Chapter 6: Ticket Vending Machines

1. Introduction

Prior to the experiments reported in this thesis, the author carried out an error analysis of the ticket vending machines installed in London Underground and overground stations. There were two observational phases, in 1990-91 and 1996. The results from the first phase (Connell 1991) formed part of the material assessed for the author's MSc (awarded by London Guildhall University). The results from both phases were published as Connell (1998). This Chapter re-examines these results in the light of Experiments 1 to 3. Since the work has been assessed elsewhere, only material sufficient to current purposes, plus new analyses, will be presented.

The vending machines study consisted of a brief initial inspection of the interfaces to the London Underground machines[1], followed by lengthy observations of all three machines in use. The results of the (analytic) inspection could thus be compared with those of the later (empirical) observations (adopting the terms used by Gray & Salzman 1998). The form of the analytic inspection was an early version of an error analysis method called Dialogue Error Analysis or DEA (Christie et al. 1995). Each empirical phase consisted of overall tallies of machine and ticket window use, followed by detailed observations of the errors made on each machine. Errors were later classified into nine major categories. The results allowed comparisons to be made between the three machines and the two phases. Measures used included machine vs. ticket window use, failure rates, user error rates and error patterns based on the nine categories.

In Chapter 5 a distinction was drawn between detection rate and hit rate. Detection rate assesses the contribution made by individual problem counts to the total UPTs for a subject population, while hit rate measures the ratio of subjects' correct predictions to total observed problems (eqs. (12) and (13)). It was claimed that hit rate is a more reliable measure of predictive ability, and that cumulative curves based on hit rate will better reflect the performance of a subject group. It was shown that a combination of hit rate and problem distribution was a better measure of combined performance than detection rate alone.

In order to assess hit rate we need a reliable tally of the problems against which predictions are to be measured. But in Experiment 3 there were only seven Test condition subjects. As the experiment stands, we do not know if additional Test subjects would have increased the numbers of observed problems at each task level. We could extrapolate the cumulative curves of observed problems until convergence occurs. But this would presume that the observed problem distribution is somehow more veridical, more like the 'real thing', than the predicted problems; that is, that the contributions of individual Test subjects were not affected by external factors (such as having to make predictions) in the same way as the

1 Study of the third machine began at a later date.

other subject groups in that experiment. In order to assess this, we need some measure of what a 'real' problem profile would be like.

The ticket vending machines study provides just such an opportunity. In the two empirical phases of this study, a total of 1205 user interactions (mainly ticket-purchase attempts) were observed, of which 378 (174 in the first phase, 204 in the second) provide detailed breakdowns of the errors recorded in around 30 hours of observations. By plotting the error distributions and cumulative curves for each of the three vending machines, we can begin to see what observed problem profiles based on large user populations might be like. In particular, we will be able to view the observed totals and combined Test condition performance from Experiment 3 in the light of the data from the earlier study.

2. The Three Machines

The two London Underground ticket vending machines were the Few Fare Machine (MFM) and the Multi Fare Machine (MFM). The overground (formerly Network South East) machine was the QuickFare (QF) machine.

