Cognitive Engineering Laboratory

Research on the Characteristics of Long-Term Adaptation

Xinyao Yu, Renée Chow, Greg A. Jamieson, Rasha Khayat, Elfreda Lau, Gerard L. Torenvliet, Kim J. Vicente, & Michael W. Carter

CEL 97-04, Volume 2

Final Contract Report
September, 1997

Prepared for
Japan Atomic Energy Research Institute

Cognitive Engineering Laboratory University of Toronto Department of Mechanical & Industrial Engineering
5 King's College Rd. Toronto, Ontario, Canada M5S 3G8
Phone: (416) 978-7399 Fax: (416) 978-3453
Email: benfica@mie.utoronto.ca URL: www.mie.utoronto.ca/labs/cel/

Cognitive Engineering Laboratory

Director: Kim J. Vicente, B.A.Sc., M.S., Ph.D.

The Cognitive Engineering Laboratory (CEL) at the University of Toronto (U of T) is located in the Department of Mechanical & Industrial Engineering, and is one of three laboratories that comprise the U of T Human Factors Research Group. CEL began in 1992 and is primarily concerned with conducting basic and applied research on how to introduce information technology into complex work environments, with a particular emphasis on power plant control rooms. Professor Vicente's areas of expertise include advanced interface design principles, the study of expertise, and cognitive work analysis. Thus, the general mission of CEL is to conduct principled investigations of the impact of information technology on human work so as to develop research findings that are both relevant and useful to industries in which such issues arise.

Current CEL Research Topics

CEL has been funded by Atomic Energy Control Board of Canada, AECL Research, Alias|Wavefront, Asea Brown Boveri Corporate Research - Heidelberg, Defense and Civil Institute for Environmental Medicine, Honeywell Technology Center, Japan Atomic Energy Research Institute, Natural Sciences and Engineering Research Council of Canada, Rotoflex International, and Westinghouse Science & Technology Center. CEL also has collaborations and close contacts with the Mitsubishi Heavy Industries and Toshiba Nuclear Energy Laboratory. Recent CEL projects include:

- Studying the interaction between interface design and adaptation in process control systems.
- Understanding control strategy differences between people of various levels of expertise within the context of process control systems.
- Developing safer and more efficient interfaces for computer-based medical devices.
- Designing novel computer interfaces to display the status of aircraft engineering systems.
- Developing and evaluating advanced user interfaces (in particular, transparent UI tools) for 3-D modelling, animation and painting systems.

CEL Technical Reports

For more information about CEL, CEL technical reports, or graduate school at the University of Toronto, please contact Dr. Kim J. Vicente at the address printed on the front of this technical report.

ABSTRACT

This final contract report describes the findings from the first year of a two-year research program investigating the characteristics of long-term operator adaptation in nuclear power plants (NPPs). The report consists of three volumes. This document, volume 2, describes the results obtained from a number of different empirical analyses of long-term operator adaptation. The measurement approaches that were identified in volume 1 as being promising were used to analyze data from a six-month longitudinal study previously conducted for JAERI. The focus was on analyzing performance during normal, start-up tasks only. A number of novel measures of performance based on a dynamic systems approach were developed. Data analyses based on these measures provided a more detailed insight into the performance differences between subjects. Several measures of operator action variability were developed, the most promising of which involved a synthesis of the dynamic systems approach and the abstraction hierarchy. This novel set of measures shed a great deal of light on the strategy differences between subjects. The best subject using the traditional interface was controlling the system at the level of actions, whereas the best subject using the EID interface was controlling the system at a higher level of abstraction, a more sophisticated form of coordination. Other measures of adaptation based on a dynamic systems approach showed that the more proficient subjects exhibit a greater degree of adaptation to the structural constraints imposed by the system. Furthermore, normative and descriptive analyses of strategy selection revealed that more proficient subjects also adapt their behaviour to the limitations of their own information processing capabilities. The implications of these findings for the design and evaluation of human machine systems in nuclear power plants (NPPs) are outlined, as are the contributions and limitations of this research.

ACKNOWLEDGEMENTS

This research project was sponsored by a contract from the Japan Atomic Energy Research Institute (Dr. Fumiya Tanabe, Contract Monitor), as well as research and equipment grants from the Natural Sciences and Engineering Research Council of Canada. We would like to thank Roger Conant, Yan Yufik, and Dr. Tanabe for their contributions and help.

