Task Analysis for Error Identification: Theory, method and validation

Chris Baber and Neville A. Stanton*

Kodak/Royal Academy Educational
Technology Research Group,
School of Electronic & Electrical Engineering,
The University of Birmingham.
Birmingham.
B15 2TT

* Department of Design
Brunel University,
Runymede Campus,
Egham,
Surrey.
TW20 0JZ

Abstract

In this paper, the underlying theory of Task Analysis for Error Identification is presented. The aim is to illustrate the development of a method that has been proposed for the evaluation of prototypical designs from the perspective of predicting human error. The paper presents the method applied to representative examples. The methodology is considered in terms of the various validation studies that we have conducted, and is discussed in the light of a specific case study.

INTRODUCTION

Task Analysis for Error Identification (TAFEI) has been in development since our first presentation in 1991 (Baber and Stanton, 1991). In this early work, our concern was with the development of a method that could describe a form of dialogue between users and products with a view to predicting likely types of human error arising from these dialogues. The dialogues were described in terms of state-space diagrams, and the technique has been applied to simple products, such as kettles (Baber and Stanton, 1994), domestic electrical products such as audio-cassette players (Baber and Stanton, 1994) and video-cassette recorders (Baber and Stanton, 1992, 1994), medical informatics (Baber and Stanton, 1999), and industrial applications, such as electricity sub-stations (Glendon and McKenna, 1995) road-cleansing vehicles (Stanton and Baber, 1993), together with numerous examples of public technology (Baber and Stanton, 1992; Stanton and Baber, 1996).

Our first detailed description of the method (Baber and Stanton, 1994), proposed that the method is based on two assumptions: that "…the problems that people have with products arises from their purposeful use of them…" [p.1923] and "…that the planning of actions, by the user, will be situated at specific points in the developing interaction between human and product." [p. 1924]. In support of these assumptions, it was proposed that human-product interaction could be considered as a form of "cooperative endeavour" (Lewis and Norman, 1986). By this we mean that a dialogue between user and product is conducted in order to achieve a specific goal, which requires the user and product need to cooperate, share information and assist each other.

In broad terms, we follow the lead of Card et al. (1986) who propose that human interaction with products represents a form of problem solving which progresses according to Newell and Simon's (1972) notion of a problem-space. During the course of a dialogue between human and product, each state offers the user potential for action and, if the user selects an appropriate action, then the dialogue progresses to the next relevant state. We have been interested in asking how can one know that the user will select the most appropriate action (in terms of their goal and the product's current state), and can one predict whether the user will select an inappropriate action?

By way of example, assume that a person is about to purchase a can from a drinks vending machine. The machine has a big label telling the person to use the correct change. The person inserts a coin that is of greater value than the can they wish to buy and the machine dispenses the can but does not give any change. While there is a large ergonomics literature concerned with the design of instructions, warnings and other forms of verbal labeling (see Wogalter et al. 1999 for coverage of recent work in this area), it is not uncommon for people to misread, ignore or otherwise not notice such instructions. Indeed, Zwaga (1988) observed that the average time to read instructions adjacent to a ticket-vending machine on the Washington Metro was 34 seconds; way too short a time to digest and comprehend instructions beyond the question of 'what do I do first?' Our concern has been with the development of a methodology to demonstrate what might happen if people fail to respond to instructions appropriately, and then to develop a theory to account for such behaviour.

METHODOLOGICAL DEVELOPMENT:
TASK ANALYSIS FOR ERROR IDENTIFICATION

The procedure for the Task Analysis for Error Identification (TAFEI) is shown in figure one. The basic stages are to produce a description of the user's interaction with the product in terms of a state-space diagram. We have opted to use a very simple representation of state-space diagrams, really providing little more than a storyboard of the states that one passes through in order to achieve a specific goal. We accept that other, more formal, techniques, could offer greater precision but do not feel that this is necessary. What is important is that the sequence of states represented ought to be necessary and sufficient to achieving a specific goal. This means that one might develop a series of such diagrams if one wanted to examine a range of interactions with the product, e.g., as a form of scenario analysis. The main reasons for specifying a user goal are to avoid the combinatorial explosion associated with state-based descriptions of dialogue, i.e., to eliminate the problem of attempting to capture every single state in a product's use, and to force the focus on the user's purposeful interaction with the product. By asking what goal the user is trying to achieve, we can focus our attention on the activities that are essential to that goal, and can consider possible deviations from these activities. Also, by focusing on specific goals, we can better propose likely Global Prototypical Routines and State-Specific Routines to explain possible causes of human error.

[FIGURE ONE ABOUT HERE]

In the initial version of TAFEI our intention was to define possible classes of human error that were possible during human interaction with products. This led to a methodology that could comfortably deal with 'slips' but not always with 'mistakes'. The methodology assumes that goal-directed human-product interaction progresses through a series of 'legal', i.e., permissible, states. In each state there are opportunities for the user to deviate from the legal path. Such actions are termed 'illegal' and our initial aim is to determine possible illegal transitions. We then seek to develop possible explanations as to why the illegal transitions might be possible and why people might perform them.

Defining User Goal and Task Sequences

We tend to use Hierarchical Task Analysis (HTA) as the primary means of describing user activity (see Shepherd, 2000). This is because the technique offers us a means of describing user actions in terms of Goals and the Tasks required to reach these Goals. Furthermore, HTA also offers comprehensive descriptions of task sequences through its use of Plans. However, it is not essential that TAFEI employs a full HTA, but it is important that the analyst is able to define a user's goals and propose the required task sequence. The proforma in figure two provides an indication of how this might be performed.

[FIGURE TWO ABOUT HERE]

The indication of a 'Source' for information in the analysis can be useful when presenting to clients, designers or other parties; if the analysis is performed solely using the handbook it might yield a different task sequence to an analysis performed using a selection of experienced users. The ideal state, of course, is to conduct as comprehensive an analysis of user activity as is feasible within the scope of the project.

Developing State-Space Diagrams

State-space diagrams can most easily be constructed by following the instructions laid out in a handbook for the product, or (even better) through speaking to the design team and asking them to describe the sequence of states that the product will pass through during a specific course of action. Thus, for the 'change greeting' example, the designers from ACME might be able to provide a flow-chart or sequence diagram to indicate how the product performs this task sequence.