2.1 The Underground Machines (FFM and MFM)

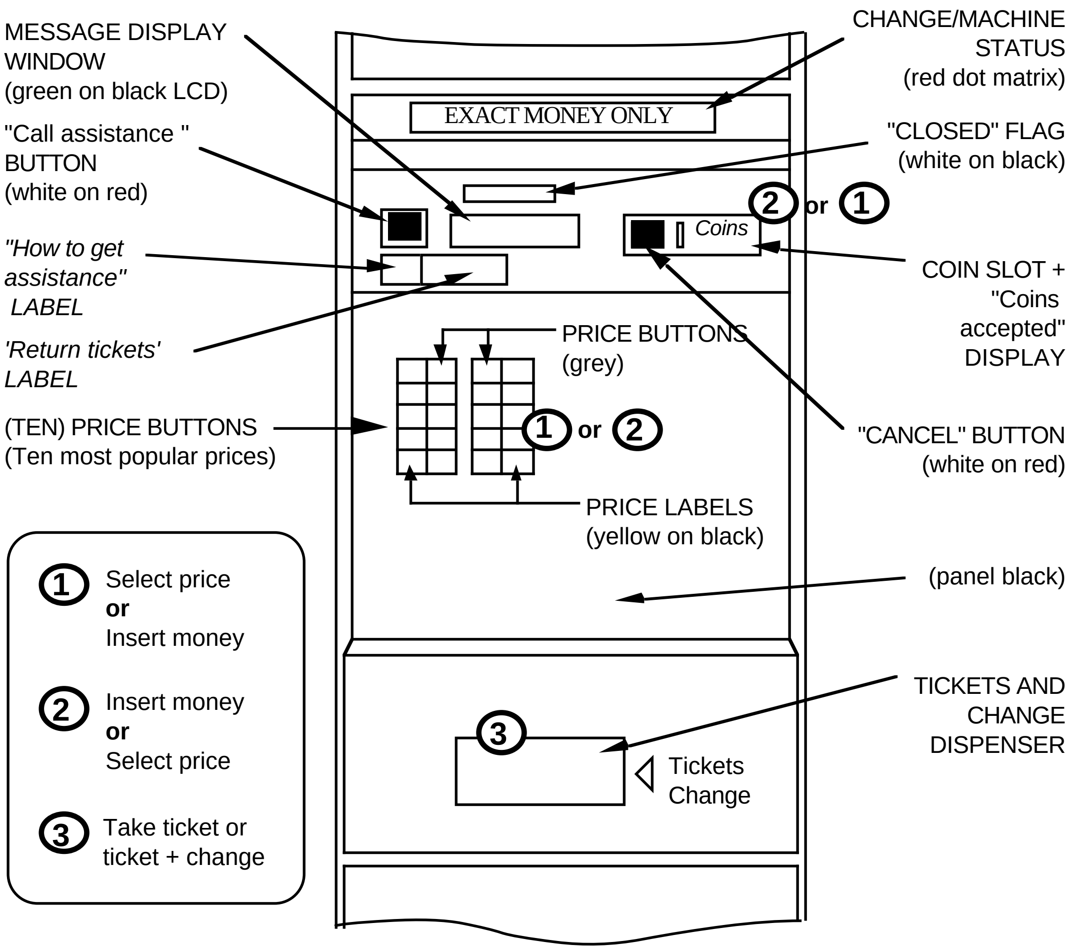

Figure 6.1. Few Fare Machine (FFM) in 1996. Colours of text and button labels are shown as [text on background]. Substantive changes from 1991 are shown in italic. Approximately to scale.

Figure 6.1 shows the smaller FFM as it appeared in 1996. The FFM enables passengers who already know the price and type of their ticket to select from a small range of ten prices. No ticket type selection is necessary. The FFM accepts only coins.

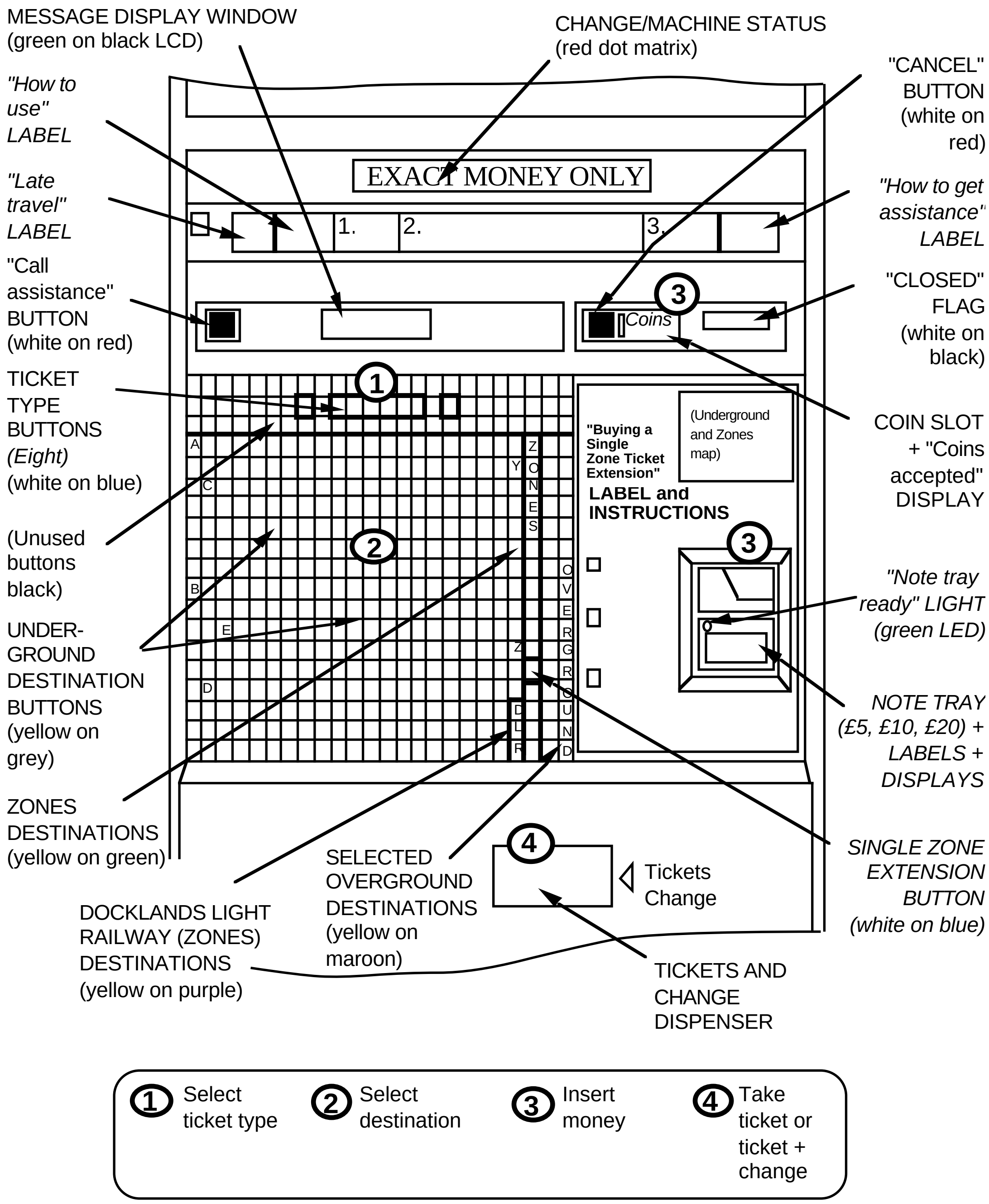

Figure 6.2. Multi Fare Machine (MFM) in 1996. Colours of text and button labels are shown as [text on background]. Substantive changes from 1991 are shown in italic. Not to scale.

Figure 6.2 shows the larger MFM as it appeared in 1996. By then improvements had been made to the labelling, the number of ticket type buttons, and the range of paper money accepted. The MFM offers the complete range of ticket types and destinations available from the underground station in question.