TABLE OF CONTENTS

OVERVIEW

A predictive model of human cognitive behavior, which includes the mental strategies used in emergency situations in nuclear power plants (NPPs), is needed in the design and evaluation of human-machine systems. Existing models, however, suffer from several limitations. In order to improve the situation, the previous research conducted over the last three years has investigated the relative impact of interface content, theoretical training, self-explanation, and visual momentum on operator adaptation in the context of the DURESS II microworld. A considerable data base has been accumulated during the past three years, which has potential to give us valuable insights with further detailed investigation.

In this year's project, methods to appropriately represent characteristics of long-term adaptation as well as the impact of individual cognitive style were investigated. The results obtained will be useful to the development of a model of human operator cognitive behavior and to the development of criteria for the design and evaluation of human-machine systems. Volume 1 describes the results of a literature review of potential measures of operator adaptation. This document, volume 2, describes the results that were obtained by applying the previously identified adaptation measures to a six-month, longitudinal data set collected for a previous contract (Christoffersen, Hunter, & Vicente, 1994). Finally, volume 3 describes the results from regression analyses that sought to uncover the relationship between individual differences — particularly, the serialist/holist cognitive style (Pask & Scott, 1972) — and various measures of performance.

1. INTRODUCTION

In volume 1 of this final contract report, we presented a literature review of approaches to the measurement of operator adaptation. The results of that review indicated that information theory, and especially dynamics, seemed to be promising measurement approaches. In this volume, we describe our efforts to apply these frameworks by conducting a number of analyses of an existing data set. The objective of this work is to determine if the measures identified in the literature review provide more insight into long-term operator adaptation. The data set that we chose to evaluate the measures should be consistent with this goal.

Fortunately, an experiment that we had already conducted with DURESS II during our earlier work for JAERI (Christoffersen et al., 1994) suits this purpose very well. First, this experiment examined operator performance on a quasi-daily basis for an unprecedented period of six months. Thus, it provides a rare opportunity to investigate adaptation over a very long period compared to that which is usual for laboratory studies.

Second, that experiment manipulated the interface that participants were given. Half of the subjects received an interface designed according to the principles of EID which contained physical and functional information (P+F), whereas the other half received a more traditional interface containing physical information (P). Thus, this data set allows us to investigate whether the proposed adaptation measures are sensitive, not just to changes with experience, but also to differences across interfaces as well as the interaction between interface and experience.

Third, because of its longitudinal nature, this experiment only included six subjects, three in each interface group. While this resulted in a decrease in statistical power, it has the advantage of making the data analysis with different adaptation measures more manageable. Because there are relatively few subjects, it is easier to examine each individually to see how our understanding of their behaviour varies as a function of the measures we use to look at the data.

Fourth, in previous analyses of the data from this experiment, we noted that the best performer in each group (AV in the P+F group, and TL in the P group) reported using very different strategies for controlling the system once they were experienced. As mentioned in volume 1, AV reported using a function-oriented strategy which focused more on the goals to be achieved rather than on exactly what actions were required to achieve those goals. In contrast, TL reported using an action-oriented strategy which focused more on a precisely defined set of actions to perform the task. Note, however, that this difference in strategies was based on verbal reports (more specifically, data from the control recipes knowledge elicitation measure), not on behavioral data. Therefore, it would be very interesting to see if any of the adaptation measures we adopt to look at this data set can distinguish between these two strategies for controlling DURESS II. From this perspective, the reported differences between AV and TL represent a prototypical case that can be used to evaluate the sensitivity and diagnosticity of our proposed adaptation measures. In summary, this six-month longitudinal study provides a particularly appropriate set of data for evaluating measures of operator adaptation.

Scope

Because this was our first comprehensive attempt at developing objective, analytical measures of operator adaptation, we thought it would be prudent to narrow down the scope of the data we examined. While data were available for three different control tasks (startup, tuning, and shutdown), we decided to focus solely on startup tasks in this investigation. Furthermore, we only examined data from normal (i.e., non-fault) trials. While the differences between individuals and interface groups are likely to be much more prominent for fault trials, we thought it would be better to start off by focusing on the analytically simpler case of normal trials. Once we acquire a deeper understanding of the comparative value of different adaptation measures, then we will be in a better position to investigate the more complex, but potentially more informative, cases of tuning and fault management (both of which can be viewed as perturbations). In next year's contract, most of our efforts will be devoted to investigating how the measures developed in this year's contract can elucidate participants' reactions to perturbations.