Begin with a listing of states that through which the products passes when the user is pursuing a defined goal. Thus, as a very simple example, assume the goal is 'change the time on an analogue wristwatch'. The initial state is the wristwatch displaying the current time, with the hands moving on the face. In order to change the time, the user needs to pull out the bezel. This action changes the state of the wristwatch in two ways: the hands stop moving, and the hands can be moved. The user then turns the bezel until the hands show the desired time. Finally, the bezel is pushed back in to return the watch to its original state. The states of the wristwatch for the goal of change time are: State0: display current time, hands moving; State1: hands not moving; State2: hands moving in response to bezel rotation, and the sequence of states for this goal are: state0, state1, state2, state0.

Having defined the sequence of states that are relevant to achieving the specific goal, the next step is to consider the subsequent states that the product can be ‘waiting for’. In other words, we propose that for each state, a product can move to a one or more subsequent states. We say that in a given state, the product is ‘Waiting for…’ an action to take them to another state. The action might be performed by the user or by the product, but that is not important at this stage. Returning to the wristwatch example, in State0 the watch is ‘waiting to have bezel pulled out one stop’ (to change the date), or ‘waiting to have bezel pulled out two stops’ (to change the time), or ‘waiting to show time’ (in which case the user does nothing). For each state, a separate item is created; figure three shows state0 of our wristwatch.

[FIGURE THREE ABOUT HERE]

Notice that the item has an identification number. It is important to note that the number *only* represents to position of that state in the sequence related to a *specific* goal; in other words, it is possible for the same state item to appear in different diagrams with different numbers. The point of the state number is to support the creation of the goal-specific transition matrix (see below). The state item also has a title for the state. We have found, that where possible, it is useful to also provide a thumb-nail sketch of the system image at that state, although this is not essential. The use of thumb-nail sketches makes TAFEI looks a little like storyboards and also focuses attention on specific features of the system image in that state. Below the state item's title, we have a set of "Waiting for…" statements, which are the possible states that a user is able to move from that state.

Developing TAFEI: mapping task sequences onto state-space diagrams

Developing the TAFEI diagram involves labeling the transitions between states in terms of user or product actions. The 'Waiting for…' statements are then connected to the state item of the subsequent state used a labeled arrow; the label on the arrow describes the user or machine action that is required to make that transition. Thus, referring back to figure three, the transition from State0 to State1 would require the user tasks of 'grasp bezel – pull bezel one stop – release bezel', which, in turn, could be represented by a notation from the HTA diagram, e.g., Plan1. In this instance, the transition from state0 to state1 would be labeled Plan1 (or P1).

Pairing user task with state transitions assumes that there is a clear synergy between the task analysis and state-space descriptions, but our experience is that the process often involves modification of both descriptions. In broad terms, the process of constructing a TAFEI diagram requires the analyst to ensure that they have produced a logical sequence of tasks in terms of the sequence of product states. In this manner, the analysis is concerned with producing a description of user-product dialogue. For a complete TAFEI, each state can be annotated using 'By Products'. A By Product is a result of the current state, e.g., a boiling kettle has By Products including steam, a VCR programmed to record a particular programme has By Products including locking the user out of further interaction with the product, and, in the wristwatch example, pulling out the bezel has a by product of stopping the hands moving.

As mentioned previously, the TAFEI diagram represents a form of story-board that relates the states of a product to user actions during the pursuit of a Goal. Thus, it can be considered analogous to cognitive walkthroughs, storyboards or the Plans of HTA. However, it is proposed that the visual nature of TAFEI, together with the combination of user and product descriptions, offers a clear account of the possible dialogue.

Developing TAFEI: constructing error matrices

Having produced a TAFEI diagram, the next stage is to consider the possibility of illegal transitions. In this instance, we consult the TAFEI diagram and ask whether it is possible to move between the states it contains. This means asking not only if one can move between state0 and state1 but also if one can move between state1 and state3 or state4 etc. The question is not whether one would want to move between these states, but whether it is possible. Here, the assumption is that if anything can go wrong, it will (Murphys Law).

We have found it beneficial to present the transitions between states in a matrix. This offers a means of ensuring that the analysis has been exhaustive (within the states required to achieve a given Goal), and gives a visual indication of the amount of possible problems in the product. Figure four shows the transition matrix for our wristwatch example.

[FIGURE FOUR ABOUT HERE]

Each cell in figure four represents a possible between states, e.g., From state0 To state1. In this stage, the analyst asks if there is anything to prevent such a transition (note: we are not asking whether such a transition is sensible or desirable only whether it is possible). Thus, at state0 it is possible to remain at state0 (by doing nothing) or move to state1 (by pulling the bezel out one stop) or move to state2 (by pulling the bezel out two stops). If the goal is to adjust the time, then only one of these transitions is desirable (or 'Legal'), i.e., from state0 to state2. This means that any other transition will not lead to towards the goal and will, hence, by 'Illegal'. The letters in the cells indicate Legal or Illegal transitions. In some instances, a transition will be impossible and this is indicated by -.

By defining Illegal transitions, we demonstrate the possibility of human error. Part of the assumptions underlying the method is that designers should then seek to eliminate Illegal transitions between states, or (where it is not possible to eliminate Illegal transitions) to indicate clearly the transition that user ought to be making. While this may sound plausible, it does highlight a significant discrepancy between TAFEI and many products on the market. TAFEI assumes that interactions are goal-based, state-transitions and that a person's interaction with a product can be improved by indicating which actions ought to be performed in order to achieve a defined goal. However, this implies that there must be some identification (in the dialogue) of the user's goal. Consumer products typically are designed to support as many transitions between states as possible (presumably on the basis that the designers did not want people to become stuck in the wrong mode, and that designers wanted to provide people with as many opportunities for action as possible). This means that, by definition almost, consumer products will yield a high proportion of Illegal transitions. It is a moot point (and beyond the scope of this paper) as to whether it is more desirable to have products which support smooth, error-free transitions between states to achieve a goal or whether the products should support flexible operation (which, in effect, means allowing multiple transitions at each state). We would argue for the former and claim that one reason for the prevalence of the latter view is that consumer products are not designed according to detailed and well-developed user models.

VALIDATING TAFEI

A significant concern in our work has been to validate the TAFEI method; after all, there seems little point in employing a method that does not pass even the basic requirements of validity and reliability (Stanton and Young, 1998, 1999a, b). Assume that a method can be considered in terms of: Face Validity (does the method appear to be dealing with the problem that it is designed to handle?); Construct Validity (does the method have an underlying theory that makes sense?); Reliability (will the method be used consistently be different practitioners?); Predictive validity (can the method predict problems that users will actually encounter?).