The FFM and MFM require different interaction (user step) sequences. On the FFM, the order of steps 1 and 2, namely **Select price** and **Insert money**, is not enforced. On the MFM, however, not only is the order of steps 1 to 3, namely **Select ticket type**, **Select destination**, **Insert money**, enforced, but re-selection is not permitted. (This order was originally (1990-91) indicated by numbered panels on the machine casing.) Both machines'

Change/Machine Status displays show whether change is currently given (these also show WAIT BY MACHINE when the 'Call assistance' button is pressed).

2.2 The Overground Machine (QF)

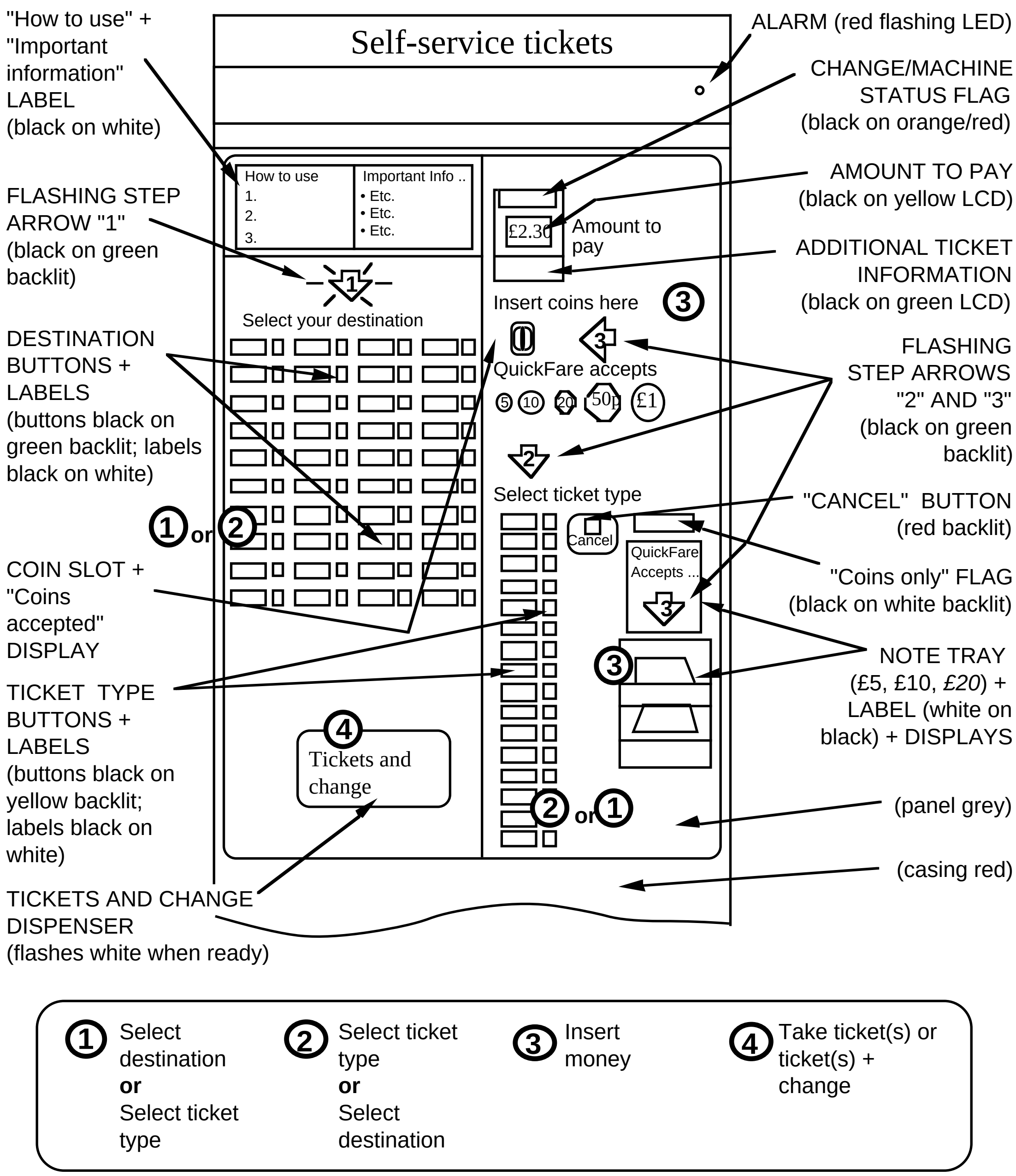

Figure 6.3. QuickFare (QF) machine in 1996. Colours of text and button labels are shown as [text on background]. Substantive changes from 1991 are shown in italic. Not to scale.

Figure 6.3 shows the QuickFare machine as it appeared in 1996. The QF offers the complete range of ticket types available from the overground station in question, plus selected destinations.

The QF and MFM are similar in that they require the user to select both a destination and a ticket type before money is inserted. However, on the QF not only is the order of steps 1 and 2, namely **Select destination**, **Select ticket type**, not enforced, but re-selection (from within the button array) is permitted at both stages. The user can make as many selections as wished, in any order including re-selection, until satisfied. On this machine the order of steps 1 and 2 is therefore only an optimal requirement.

3. Method

3.1 Station Locations

The initial (analytic) inspection was carried out at Oakwood underground station. The underground stations used for the (empirical) observation sessions were Arnos Grove and Highbury & Islington. The overground station was Waterloo. By 1996 Arnos Grove had two FFMs and one MFM, while Highbury & Islington had two of each type. By 1996 the number of QF machines at Waterloo had increased from four to eight.