We also deliberately avoided conducting any statistical tests in our analyses. As mentioned in the proposal for next year's contract, there are problems with traditional statistical tests that we have not yet resolved. One of the work items we will be addressing in next year's project is to review the statistical literature to identify more meaningful and more appropriate statistical techniques. At that time, we can use the results from this year's project to evaluate the suitability of different statistical methods. Thus, the analyses presented here are purely descriptive and do not contain any inferential statistics.

Outline

The remainder of this volume is organized as follows. First, we will briefly describe the data analysis tool that we developed as part of this project. Second, we will present several analyses of subjects' performance using measures based on a dynamic systems approach. Although these analyses do not examine adaptation per se, they do show which subjects are more proficient than others. These differences provide a useful context for understanding differences in adaptation. Third, we present several analyses of the variability in operator actions. As mentioned in volume 1, action variability is one of the ways in which we expect to be able to differentiate between different forms of adaptation. Some of these analyses turned out to be

particularly sensitive and important. Fourth, we describe the results of several analyses whose aim was to determine to what extent operators' actions were adapted to the state of DURESS II. As we will see, these analyses were not as informative as we had hoped. Fifth, we will present a complementary set of analyses whose aim was to determine to what extent operators' actions were adapted to the structure of DURESS II. These analyses showed differences between more and less proficient operators, as well as revealing some differences in strategies as well. Sixth, we present a final set of analyses that evaluated the extent to which operators adapted their behavior, not just to the structural constraints of DURESS II, but also to the psychological constraints imposed by their own information processing limitations. Finally, we conclude by describing the contributions of this research, its implications for the design and evaluation of human-machine systems in NPPs, its limitations, and a set of future research topics.

2. DATA ANALYSIS TOOL

As part of our data analysis effort, we developed a software tool, ADAPT (Adaptation Data Analysis & Processing Tool), that is described in full detail in a companion report (Yu, Khan, Lau, Vicente, & Carter, 1997). This tool serves several important roles. First, after pre-processing the DURESS II binary log files into a format that is more readable, ADAPT interpolates the data in the log files so that the state of the simulation is available between operator actions (the DURESS II simulator only logs data at the time of an operator action). Second, the tool provides a library of data analysis functions, each of which represents a different way of looking at a particular data set (i.e., a different set of measures). Third, ADAPT generates output files with the results of each analysis, either in graphical or alphanumeric format. For all of these reasons, ADAPT provides a valuable resource for this and future projects since it provides a relatively efficient and flexible means of analyzing data from any experiment with DURESS II from a number of different perspectives.

The user's manual for ADAPT (Yu et al., 1997) provides a detailed description of precisely how almost all of the measures described in the remainder of this volume were calculated. Thus, we will not repeat that information here. In addition to the measures reported here, ADAPT is capable of performing a larger suite of analyses. Up to this point, mainly those data analysis functions used in this research have been document in the user's manual. As this research

programme progresses, an increasing number of these functions will be documented and used in analysis.

3. ANALYSES OF PERFORMANCE

In this section, we present a number of analyses, not of adaptation per se, but of performance. These analyses will shed some light on which subjects are the most proficient at normal startup trials. In addition, we will be able to see how subjects' performance changed over the course of the experiment. Both of these sets of findings will provide some background information that will be useful in interpreting the various adaptation measures that are discussed in the remainder of this document.

Previous Analyses

Before presenting the new analyses we have performed, we will very briefly summarize the results from some of the analyses we had already conducted before the present project (Christoffersen et al., 1994; Howie, 1996). Table 1a shows the results for the first and last 22 normal trials, not including the transfer trials. The means and standard deviations for the time to complete startup tasks are indicated. Beginning with within-group comparisons, AV and TL are clearly the fastest in their respective groups. AS is clearly the slowest in the P+F group. ML was the slowest in the P group early on, whereas WL was the slowest at the end of the experiment. As for between-group comparisons, the following pairs of subjects demonstrated very similar absolute levels of mean performance: AV and TL, IS and ML, and AS and WL. These subject pairs can be considered cohorts in the sense that they exhibit a comparable level of performance with different interfaces after extensive practice. Note, however, that the P+F subjects demonstrated a higher level of consistency. This variability effect was statistically significant and was observed throughout the experiment, up to the transfer trials. At this point the P subjects, who were now using the P+F interface, became less variable than the subjects who were now using the P interface (Christoffersen et al., 1994).