In terms of Face and Construct Validity, it is not easy to collect data concerning these factors. Given that the technique is grounded in a psychological theory of how people use products, we claim that it has strong construct validity. We have presented the technique to other practitioners at Workshops and Conferences, as well as through our teaching. We have a growing collection of projects in which TAFEI has been employed and which the authors of these studies feel that the method has been useful. All of this evidence is presented in support of strong face validity.

Our main efforts have been toward establishing the Reliability and Predictive Validity of TAFEI. We have sought to consider reliability from three main perspectives: comparative reliability relates to the performance of TAFEI in comparison with other methods, inter-rater reliability relates to the performance of analysts using TAFEI, and predictive validity relates to the performance of TAFEI relative to actual observations. We feel that it is important for a method that claims to predict human error can be tested against actual errors that have been observed, and so the latter measure (predictive validity) is central to our assessment.

Two principal studies have been conducted to compare TAFEI's predictions with observations of actual performance. The first study looked at the application of TAFEI by experts, i.e., the techniques developers, while the second study looked at the use of the method by novices. In the first study, TAFEI was used to predict types of human error that might be encountered in the use of a ticket-vending machine (Baber and Stanton, 1996). These predictions were compared against a set of human errors that had been observed during 48 hours of over 300 observations of people using the ticket-vending machines on underground railway stations (an account of the observation study can be found in Baber and Parker, 1994). In broad terms, TAFEI performed well; predicting 10 out of the 15 types of error that had been observed, i.e., around 67% accuracy. This figure, produced by an analysis conducted by the authors, compares favourably with the accuracy of 68.6% claimed by Holnagel (1993) in the validation of CREAM. Thus, it would seem that, while TAFEI does not predict 100% of errors, it produces an acceptably high match.

We have also compared TAFEI performance with other human error prediction techniques, namely SHERPA and heuristic evaluation. Baber and Stanton (1996) found that SHERPA yielded 12 of the 15 error types, i.e., 80%. Stanton and Stevenage (1998) found that SHERPA had an accuracy of 75%. While SHERPA performs very well, it also generates a high number of 'false alarms' [a mean of 15.4 false alarms], i.e., it predicts errors that do not occur.

A significant issue is how well a method performs in the hands of people other than its developer. Stanton and Baber (2001) report a study in which novice users are trained to use TAFEI, SHERPA and heuristic evaluation. In this study, novice users were trained in one of the three methods. Heuristic evaluation was included to represent a 'no method' condition, in that participants were given a lecture of human error (together with Reason's (1990) GEMS model) and allowed to apply this as they saw fit. The study compares observed errors with predicted errors in a task involving the use of a vending machine. The mean accuracy of the performance of novice users is SHERPA: 68%; TAFEI: 48%; Heuristic: 27%. In broad terms, the novices performed at around 20% lower than experts. However, it is interesting to note that the heuristic evaluation was the least accurate. The study showed that with three sessions for the methods, performance improved. For example, sensitivity metrics for TAFEI increased from 0.73 in trial one to 0.79 in trial three.

In terms of inter-rater reliability, Baber and Stanton (1996) report a correlation of around r = 0.9 between two expert users for TAFEI and SHERPA. In Stanton and Baber (submitted) the novices show correlations of 0.79 for TAFEI and 0.73 for SHERPA.

The work discussed in this section is still, of course, in progress. However, it is proposed that ergonomics methods do not tend to seek to measure reliability (Stanton and Young, 1999). This means that any attempt to validate a method represents a step forward for ergonomics. We also feel that, despite the apparently low levels of accuracy for novice users, TAFEI still represents a viable method for evaluating products and predicting human error (and is demonstrably superior to simply using heuristic evaluation).

THEORETICAL DEVELOPMENTS:

REWRITABLE ROUTINES

Assuming that a human-product dialogue progresses through a series of states as a cooperative endeavour allows us to ask how the user and product move between states. We assume that the product's "system image" (Norman, 1988; Dix et al., 1992) offers the user clues and hints as to which actions should be performed in a given state, and that the user has a set of "mental models" that guide actions and support interpretation of the system image.

The system image should clearly and unambiguously inform the user only of those actions that make sense in a given sequence of tasks. This means that we often take issue with the plethora of personal and public technologies that have sought to present the user with all possible options at each state; not only could this bombard the user with confusing information, and increase the time that a novice would take to select an option, but also increases the possibility of the person making a mistake in their interaction with the product. Ideally, products ought to have a small set of options at each state and a clearly identified task-flow between states. By way of example, Baber and Parker (1994) evaluated the ticket-vending machines used by London Underground Limited using TAFEI and field-studies. On the basis of the data collected, it was proposed that a revised TAFEI could be created, i.e., in which one limited options at each state and clarified the task-flow. This resulted in a change of design from a button-based to a touch-screen based ticket-vending machine. The touch-screen based design was then developed by Westinghouse-Cubic and has been introduced throughout London

Underground. TAFEI was a significant component in developing the touch-screen system, and was used to model system interaction, i.e., between the ticket purchaser and the vending machine, prior to building the device.

In addition to system image, we assume that users employ some form of "mental model" (although we are not comfortable with this term and prefer 'rewritable routines'). We propose that users of products make use of "…highly fragmented knowledge of the product, [based] on a variety of metaphors and…heavily influenced by the system image". (Baber and Stanton, 2001). Such knowledge is assumed to take the form of primarily routines for action, i.e., they function to guide the user's actions. These routines are analogous to the notion of Scripts in cognitive psychology (Schank and Abelson, 1977). Users can make use of "Global Prototypical Routines" (Baber and Stanton, 2001) that represent stereotypical responses to system images that a person has learned, acquired or otherwise developed. Examples of Global Prototypical Routines can be found in the literature on stimulus-response compatibility, e.g., people exhibit a strong stereotyped response to turn a tap (faucet) anti-clockwise to turn it on or to increase water-flow (Sanders and McCormick, 1992). It is not important whether such responses are *always* correct, but it is important that people will tend to try such responses before other actions. Indeed, work by Rosenbaum (1992) suggests that the decisions to perform a particular action are made even before the person makes contact with an object, e.g., Rosenbaum (1992) asked participants to reach for a control lever that they needed to move in a given direction, and found that the orientation of the wrist indicated the direction of movement (in turns of optimizing torque) prior to gripping the lever. One can replicate these observations by simply stopping oneself prior to gripping a door handle and observing the position of one's hand, i.e., does the hand represent an optimal orientation to grasp the door handle or does it represent an orientation that will result in a comfortable position once the handle is turned? Thus, we propose that people employ Global Prototypical Routines, i.e., a repertoire of stereotypical responses that allow them to perform repetitive and mundane activities with little or no conscious effort. It is only when the action breaks down that one becomes aware of the selection of response.