3.2 Procedure

3.2.1 Initial Inspection (1990)

The initial inspection, of the underground machines only, took place in a single two-hour session. The analyst (the author) attempted to predict the range of errors which might later be observed on each of the two machines.

The procedure used was an early version of Dialogue Error Analysis or DEA (Christie et al. 1995). It involved the identification and later prioritisation of the likely errors arising at each step of the user task (deemed to be the purchase of a ticket). This version of DEA included the assessment of immediate (contingent) and primary (underlying) causes deemed responsible for each error.

The outcome of the initial inspection will be summarised in the Results.

3.2.2 Observation Sessions (1990-91 and 1996)

Each observational session was in two parts: first, overall observations of machine and ticket window use, including successes and failures; second, detailed observations of the errors occurring on each machine. All observations were carried out by a single observer (the author), from behind, and at a sufficient distance to avoid interaction with users. Ticket windows remained open throughout.

In the first part of each session the following were recorded.

1. For all three machines: the number of users who succeeded and failed in getting a ticket or tickets out of the machine.
2. For the underground machines: the number of users who used a ticket window, including those who failed to get a ticket or tickets. (Ticket window use could not be recorded at Waterloo.)

In the second part of each session attention was directed at those attempts that involved errors, whether successfully or not. When there was a failure, one error - the last or only - was deemed to be responsible: this was called the critical error. When there was a success that involved one or more errors, all such errors were deemed to be non-critical.

For each attempt involving error, recorded data included the following.

1. The errors themselves. Occurrences were logged against all known errors, care being taken to note any that were novel (Other) or whose causes could not at attributed at that time (Unknown). Errors were not at that stage categorised.
2. Whether each error was critical or non-critical.
3. The numbers of each error.

3.3 Error Categories

All observed errors were later assigned to one of the following nine categories. Each category placement represents the best attempt at an explanation for that error (in Rasmussen's 1982, 1987 terms, the 'stopping point' for a possible causal sequence).

Timeout (T): This occurred when a user failed to respond to within the time limit for a machine to detect continuous user input. Forced return was made to rest state, requiring a restart.

Change availability (C): Either the user had no (or insufficient) change of the correct type for that machine, or attempt had been made to insert money of the wrong type. If the right type and quantity could not be found in time (or Cancel pressed), the user would be timed out.

Money returned (R): This occurred when no change was being given and money was inserted to more than the ticket price. The consequence on all three machines was that all money was returned, regardless of the margin between price and amount inserted.

Step order wrong (OR): User actions were not in the order prescribed for that machine. Typically order errors occurred at step 1 on both the MFM and the QF. The consequences depended on the machine type: on the MFM the only solution was to press Cancel and start again, but on the QF the selection is accepted. (The status of the QF error is therefore debatable; in the earlier analyses it was assigned lower severity than the MFM equivalent.)

Selection wrong (S): This represents the variety of incorrect (but in the right order) selections that could be made on the three machines. On the MFM only solution was to press Cancel and start again, while on the QF the user could re-select and continue.

Mechanical (M): This included the various mechanical faults (e.g. refusal of money of correct type, coin slot jammed shut) which are familiar with vending machines.

Availability (A): The required ticket type, destination/zone or price was not currently available on that machine. Typically, users were seen to run a finger over the appropriate button array, then give up.

Other (OT): Miscellaneous errors which were not deemed worthy of a category of their own. On the MFM, it included an unexpected piece of user behaviour: users were seen to desert the MFM for the FFM, having found the price of their ticket, even when the MFM *was* giving change.

Unknown (U): Cases where users were seen to give up on a machine (and go straight to a ticket window), having made no other attributable error. Most examples occurred on the MFM and QF: it is thought likely (but nevertheless recorded as Unknown) that users were comparing prices before purchase.

4. Results

4.1 Overall observations

Table 6-1 summarises the overall numbers of machine attempts, successes and failures on all three machines in each of the two observation phases.

	1990 - 91				1996			
	FFM	MFM	Total	QF	FFM	MFM	Total	QF
No. of users going to ticket window (not using machine)	-	-	166	[no data]	-	-	120	[no data]
No. of machine attempts	83	85	168	248	73	82	155	256
Failure rate (%)	1.2	28.2	14.9	15.3	4.1	14.6	9.7	18.4

Table 6-1. Overall machine use during the two phases. Failure rate is the ratio of failures to attempts.

Original (Connell 1991, 1998) analysis revealed that between the two phases the willingness of passengers to use the underground machines had improved relative to ticket window use. The failure rate on the MFM had also declined relative to the other two machines, even though at 15% it might still be considered too high. The failure rate on the QF had increased from 15% to 18%, while that on the FFM had increased from 1% to 4%.

4.2 Error Observations

4.2.1 Errors per User

Table 6-2 shows the mean observed error rates (for unsuccessful attempts) in both phases and detection rates for all three machines in 1996. (Detection rates cannot now be determined for 1990-91 since the original data is no longer available.) Detection rate is here used in its original sense (see Chapter 4, eq (2)), as the mean ratio of observed errors (problems) per user to total errors for all users.

	1990 - 91			1996		
	FFM	MFM	QF	FFM [5]	MFM [12]	QF [13]
Errors per attempt	1.0	1.3	1.2	1.0	1.2	1.4
Detection rate (%)	-	-	-	20.00	9.66	10.43

Table 6-2. Mean errors per unsuccessful attempt (both phases) and detection rates (1996). Each attempt involves a single user. Figures in [brackets] are the number of sole observed errors (UPTs) on each machine.