Table 1b shows the results for the routine and non-routine faults, again measured by time to complete the trial. These data are a better indication of skill and knowledge in that fault trials, especially the non-routine type, were more challenging than normal trials. Here we find that AV was again the fastest in the P+F group. TL was the fastest in the P group for routine faults, but WL was the fastest for non-routine faults. AS and ML were the slowest again in the their

respective groups. In contrast to the normal trials, the fastest subjects in each group are no longer comparable; AV was by far the fastest overall. In fact, if one matches the subjects across groups as cohorts in terms of performance, one finds that the P+F subject was always faster than the respective P subject. This pattern held even for the slowest pair of subjects; AS was faster than ML, who was only able to complete one of the three non-routine faults.

		Trials 1-22		Trials 196-217	
Group	**Subject**	**Mean**	**SD**	**Mean**	**SD**
P+F	AS	860.2	441.3	437.2	31.3
	AV	517.3	149.0	353.5	16.1
	IS	644.4	125.8	399.0	17.5
P	ML	660.3	261.8	390.2	48.0
	TL	493.6	80.2	357.5	20.5
	WL	624.3	170.4	437.2	97.9

Table 1a: Summary of performance differences on normal trials. Average startup task completion for first and last 22 normal trials.

		Routine		Non-Routine	
Group	**Subject**	**Mean**	**SD**	**Mean**	**SD**
P+F	AS	868.3	362.6	1166.0	164.0
	AV	570.0	127.3	740.3	89.3
	IS	667.0	140.8	1030.5	101.1
P	ML	883.3	314.4	1255.0	*
	TL	686.5	264.3	1172.0	93.3
	WL	687.8	164.6	1022.0	53.7

* This subject was only able to successfully complete 1 of the 3 non-routine fault trials.

Table 1b: Summary of performance differences on fault trials. Average trial completion time for routine and non-routine faults.

With these relative differences in mind, we now turn to the new performance analyses we have conducted as part of this project.

State Space Measures

If we adopt a dynamical systems perspective to performance measurement (see volume 1), we can characterize the behavior of each reservoir in DURESS II as a trajectory in a two-

dimensional goal space defined by the temperature and output demand flowrate for that reservoir (cf. Sanderson, Verhage, & Fuld, 1989; Howie & Vicente, in press). For a successful startup trial, the trajectory would begin at the origin (because the system is shut down at the start of the trial) and end at the goal state. By normalizing the temperature and output flow values with respect to the target values for that trials, we can meaningfully compare trajectories across trials.

Trajectories in goal space. The result of such an analysis can be depicted as trajectories in goal space. An example is given in Figure 1, which shows the first trial successfully completed by each of the P+F subjects. The trajectories are clearly inefficient, taking a wild, disorderly path until the settle at the goal point (1,1). This is not surprising given that the subjects were just learning to control the system.

Line length of trajectories in goal space. Although the goal space graphs provide important insights, they are inefficient because they show the behavior of only one trial on each graph. Furthermore, they only provide us with qualitative insights into performance based on visual inspection of the graphs. To overcome these two limitations, we used the length of the path in the goal space as a quantitative measure of performance (Howie & Vicente, in press). We expected that the path lengths for more proficient subjects should be shorter than those for less proficient subjects, and that the path lengths of all subjects should decrease with experience.

Figures 2 and 3 show the results of this analysis for the P+F and P groups, respectively, using graphs of line length in goal space vs. trial for each reservoir. There was a marked improvement in consistency with experience for most subjects, except ML who was the worst subject overall on this measure. As in the measures presented above, TL was clearly the best in the P group. Interestingly, TL seems to exhibit an abrupt improvement in performance on the upper reservoir starting around trial #100, suggesting a change in strategy. We will revisit this point later in this section. As for the P+F group, IS was the most consistent by the end of the experiment. Comparing Figures 2 and 3, it seems that the P+F group as a whole was more consistent than the P group, particularly by the end of the study.