In addition to Global Prototypical Routines we propose that people also have State-Specific Routines. "Interpretation of the system image in terms of the current goal state might draw on knowledge related to other products, i.e., through analogy or metaphor, in order to infer an appropriate action." (Baber and Stanton, 2001). In other words, these routines are heavily dependent upon the information and clues available through the system image. An example of how state-specific routines might function is in the use of mobile telephone handsets: in order to switch on some models of mobile telephone, the user presses a button that shows an image of a handset lying flat, and to make a call they press a button that shows an image of a raised handset. Faced with these images, a novice user could easily opt to select the raised handset image to turn on the telephone, on the assumption that one lifts the handset on a 'normal' telephone. What is important for our theory, is that these State-Specific Routines are developed and employed only for a specific state, and that once the user has moved beyond that state, these routines are no longer required. Hence, we assume that such routines are rewritable.

By rewritable we mean transitory and easily over-written. The underlying notion here is of the articulatory loop in working memory (Baddeley, 1986). The articulatory loop has a limited capacity, perhaps equivalent to 2 to 4 seconds of speech, and can be disrupted by competing information. While we sympathise with the GOMS notion (Card et al., 1986) that a set of procedures is 'unpacked' through a limited working memory (in the case of GOMS, a pop-stack), we feel that the notion of a complete set of procedures in place prior to the interaction overly idealises human-product interaction. It is not, in our opinion, common for a user of a product to have formulated a complete and precise procedure for performing a specific activity, and for a dialogue to be the simple unpacking of this procedure. Rather interaction progresses through a series of jolting starts and stops, with opportunistic (situated) decision-making and problem-solving leading to dead-ends, restarts and other forms of error (cf. Suchman, 1988). Of course, with very little practice we become proficient a performing a particular procedural sequences, but this might represent the compiling of routines into a larger routine. Performing a different activity, or performing the activity on a different machine or coming back after a lengthy absence, might return us to the hesitant and error-prone performance of the novice. In a study of Automated-teller Machine (ATM) use, Burford and Baber (1994) observed variation in the time that people spent performing simple operations, and the types of error made. One common form of error arose from incorrect insertion of a card into the ATM, and this could be related to the users having experience of machines that required alternative orientations. In this example, 'card_insertion' could be said to be covered by a State-Specific Routine that stated, for instance, "insert card with magnetic strip up and to the right". Given the presence of such a routine, it was not necessary for the users to either read instructions or pay much attention to the process of card insertion; they *knew* what to do, and only needed to revise their routine in the face of failure.

To recapitulate so far, it is our contention that human interaction with products progresses through a sequence of states, and that people interpret each state according to State-Specific Routines and relate product state to their goals through Global Prototypical Routines. This means that analyzing human-product interaction in terms of sequences of states ought to allow us to ask questions about erroneous progression between states. We also assume that we are all novices when it comes to public and personal technologies; few of us read instructions, we seldom receive any training and our performance is often the result of over-learned routines that are sufficient to achieve a small set of goals. In other words, very few of us are capable of fully exploiting the huge range of functions offered by the products that we most often use.

Figure five presents a schematic of how we propose that the concept of Rewritable Routines can be applied. The current state of a product is reflected through its System Image (1). The System Image could evoke a set of Global Prototypical Routines (2), e.g., if the System Image includes buttons, then Routines for pressing buttons are evoked. This reflects assumptions from ecological psychology relating to the Affordances of the system image (3). In addition, the interpretation of the System Image leads to the generation of State-Specific Routines (4), which allow the user to determine the likely procedures associated with specific features of the system image. A comparator (5) then relates possible routines to the user Goal (0), prior to the selection of action (6). What is important in this proposed model is that the comparator operates within very constrained limits (i.e., both in terms of capacity and duration). Thus, the routine selected to handle one state will be overwritten as soon as the product moves into a subsequent state (hence, the routines are rewritable).

[INSERT FIGURE FIVE ABOUT HERE]

A SIMPLE EXAMPLE OF REWRITABLE ROUTINES

Imagine a set of Dishwasher controls as shown in figure six. Imagine that you have first encountered these controls in a holiday home, after a particular heavy meal. You have loaded the Dishwasher, added the washing tablet and closed the door. Which wash cycle do you select?

[INSERT FIGURE SIX ABOUT HERE]

You might try to recall the settings on your Dishwasher at home, if you are fortunate enough to own one. Otherwise, a Dishwasher might be a luxury item that you only encounter on holiday or in other people's homes. In this instance, you have no previous experience. In either case, it is probable that you need to spend a little time considering the system image and reformulating your goal: to get the dishes clean, or perhaps, to get this thing working so that I can return to the sofa with a glass of wine.

Given that figure six is not intended to represent a real product, there is not a ‘correct’ answer to the problem. The point is simply to ask what do you do when confronted with an unfamiliar interface?

In our analysis, we would construct a TAFEI diagram of the product, as shown in figure seven.

[FIGURE SEVEN ABOUT HERE]

Figure seven shows the TAFEI diagram for the Goal ‘Load machine and switch (without changing any setting)’. While transitions from state 1 move to new states, it is likely that the system image will remain relatively constant. In other words, the interface that is shown in figure six will continue to be the user’s point of reference, with states {4-8} being small modifications (related to movement of the sliders) of the original interface. Only states 1, 2 and 3 lead to significantly different states, i.e., in 2 the user’s focus will be on loading the machine and on 3 the user will be ‘locked-out’ of machine functioning as the wash cycle begins.

A Transition Matrix arising from figure seven would yield erroneous transitions from state0 to state3, i.e., switching on an empty dishwasher. We might propose that there could be a display to indicate whether the dishwasher is loaded in order to prevent this slip.