Error rate: Original analyses within phase revealed that in 1996, QF users were making significantly more errors per attempt than those on either of the other two machines, whereas in 1990-91 the rate of QF error-making appeared to be no greater than that on the other machines. In 1996, as in 1990-91, the MFM error rate was significantly higher than that on the FFM.

Detection rate: Subsequent analysis shows that in 1996, the FFM detection rate was significantly higher than that for either of the other two machines (FFM/MFM/QF: one-way ANOVA, F[2,161]=24.30, p<0.001; FFM/MFM and FFM/QF: p<0.001 one-tailed and two-tailed). The similarity of the FFM error rates in the two phases (and the small number of errors on that machine) suggests that this was also true in 1990.

4.2.2 Error Distributions and Cumulative Curves

Figure 6.4 shows the distributions and cumulative curves (actual and probability) of observed errors on the three machines in 1996 (data from phase 1 is no longer available). It is clear that in spite of the relative lack of skewing in the distributions (no preponderance of single-incidence errors, for example), the contributions of individual users were insufficient for either of the larger two machines to comply with a ‘3 to 5’ (between 3 and 5 users to reveal 75% of problems). In contrast to the FFM, with its very few failures (Table 6-1) and higher detection rate (Table 6-2), the numbers of users required to reveal 75% of observed errors on the MFM and the QF were around 25 and 30 respectively. Further, all three curves can be seen to diverge from their theoretical equivalents. (Similar patterns are revealed for critical errors.)

Thus even given the large numbers of users and the relatively objective nature of the data-gathering, the two larger machines appear to have exhibited observed error patterns more like those generated by the inspections in Experiments 1 to 3 than is claimed for user testing (in e.g. Nielsen 1994c, Nielsen 1994b, Virzi 1992). The likely reasons for this will be taken up in the Discussion.

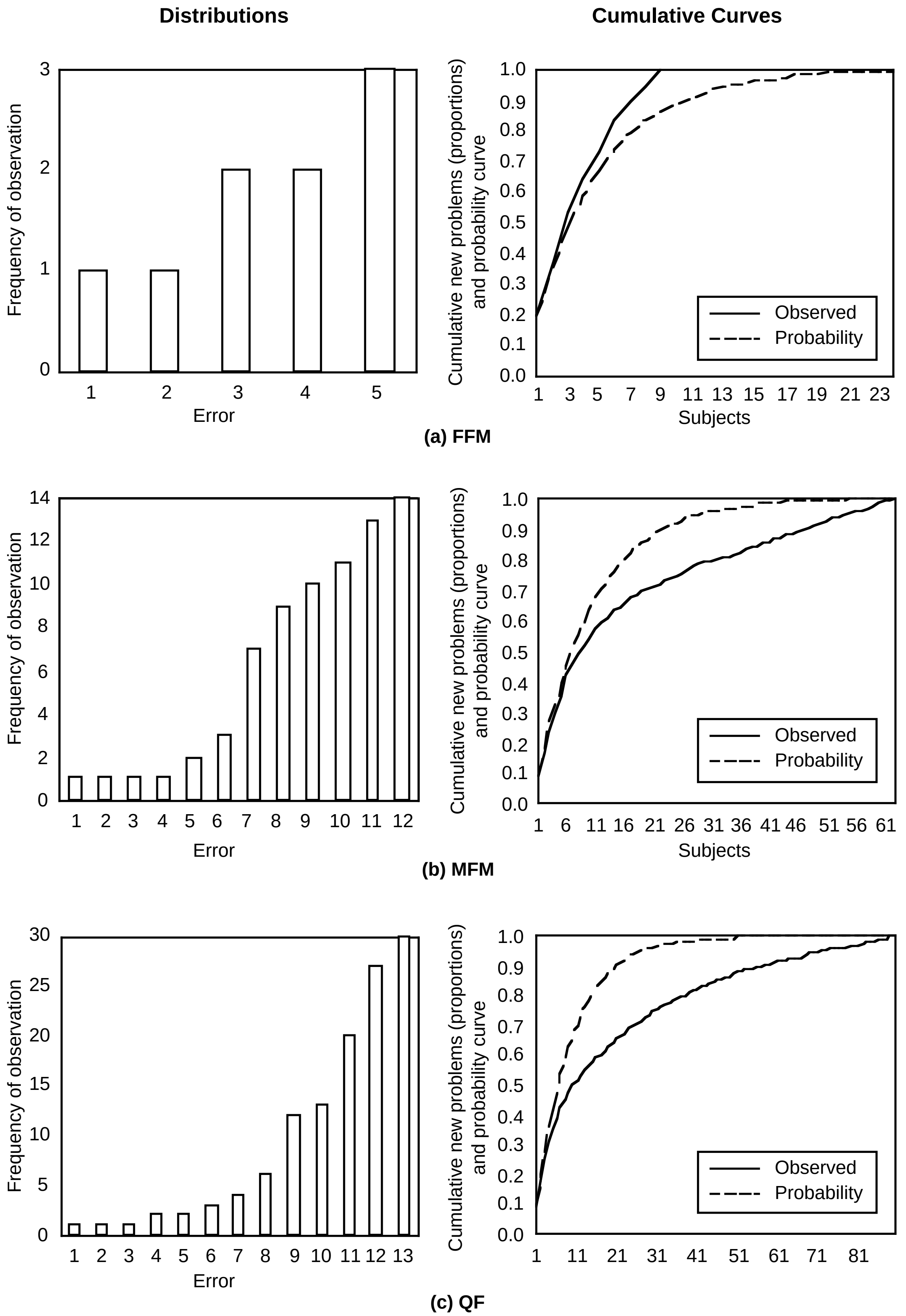

Figure 6.4. Phase 2 (1996). Error distributions and cumulative curves (observed and probability) for the (a) Few Fare Machine (FFM), (b) Multi Fare Machine (MFM) and (c) QuickFare machine (QF).