There is another interesting piece of information that can be obtained from Figures 2 and 3 if we take into account the minimum line length. Recall that the trajectories begin at the origin and end approximately at (1,1). Thus, the minimum line length that can be possibly be attained is 1.414 (i.e., $\sqrt{2}$). This value would be obtained if subjects moved along the temperature and output demand axes at an equal rate. The data reveal that this minimum value was never

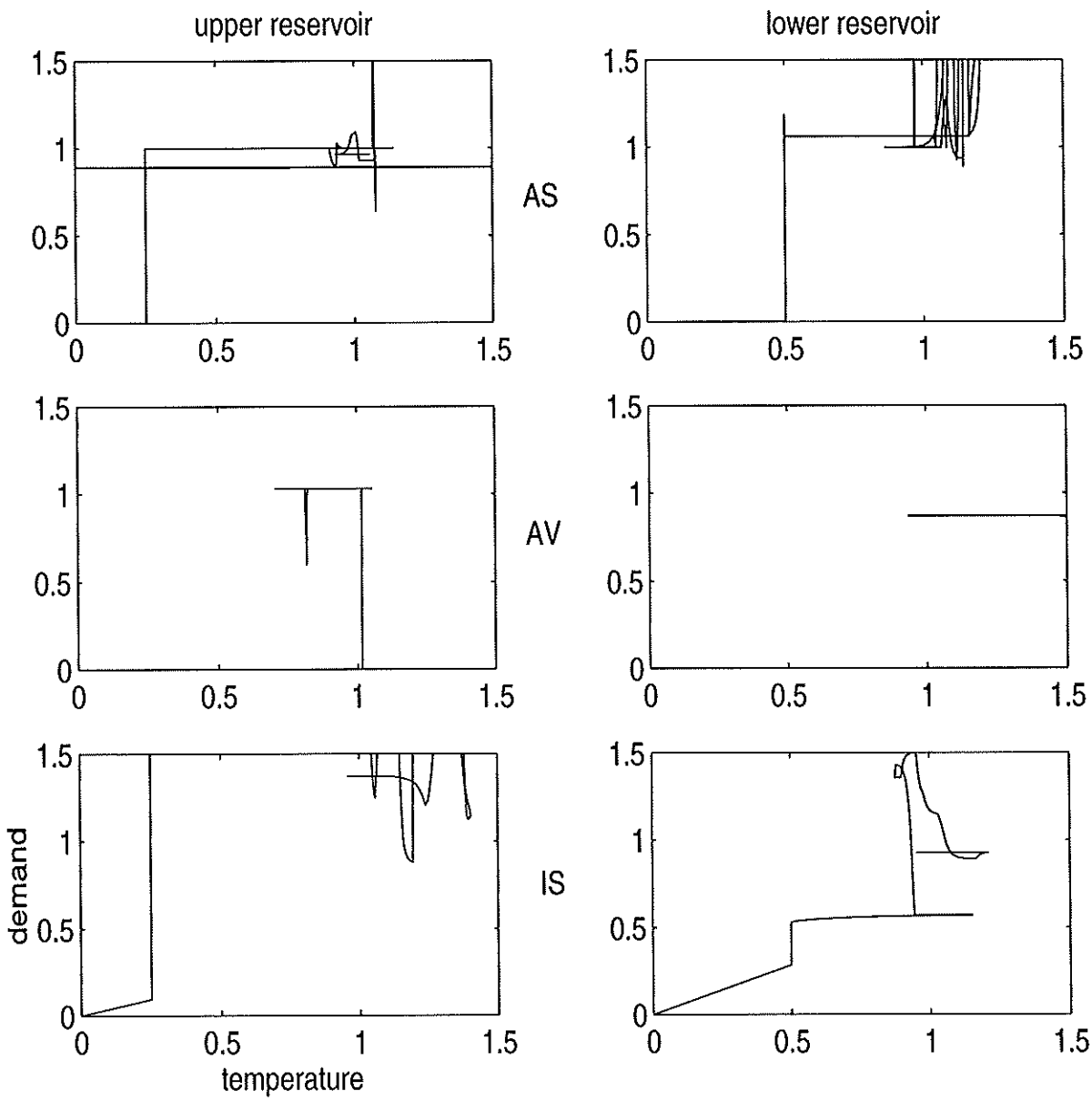

Figure 1: Trajectories in goal space of the first successful trial of P+F subjects