Thus, from the TAFEI analysis it is possible to propose various slips that could result in user error (and the user failing to achieve a particular goal). However, suppose that the current setting (as shown in figure six) is for lightly soiled glasses rather than for a full dinner service after a meal, and that should the user follow the TAFEI shown in figure seven, they will have to rewash the contents of the dishwasher. From this assumption, we note the user of our hypothetical dishwasher is able to erroneously press 'Wash' without altering any of the other settings. In order to explain why this might occur, one can propose the following Global Prototypical (GPR) and State-specific (SSR) routines:

i. GPR1: Assume Default – the user assumes that the current settings represents a system default (rather than merely the last wash), and that leaving the controls alone will call on these defaults;

ii. GPR2: Assume Dangerous Consequences – the user assumes that any change to the current settings could cause all manner of problems and that it would be far safer to only activate the 'Wash' button;

iii. GPR3: Assume Lack of Knowledge – the user assumes that, as they do not understand what the controls actually do, it would be sensible not to adjust them;

iv. SSR1: The numbers on the 'Detergent Strength' represent a characteristic of the detergent, i.e., there should be a corresponding number on the side of the packet (as opposed to representing the amount of detergent added to the wash);

v. SSR2: The numbers on the ‘Finish’ scale represent increasing quality of finish to the wash, i.e., with 3 representing high quality finish (as opposed to representing the length of spin).

Each routine represents a modification on the adage ‘if it aint broke don’t fix it’ (although, of course, the user has no idea if the machine is or is not broke). Each routine represents a credible explanation as to why the user does not alter any settings, and each routine raises questions about the system image. In other words, by raising the possibility of user error and providing possible routines, we have effectively conducted an evaluation of the interface in a form that begs redesign. Questions arise from these routines, such as how can the user be informed as to the system default or the need to change settings?, should casual users be allowed access to all settings?, could the controls be more clearly labeled in terms of their function rather than name?

FURTHER DEVELOPMENTS OF TAFEI AND REWRITABLE ROUTINES

Having briefly explained the concept of Rewritable Routines in the previous section, it is clear that our next phase of work is to test the assumptions that underly this theory. One way in which this can be explored is through the use of cognitive walkthrough (or verbal protocol) studies of people interacting with products. Indeed, the report of two people using a photocopying machine in Suchman (1988) can be read in the light of Rewritable

Routines. From such studies, we can refine the theory and consider such points as how many Routines do people seem to run at once? If one considers the problem solving literature, then it is credible that people only hold a very small number (i.e., no more than 2) hypothesis, explanations, routines etc. in mind at any one time (Evans, 1989). We would hazard a guess that a similar limitation is at play when people interact with technology. Furthermore, we would be interested in extending the notion of compatibility as a means of explaining how people interpret and make use of system image.

One development that we have been working on over the past two or three years has been the use of TAFEI to predict user performance. We assume that the principal metrics of user performance are time and error, and have developed TAFEI to offer a means of predicting performance time in terms of possible errors. In the initial model, we assume that erroneous actions can be corrected in one move (by returning to the start state, e.g., by pressing a Cancel button). However, it is a simple matter to develop the model to allow people to move to the previous state, e.g., by pressing an Undo button.

Baber and Stanton (2001) describe the use of a portable CD player in terms of predicting transaction time. We consider the simple task of setting the product to repeat play (which involves several keypresses to be performed in a specific sequence). Assuming that keypressing can be allotted a finite time (e.g., Card et al., Olson and Olson, Baber and Mellor), it is possible to propose that the time to perform the activity is simple a sum of these times. However, on the basis of TAFEI, we assume that each possible state can lead to several other states (only one of which is Legal). Thus, by assigning probabilities of transition to different states, it is possible to modify the model to reflect both time and error. As shown in more detail in Baber and Stanton (2001), we take probabilities from HEART and apply these to reflect 'expert' or 'novice' performance. In other words, we assume that the basic actions, i.e., press a button, will take the same amount of time whatever one's level of expertise. What will influence performance will be the time taken to undo errors.

The use of error probabilities allows us to develop predictive models of user performance that can handle variation in level of expertise; we assume that novice progress more slowly as a result of making more mistakes, i.e., from having incomplete and poorly executed procedures, rather than simply going slowly. Work on skilled performance suggests that experts and novices perform basic actions at the same speed, but that differences in performance arise from experts being able to chunk these actions into coherent procedures.

DISCUSSION

TAFEI represents a theory-based account of why people make mistakes when using personal and public technology, coupled with a method for predicting how and when people would make such mistakes. The work reviewed in this paper has covered a wide range of application domains. Throughout the development of the method we have sought to validate the method, i.e., to demonstrate that the predictions are reliable and useful. At present TAFEI appears to be able to match between 48% - 68% of predicted with observed errors, and we are currently working on the development of procedures that should allow practitioners to function at the higher end of this range. Of course, one might ask whether a method that effectively misses half of possible human error is useful. Our answer to this is two-fold:

i. as mentioned in the Introduction, the prediction of human error is problematic in that one is never certain that one will be able to describe all possible errors that people can make, and, consequently, any technique that can capture some of these errors might be useful;

ii. TAFEI is intended to function as a prototyping tool (following a technique we call analytical prototyping), which means that the purpose of the technique is to provide a vehicle for rapid assessment of designs prior to commencing user trials. This means that the current use of the method is to provide a 'broad brush' assessment of the potential for a user-product dialogue to fail in order to stimulate ideas for redesign or modification.

With these points in mind, we argue that the earlier in the design process that techniques such as TAFEI are pressed into service, the more beneficial they will be.

REFERENCES

Baber, C. and Parker, A. (1994) Buying a ticket on the Underground In Ed. S.A. Robertson *Contemporary Ergonomics 1994* London: Taylor and Francis 46 – 51

Baber, C. and Stanton, N.A. (1991) Task Analysis for Error Identification In Ed. E. J. Lovesey *Contemporary Ergonomics* London: Taylor and Francis 67 – 71

Baber, C. and Stanton, N. A. (1992) Defining "problem spaces" in VCR use: the application of task analysis for error identification In Ed. E.J. Lovesey *Contemporary Ergonomics 1992* London: Taylor and Francis 418 - 423

Baber, C. and Stanton, N.A. (1994) Task analysis for error identification, *Ergonomics 37 (11)* 1923-1942

Baber, C. and Stanton. N.A. (1996) Human error identification techniques applied to public technology: predictions compared with observed use *Applied Ergonomics* 119-131

Baber, C. and Stanton, N.A. (1998) Rewritable routines in human interaction with public technology *International Journal of Cognitive Ergonomics 1 (4)* 337-349

Baber, C. and Stanton, N. (1999) Analytical prototyping, In J.M. Noyes and M. Cook (eds.) *Interface Technology*, Baldock: Research Studies Press. 175-194

Baber, C. and Stanton, N.A. (2001) Analytical prototyping of personal technologies: using predictions of time and error to evaluate user interfaces, In M. Hirose (ed) *Interact '01*, Amsterdam: IOS Press 585-592

Baddeley, A. (1986) *Working Memory*, Oxford: Oxford University Press

Burford, B.C. and Baber, C. (1994) A user-centred evaluation of a simulated adaptive autoteller In Ed. S.A. Robertson *Contemporary Ergonomics 1994* London: Taylor and Francis 64 - 69

Card, S.K., Moran, T.P. and Newell, A. (1983) *The Psychology of Human-Computer Interaction*, Hillsdale, NJ: LEA

Dix, A., Finlay, J., Abowd, G. and Beale, R. (1992) *Human Computer Interaction*, London: Prentice Hall

Embrey, D. E. (1986). SHERPA: A systematic human error reduction and prediction approach. Paper presented at the International Meeting on Advances in Nuclear

Power Systems, Knoxville, Tennessee.Glendon and McKenna (1995)

Evans, J. St.-B.T., 1989, *Bias in Human Reasoning*, Hove: LEA

Hollnagel, E. (1993). Human reliability analysis: Context and control. London: Academic Press.