4.3 Predicted Versus Observed Errors

In this Section we will compare the errors observed on the MFM and FFM in 1990-91 with those predicted in the initial (1990) inspection. This analysis will allow us to assess the relationship between hits, misses and FPs on those machines. (Due to the six-year interval between the initial inspection and phase 2, these comparisons will be done for phase 1 only.)

The results of the initial inspection (Dialogue Error Analysis) on the underground machines are summarised in Table 6-3. Items E1 to E11 represent the full set of errors (some common to both machines) which were predicted at that time. Priorities are the product of observed frequency and assigned seriousness. The frequency [1...4] of an error was deemed to be the same as the frequency of the task step with which it was associated. Seriousness [1...4] was assigned according to the consequences for the user of the error in question. In this case, all steps were deemed necessary for successful operation, so priority could be assigned according to seriousness alone.

Error	Description	Priority
FFM		
E4	More than price inserted when no change given	4
E2	No change of correct type	3
E1, E5	Timeout (first or all money not inserted in time)	2
E6	"Call assistance" button pressed for "Cancel" button	2
E3	Wrong coin(s) used	1
MFM		
E4	More than price inserted when no change given	4
E10	Note not accepted	4
E11	Coin slot jammed	4
E2	No change of correct type	3
E9, E5	Timeout (destination or all money not inserted in time)	2
E6	"Call assistance" button pressed for "Cancel" button	2
E7	Destination/zones button selected for type button	2
E8	Ticket type button selected for destination/zones button	2
E3	Wrong coin(s)/note used	1

Table 6-3. Prioritised [1.. 4] (low .. high) error listing resulting from the initial (1990) inspection of the underground machines (FFM and MFM).

Table 6-4 shows all the critical and non-critical errors which were observed on the MFM in 1991, now grouped according to the nine categories described above. Included are the predictions featured in Table 6-3.

Description of error	Category code	Predicted ?	Total
Timeout (coin/note insert / all money in / button press not in time	T	E9, E5	12
No or insufficient change (of correct type)	C	E2	10
Coin(s)/note of wrong type used	C	E3	4
More than ticket price inserted when no change given	R	E4	4
Destination/zones button pressed instead of ticket type	OR	E7	15
Attempt to insert money at start	OR		4
Wrong destination/zone selected	S		7
Ticket type button pressed (again) instead of destn/zone (i.e. wrong type selected)	S	E8	5
Coin/note (of correct type) rejected	M	E10	2
No coin/note (of correct type) accepted	M		1
Coin slot jammed	M	E11	3
Destination / type / zone(s) not available	A		2
Used as price-finding machine when change given	OT		3
Call Assistance pressed for Cancel	OT	E6	1
"Wait by machine" showing (not Cancelled)	OT		1
Unknown (give up ? experimenting ?)	U		5
Total	16	9	79

Table 6-4. MFM predicted and observed errors, phase 1 (1990-91). Predicted errors refer to Table 6-3. T = Timeout; C = Change availability; R = Money returned; OR = Step order wrong; S = Selection wrong; M = Mechanical; A = Availability; OT = Other; U = Unknown.

Out of the 16 errors (including Unknown[2]) which were observed, 9 had been successfully predicted (hits). All of the initially predicted errors were observed at least once in 1991 (i.e. zero false positives). Using the terminology introduced in Chapter 5, this yields a hit rate (%hits/observed) of 56.3%, with accuracy (%hits/predictions) and redundancy (%FPs/predictions) of 100% and 0% respectively. See Table 6.5.

	Observed	Not observed	
Predicted	9 (hits)	0 (FPs)	9
Not predicted	7 (misses)		
	16		

Table 6-5. MFM hits, misses and false positives (FPs) between the initial inspection (1990) and phase 1 (1990-91).

Similar analysis of the FFM predictions in Table 6-3 (observed errors not illustrated) yields a hit rate of 25%, with accuracy and redundancy of 40% and 60% respectively. See Table 6-6. However, the low failure rate of 1% on this machine (Table 6-1) meant that only 9 errors in total were observed compared to 79 on the MFM. Thus predicted - observed comparisons for this machine may be less reliable.

	Observed	Not observed	
Predicted	2(hits)	3 (FPs)	5
Not predicted	6 (misses)		
	8		

Table 6-6. FFM hits, misses and false positives (FPs) between the initial inspection (1990) and phase 1 (1990-91).

2 In Connell (1998) Unknown errors were not included in that predicted-observed comparison.

While 56% and 25% are respectable hit rates (with caveats concerning the paucity of FFM data), it was shown in Connell (1998) that they could be increased to 86% and 71% respectively by comparing predicted and observed error *categories* (as listed in Section 3.3 and Table 6.4) rather than *actual* errors. It was also shown that most predicted errors fell within one or two priority points of the observed errors of the same type (observed priority was taken as the product of the frequency of occurrence and the ratio of critical to total errors). In Section 6 these results will be discussed in the light of the types vs. instances issue introduced in Chapter 4.

4.4 Experiment 3: Test Condition

In this Section we will compare the pattern of results just generated for the ticket vending machines with that for the Test condition in Experiment 3.

