubleshooting. Ergonomics 17, 293–307 (1974)

8. Hollnagel, E., Mancini, G., Woods, D.: Cognitive Engineering in Complex Dynamic Worlds. Academic Press, London (1988)
9. Hollnagel, E.: Cognitive reliability and error analysis method. Elsevier, Oxford, UK (1998)
10. Broadbent, D., Cooper, P., FitzGerald, P., Parkes, K.: The Cognitive Failures Questionnaire (CFQ) and its correlates. British Journal of Clinical Psychology 21, 1–16 (1982)
11. Ericsson, K.A., Simon, H.A.: Protocol analysis: Verbal reports as data (rev. ed.). MIT Press, Cambridge, MA (1993)
12. Perkins, D.: The mind's best work. Harvard University Press, Cambridge, MA (1981)
13. Altmann, E., Trafton, J.: Memory for goals: an activation-based model. Cognitive Science 26(1), 39–83 (2002)
14. Mandler, G.: Consciousness: Respectable, useful, and probably necessary. In: Solso, R. (ed.) Information Processing and Cognition, Erlbaum, NJ (1975)
15. Norman, D.: Categorization of action slips. Psychological Review 88(1), 1–15 (1981)
16. Miller, G.: Human memory and the storage of information. IRE Transactions on Information Theory, IT-2(3), 128–137 (1956)
17. Hoc, J., Leplat, J.: Evaluation of different modalities of verbalization in a sorting task. International Journal of Man-Machine Studies 18, 283–306 (1983)
18. Fischhoff, B.: Hindsight ≠ foresight: the effect of outcome knowledge on judgement under uncertainty. Journal of Experimental Psychology: Human Perception and Performance 1, 288–299 (1975)
19. Flach, J.M.: For those condemned to live in the future. Quality and Safety in Health Care 12, 304–312 (2003)
20. Woods, D. D., Johannesen, L. J., Cook, R. I., Sarter, N. B.: Behind human error: Cognitive systems, computers, complexity and hindsight. CSERIAC, 94–101 (1994)
21. Blandford, A., Back, J., Curzon, P., Li, S., Ruksenas, R.: Reasoning about human error by modeling cognition and interaction. In: Proc. Resilience Engineering Symposium (2006)