Lewis, C. and Norman, D.A. (1986) Designing for error, In D.A. Norman and S.W.Draper (eds) *User Centred System Design,* Hillsdale, NJ: LEA, 411-432

Newell, A. & Simon, H. A. (1972) *Human Problem Solving.* Prentice Hall: Englewood Cliffs, NJ.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic Books

Reason, J. (1990). *Human error.* Cambridge: Cambridge University Press.

Rosenbaum, D.A. (1992) *Human Motor Control*, San Diego, CA: Academic Press

Sanders, M.S. and McCormick, E.J. (1992) *Human Factors in Engineering and Design*, New York: McGraw Hill

Schank, R.C. and Abelson, R., 1977, *Scripts,Plans, Goals and Understanding*, Hillsdale, N.J.: LEA

Stanton, N. A. & Baber, C. (1993) Task analysis for error identification. Paper presented to the *joint MRC/RNPRC/APRC Workshop on Task Analysis*, 5-6 April, Scarman House, University of Warwick.

Stanton, N.A. and Baber, C. (1996) A systems approach to human error *identification* Safety Science 22 (1-3) 215-228

Stanton, N.A. and Baber, C. (1998) A systems analysis of consumer products, In N.A. Stanton (ed.) *Human Factors in Consumer Products* London: Taylor and Francis 75-90

Stanton, N.A. and Baber, C. (2001) Error by design. Design Studies (in press)

Stanton, N. A. and Stevenage, S. (1998) Learning to predict human error: issues of reliability, validity and acceptability. *Ergonomics* 41 (11), 1737-1756

Stanton, N. A. & Young, M. (1998) Is utility in the mind of the beholder? A review of ergonomics methods. *Applied Ergonomics.* 29 (1) 41-54

Stanton, N. A. & Young, M. (1999, a) What price ergonomics? *Nature* 399, 197-198

Stanton, N. A. & Young, M. (1999, b) *A Guide to Methodology in Ergonomics: Designing for Human Use.* Taylor & Francis: London.

Suchman, L.A., 1988, *Plans and Situated Actions*, Cambridge: Cambridge University Press

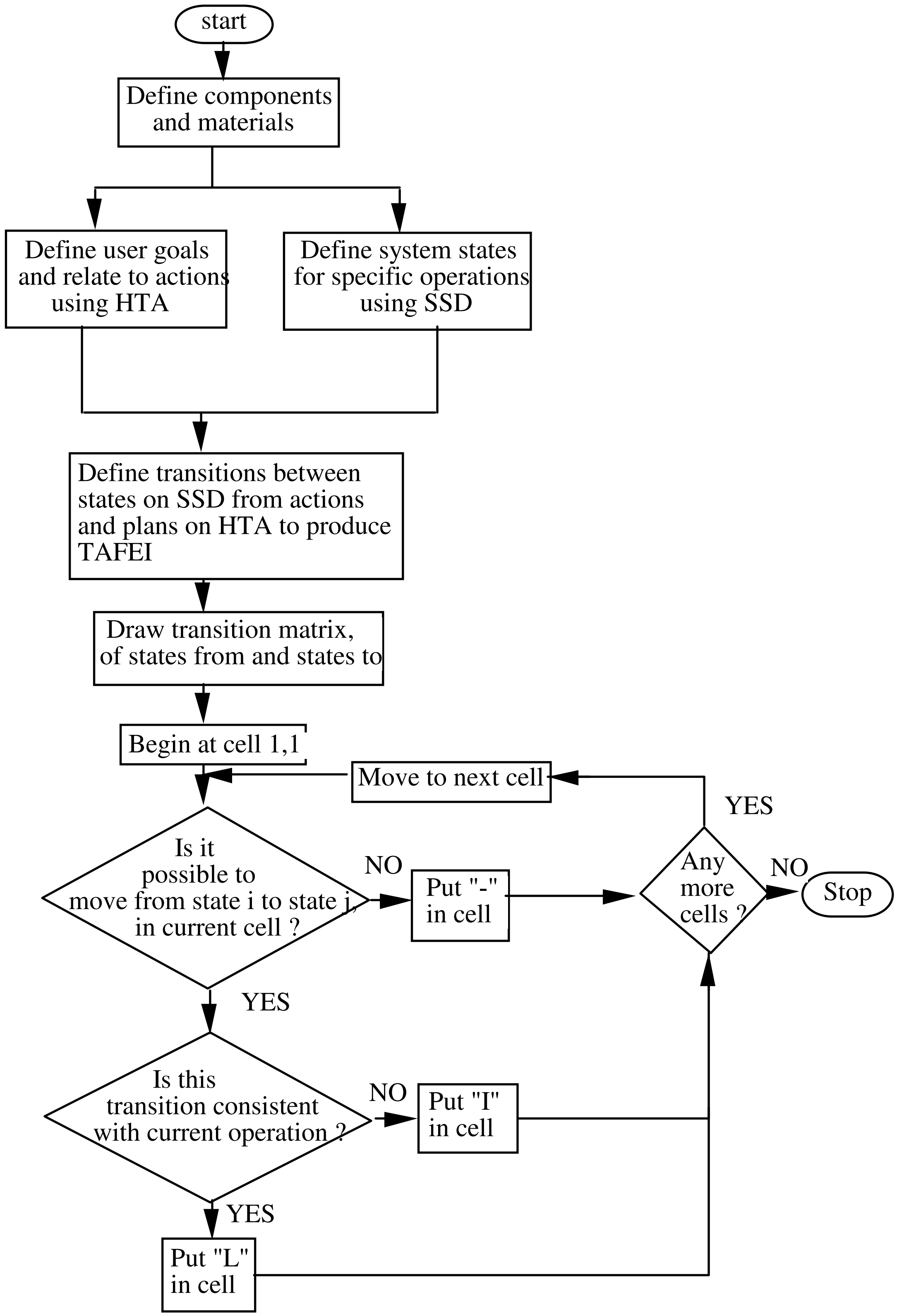
Wogalter, M., deJoy, D. and Laughery, K., 1999, *Warnings and Risk Communication*, London: Taylor and Francis