We saw in Chapter 5, Figure 5.2 that in Experiment 3 the distributions of predicted and observed problems were comparable on the Skill and Rule level tasks but not the Knowledge level task. It was shown in Figure 5.5 of the same Chapter that only at the Rule level and in the Skill Heuristic condition were hit rates high enough to reach the corresponding targets of observed problems. However, it is possible that more Test subjects would have revealed additional problems than those uncovered by the seven who took part (thus pushing down hit rates still further).

We saw in Section 4.2.2 above that the patterns of observed errors on the MFM and QF were such that large numbers of users would have been required to achieve convergence. If this is so for the relatively simple interfaces and closed tasks embodied by ticket vending machines, it is possible that more complex interfaces and tasks would require at least similar numbers of inexperienced users to reveal all potential problems. Thus at the Experiment 3 Knowledge level, and perhaps at the Skill level also, we might expect further Test condition novices to exhibit a comparable lack of convergence. (Rule level convergence was shown to occur within 1 or 2 subjects.)

Figure 6.5 shows the proportionalised cumulative curves for Experiment 3 Skill and Knowledge level Test condition subjects. We can see that even though the detection rates - 37% and 31% respectively - are high enough for a ‘3 to 5’, in both cases the curves begin to diverge from their ideal (probability) equivalents after 5 or 6 subjects. This *implies* that further subjects would have been required to exhaust the potential problems which these tasks embodied.

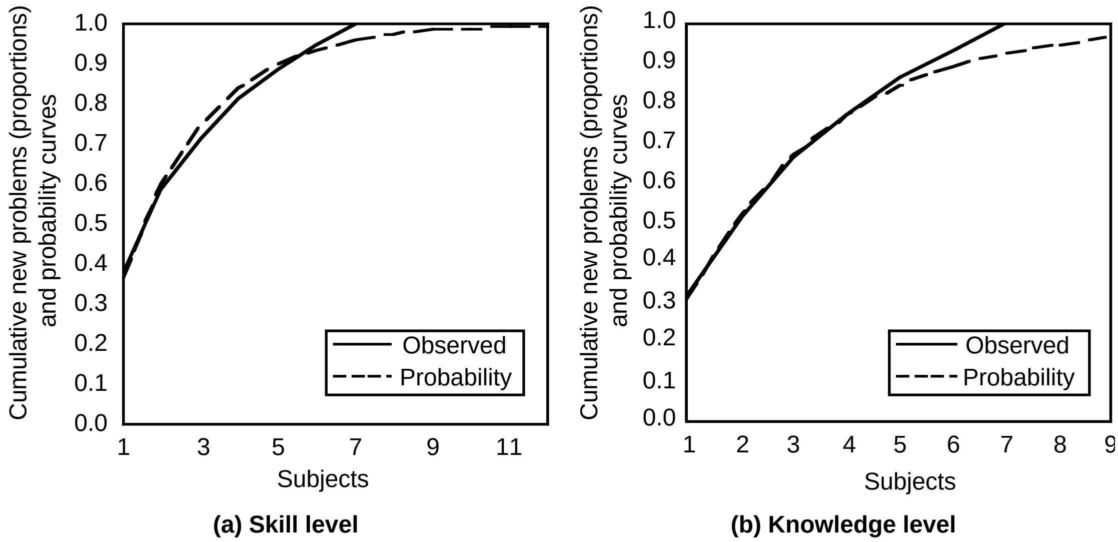

Figure 6.5. Cumulative and probability curves for Experiment 3 Test condition subjects, on (a) Skill and (b) Knowledge level tasks.

If this analysis is correct, at these task levels the number of Test subjects needed to produce reliable totals against which to assess predictive problem counts would have been considerably more than those used in Experiment 3. (Extrapolation of the absolute Skill and Knowledge curves suggests that 75% of the new totals would have required around 19 and 17 subjects respectively.) This *suggests* that observed problem totals based on all but very simple tasks may be unreliable unless derived from large subject populations, and that the totals (and hence hit rates) from the Skill and Knowledge levels in Experiment 3 are also suspect. This issue will be taken up in the Discussion.

5. Summary of Results

1. The original (Connell 1991, 1998) analysis showed that use of the two London Underground ticket vending machines (FFM and MFM) had increased between the two observational phases (1990-91 and 1996).

2. The original analysis showed that between the two phases the MFM failure rate had declined (from 28% to 15%) relative to the other two machines, whereas that on the other machines (FFM and QF) had increased (from 1% to 4% and from 15% to 18%, respectively).

3. The original analysis revealed that in 1996 the rate of error-making per user was significantly greater on the QF than on the other two machines, whereas in 1990-91 it had not been. In both phases the MFM error-rate was greater than that on the FFM.

4. Subsequent analysis for this thesis revealed that in 1996 the detection rate of FFM users was significantly greater than that on both the other two machines. It is likely that this was also true in 1990-91.

5. The detection rate differences were reflected in the cumulative curves of observed errors. Only on the FFM did convergence occur early enough for a '3 to 5'. All three curves showed divergence from their theoretical (probability) equivalents.

6. The relationship between predicted (initial inspection) and observed (phase 1) errors on the underground machines was re-assessed in the light of Experiment 3. MFM hit rate and accuracy were shown to be more than twice that on the FFM (though very few FFM errors had been recorded). It was recalled that higher hit rates could be generated by counting by error categories (types) rather than actual errors (instances).