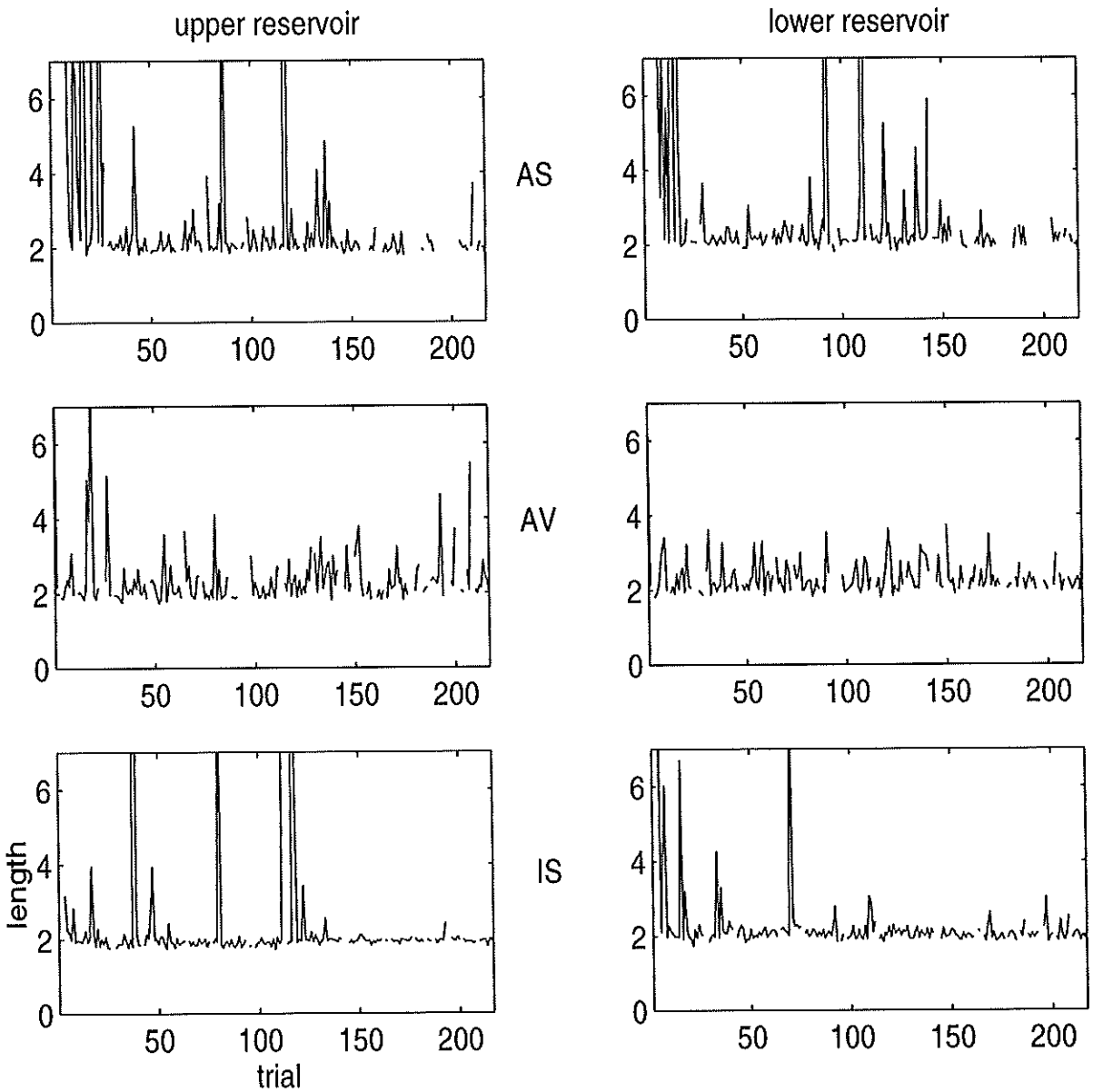

Figure 2: Lengths of trajectories in goal space of P+F subjects

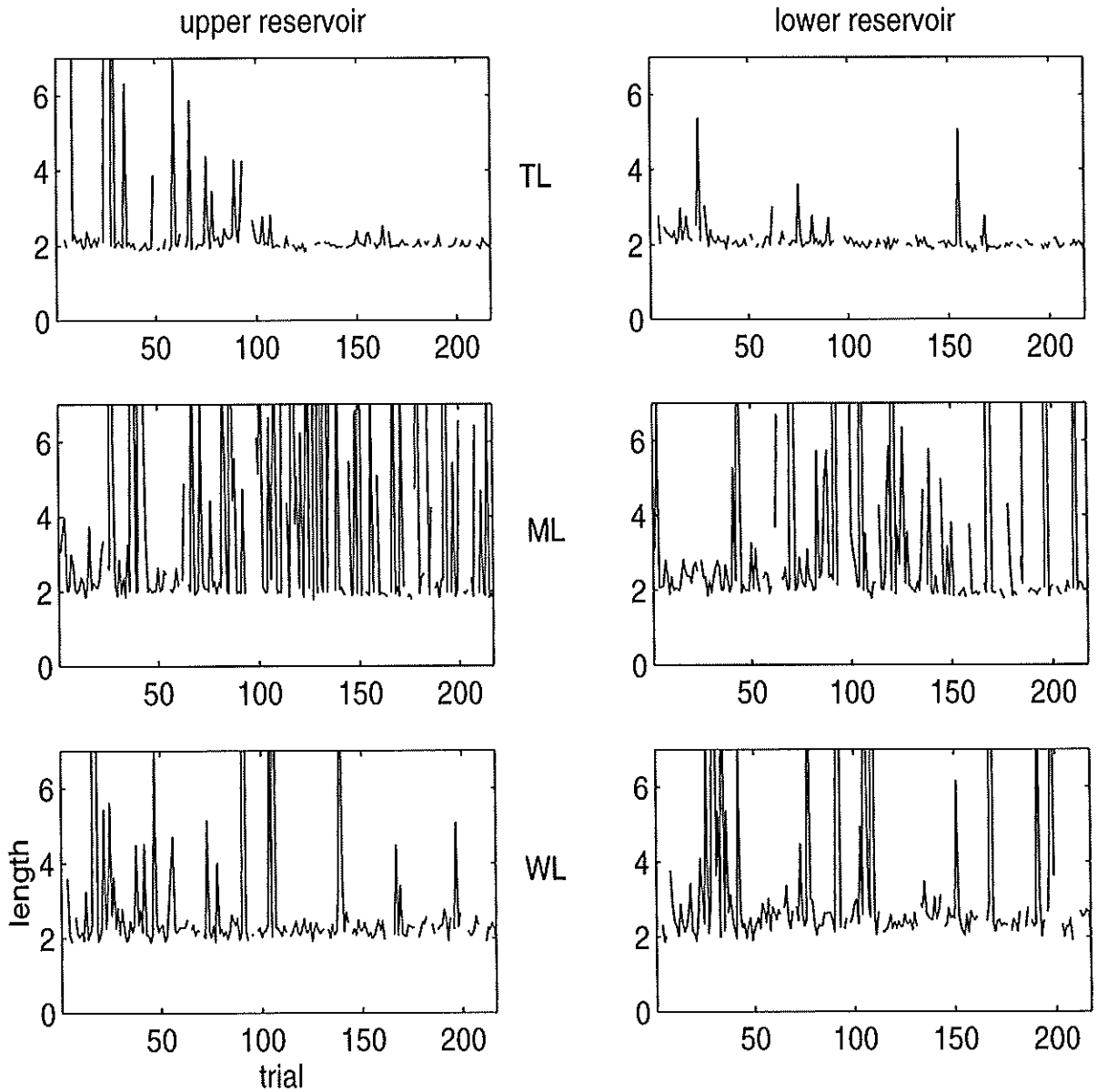

Figure 3: Lengths of trajectories in goal space of P subjects

approached by any subject. The shortest path lengths seem to be just below 2. Interestingly, 2 is the line length we would expect to obtain if subjects were following a sequential, univariate control strategy, moving first along one axis and only then along the other. The individual goal space graph trajectories reinforce this interpretation. Most subjects usually waited until the temperature was well on its way to the target value before they opened the output values to satisfy the output demand goals.

Distance to goal. Another way of deriving a higher-order measure from the goal space graphs is to calculate the distance to the goal as a function of time for each trial (Shaw, Kadar, Sim, & Repperger, 1992; Howie & Vicente, in press). Again, we can normalize the data by dividing them by the goal target value for each trial, thereby allowing for meaningful comparisons across trials. At the beginning of the trial, subjects are at a distance of 1.414 from the goal state for each reservoir (see Figure 1). By the end of the trial, they should be at a distance of approximately zero. Figure 4 shows an example of distance to goal graphs for the first normal trial completed by each subject in the P+F group. As we would expect of novices, the distance to goal paths take a rather inefficient, and in some cases meandering, path.

Area under distance to goal. As before, we can summarize the information in these graphs by deriving a higher-level measure of performance. In this case, we can calculate the area under the distance to goal graphs to create a measure of efficiency (Howie & Vicente, in press). We would expect that more proficient subjects would exhibit smaller areas, and that the areas of all subjects should decrease with experience.