Zwaga, H., 1988, Buying a ticket for the Washington Metro, In E.D. Megaw (ed) *Contemporary Ergonomics* 1988, London: Taylor and Francis, 273-278

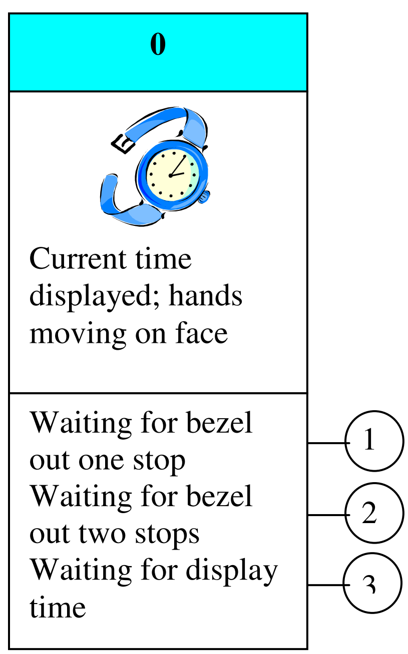
Chris Baber graduated from Keele University in 1987, with a BA(Hons) in Psychology & English. He then went to Aston University to read for a PhD in Speech Technology in Control Room Systems. He obtained his PhD in 1990, and his thesis was subsequently published by Ellis Horwood. He began lecturing at The University of Birmingham in 1990, initially on the MSc Work Design and Ergonomics, and subsequently moving to Electronic, Electrical and Computing Engineering. His current research interests concern human interaction with future technologies (he is working on wearable computers, augmented reality and speech technology), and modelling of human performance (particularly in terms of time and error). He has written or edited four books, some 40 refereed journal papers and 20 book chapters, and in excess of 100 conference contributions. When he is not being an academic, he is the father of four children, a player of the guitar and double bass, a listener of modern jazz music and a reader of contemporary fiction.

Neville A. Stanton is a Professor of Human-Centred Design in the Department of Design at Brunel University. He received his B.Sc. from the University of Hull and his M.Phil. and Ph.D. for the University of Aston in Birmingham, UK. His research interests span most psychological aspects of human interaction with technological systems He has published over 50 international academic journal papers and six books. He was a Visiting Fellow of the Department of Design & Environmental Analysis at Cornell University in 1998. He was awarded the IEE Informatics Divisional Premium Award in 1998 and the Ergonomics Society's Otto Edholm Award in 2001. Professor Stanton is on the editorial boards of "Ergonomics", "Theoretical Issues in Ergonomics Science" and the "International Journal of Human Computer Interaction".

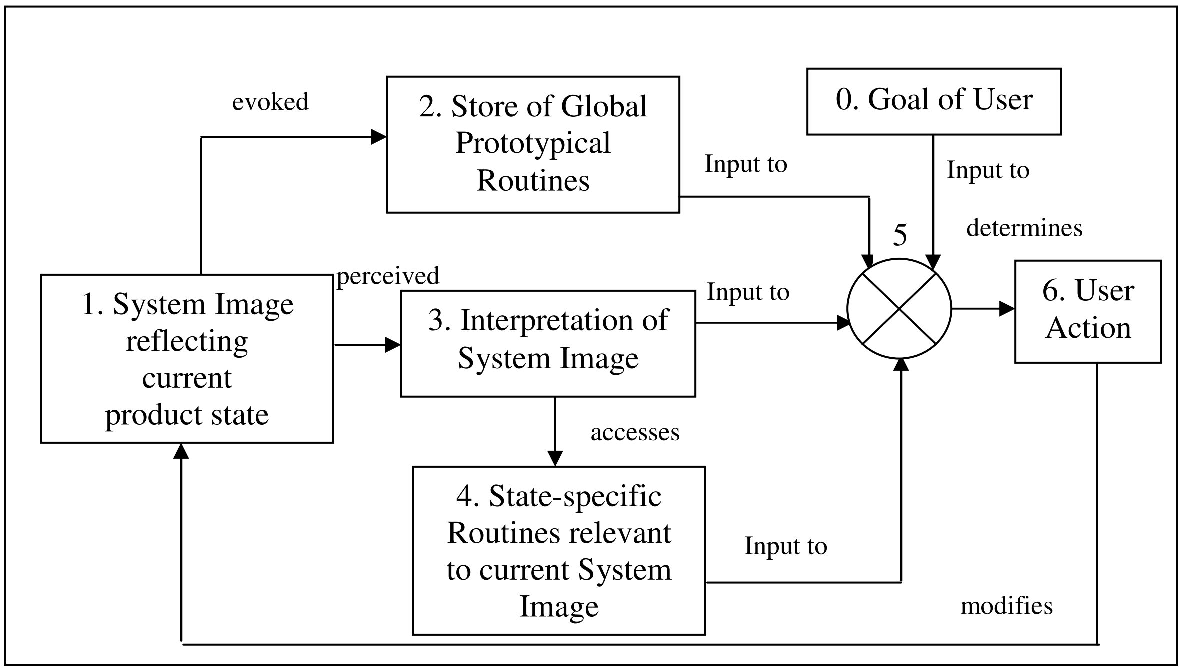
Figures

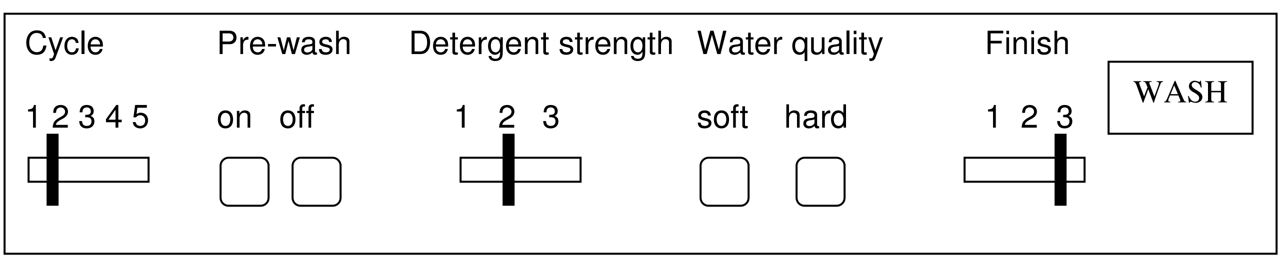

start
Define components and materials
Define user goals and relate to actions using HTA
Define system states for specific operations using SSD
Define transitions between states on SSD from actions and plans on HTA to produce TAFEI
Draw transition matrix, of states from and states to
Begin at cell 1,1
Move to next cell
YES
Is it possible to move from state i to state j, in current cell ?
NO
Put "-" in cell
Any more cells ?
NO
Stop
YES
Is this transition consistent with current operation ?
NO
Put "I" in cell
YES
Put "L" in cell