7. The cumulative curves of Experiment 3 Test condition subjects were compared with those of the larger numbers of vending machine users. It was suggested that in Experiment 3 the observed problem totals at the Skill and Knowledge level may have been larger than those derived from the seven test subjects, thus driving down the associated hit rates still further than reported in that Chapter.

6. Discussion

This fresh look at the earlier results has served two purposes. First, it was possible to put the vending machines data in the context of the analysis presented in this thesis. Second, the Experiment 3 Test condition data was compared with that derived from the large number of vending machines observations.

6.1 Vending Machines Data

The original data showed that although the use of the underground (and probably overground) machines had increased between the two observational phases, failure rates had declined by almost half on the MFM while those on the other machines had increased. The FFM failure rate had increased threefold, though to just 4%. The FFM and MFM changes may be due to a practice effect, in that by 1996 the expectation of these machines had outstripped what they could deliver: the FFM's limited ticket range, the MFM's inflexibility in regard to re-selection and money acceptance. More users appeared to be prepared to put up with the MFM's deficiencies (though still with only 85% success), while the FFM had encouraged three times more people to look for tickets not on offer. The QF's *apparent* selection and order requirements continued to cause sub-optimal performance, though these errors are not enforced (thus representing good examples of the 'task fit (poor support)' error genotype outlined in Sutcliffe et al. 2000).

The above failure rates masked differences in the rate of error-making on each machine. The (mostly non-critical) unenforced QF errors were probably responsible for the higher per-user error rate on this machine in 1996, while, as we would expect, the MFM error rate was consistently higher than that on the FFM (and in both phases there was a greater proportion of critical MFM than critical QF errors). However, subsequent analysis shows that in 1996 the FFM detection rate (20%) was significantly higher than that on either of the other machines (10% and 11% respectively). This is reflected in the cumulative curves for the three machines, which reveal that only FFM users managed to achieve convergence early enough for a '3 to 5'. In addition, all three curves diverged from their theoretical (probability) equivalents. Thus, once more, this analysis has failed to confirm the predictions of Nielsen and others for user testing, in that only the relatively simple and error-free FFM generated a curve sufficient to support a '3 to 5'. This is so even given the large numbers of users and the extensive observational data gathered in phase 2. Attempt will be made in Chapter 8 to account for the form of these and the other curves presented in this thesis.

The Experiment 1 and 2 expedient of counting by type rather than instance, whereby problem counts deriving from higher-level categorisation were shown to generate detection rates sufficient for a '3 to 5', was seen to have had a parallel in the vending machines study. The original data had shown that hit rates of 25% and 56%, generated by comparing the predicted and observed errors on the FFM and MFM respectively, could be increased to 71% and 86% by using the nine error categories rather than the actual ('raw') errors. It had also been shown that the predictions fell within one or two priority points of nine error types. We can now see that hit rates such as these would generate cumulative curves better than those for the Rule level tasks in Experiment 3 (which, in turn, resemble the predictions made by Nielsen (1992) for 'double specialists'). With due caveats concerning the small number of observed FFM errors, it is likely that the 'raw' hit rates are more representative of the reality, namely, that on their own, inspection methods such as Dialogue Error Analysis are unlikely to generate hit rates much better than 50%.

6.2 Experiment 3 Test Data

In Experiments 1 and 2 failure rate was not measurable due to the deliberately open-ended task requirements. However, Experiment 3 offers limited basis for comparison. We saw in Chapter 5 that no subjects failed to complete all seven tasks in that experiment, even though part of the experimental manipulation was that subjects were not expected to find the Rule level shortcuts or complete the Knowledge level filtering without assistance. Inspection of the problem records reveals that 100% of Rule and Knowledge subjects (in all three conditions) duly failed. However, there were some differences in performance at the Skill level for Test condition subjects. (Heuristic and Principle subjects' failures were recorded only in the context of their predictions, so will not be reported.) On Skill tasks 2 and 3 (move cells, emphasise words), 3 out of 7 (43%) Test subjects left errors (e.g. wrong cell placement, spaces also made bold), while on task 1, (move sentences) 2 subjects (29%) left errors (e.g. spaces not retained). (There were no failures on task 5.) This very limited data implies that success on even simple Skill level tasks cannot be taken for granted, unless the desired outcome is as obvious to subjects as the delivery of a train ticket.

Finally, it was seen in Figure 6.5 that the Experiment 3 Skill and Knowledge level curves diverged from their theoretical equivalents after 5 or 6 subjects, implying that additional Test subjects would have uncovered further problems. Since the MFM and QF curves also deviated from the theoretical form, it is likely that this much smaller study is not unrepresentative. If so, then perhaps 35 rather than 21 Skill problems and 25 rather than 14 Knowledge problems (extrapolating the actual curves in Figure 5.4 of Chapter 5) might have been revealed by user numbers comparable with those in the vending machines study. This would make the Experiment 3 hit rates still lower than those reported in Chapter 5. However, comparison of Figures 6.4 and 6.5 of this Chapter shows that the vending machines curves run inside their corresponding theoretical (probability) forms, while those for Experiment 3 run outside. It is the author’s belief that these different forms of divergence from the ideal model are indicative of particular problem distributions. However, this hypothesis remains untested.

Summary of Chapter 6

The results of an earlier study by the author of the London Underground and overground ticket vending machines (the FFM, MFM and QF) were summarised and re-examined in the light of previous Chapters. Discussion focused on the detection rates and cumulative curves from the three machines, along with predicted and observed errors vis-a-vis the instances vs. types issue introduced in Chapter 4. The reliability of the Experiment 3 Test condition data was also examined in relation to the vending machines study.