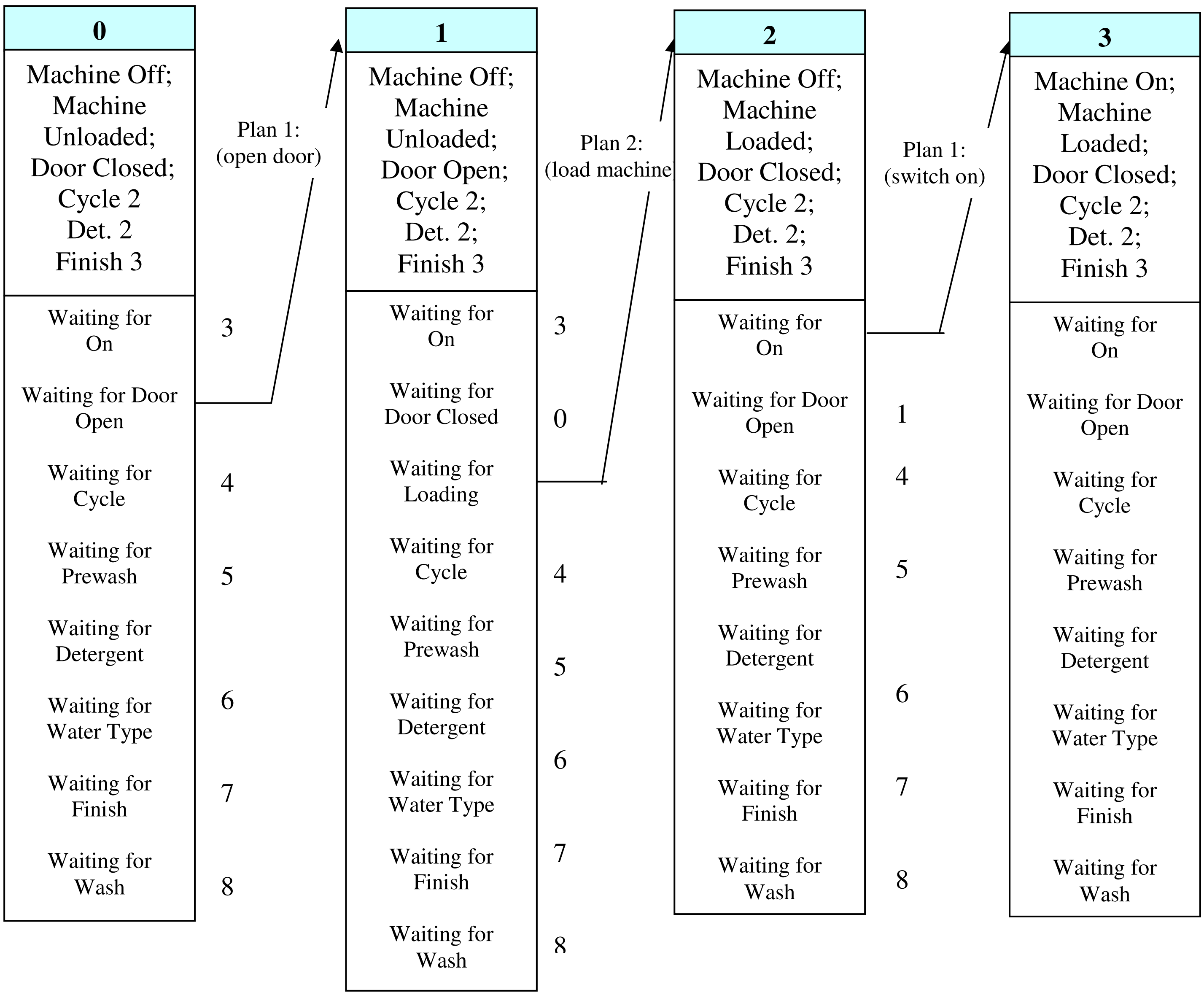
Product:	ACME Answerphone
Primary Goal:	Change greeting message
Subsidiary Goals:	Present a professional image to possible clients
Task Sequence:	1. Lift handset 2. Lift cover of answerphone 3. Press # on keypad 4. Press R (and hold) on answerphone 5. Listen for beep 6. Speak message 7. Release R 8. Press # on keypad
Source:	Procedure written in handbook followed by analyst and notes made.



States to States from	0	1	2
0	I	I	L
1	I	I	I
2	L	I	I






0
Machine Off; Machine Unloaded; Door Closed; Cycle 2 Det. 2 Finish 3
Waiting for On
Waiting for Door Open
Waiting for Cycle
Waiting for Prewash
Waiting for Detergent
Waiting for Water Type
Waiting for Finish
Waiting for Wash
3
4
5
6
7
8
Plan 1: (open door)
1
Machine Off; Machine Unloaded; Door Open; Cycle 2; Det. 2; Finish 3
Waiting for On
Waiting for Door Closed
Waiting for Loading
Waiting for Cycle
Waiting for Prewash
Waiting for Detergent
Waiting for Water Type
Waiting for Finish
Waiting for Wash
3
0
4
5
6
7
8
Plan 2: (load machine
2
Machine Off; Machine Loaded; Door Closed; Cycle 2; Det. 2; Finish 3
Waiting for On
Waiting for Door Open
Waiting for Cycle
Waiting for Prewash
Waiting for Detergent
Waiting for Water Type
Waiting for Finish
Waiting for Wash
1
4
5
6
7
8
Plan 1: (switch on)
3
Machine On; Machine Loaded; Door Closed; Cycle 2; Det. 2; Finish 3
Waiting for On
Waiting for Door Open
Waiting for Cycle
Waiting for Prewash
Waiting for Detergent
Waiting for Water Type
Waiting for Finish
Waiting for Wash