Figure 5 shows the results for the P+F group while Figure 6 shows the results for the P group. All subjects show signs of improvement with practice, some (e.g., IS) more than others (e.g., ML). TL is the best subject in the P group, while AV is the best in the P+F group. The performance of AV in particular is notable, as evidenced by the consistently low values he exhibits throughout much of the experiment. In fact, the P+F group as a whole seems to be more consistent than the P group.

Summary. These state space analyses of performance led to the following observations:

- subjects' performance improved over the course of the experiment
- TL was the most proficient subject in the P group
- AV and IS were the most proficient subjects in the P+F group
- TL seemed to change his strategy at about trial #100

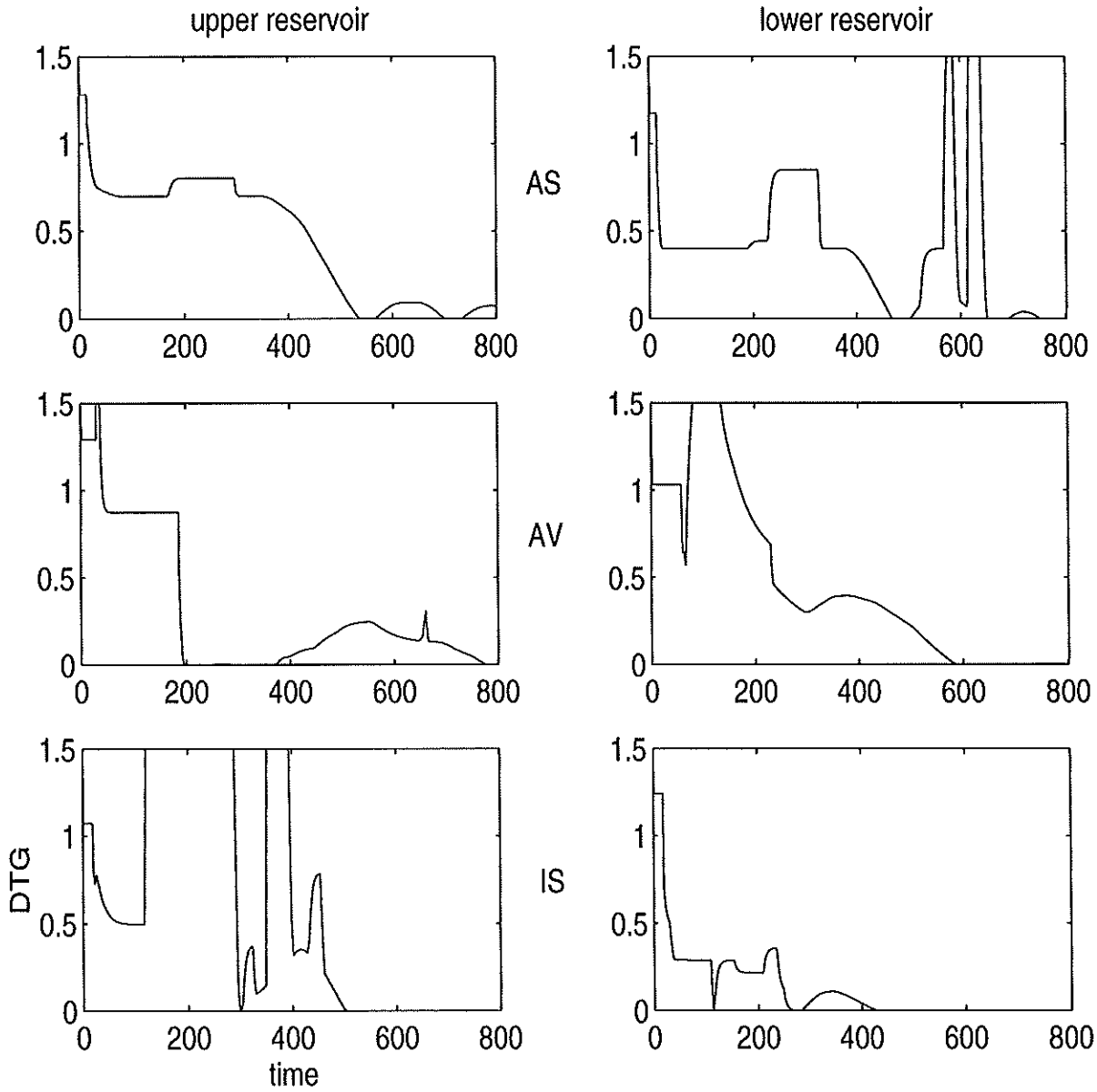

Figure 4: Distance to goal of the first successful trial of P+F subjects

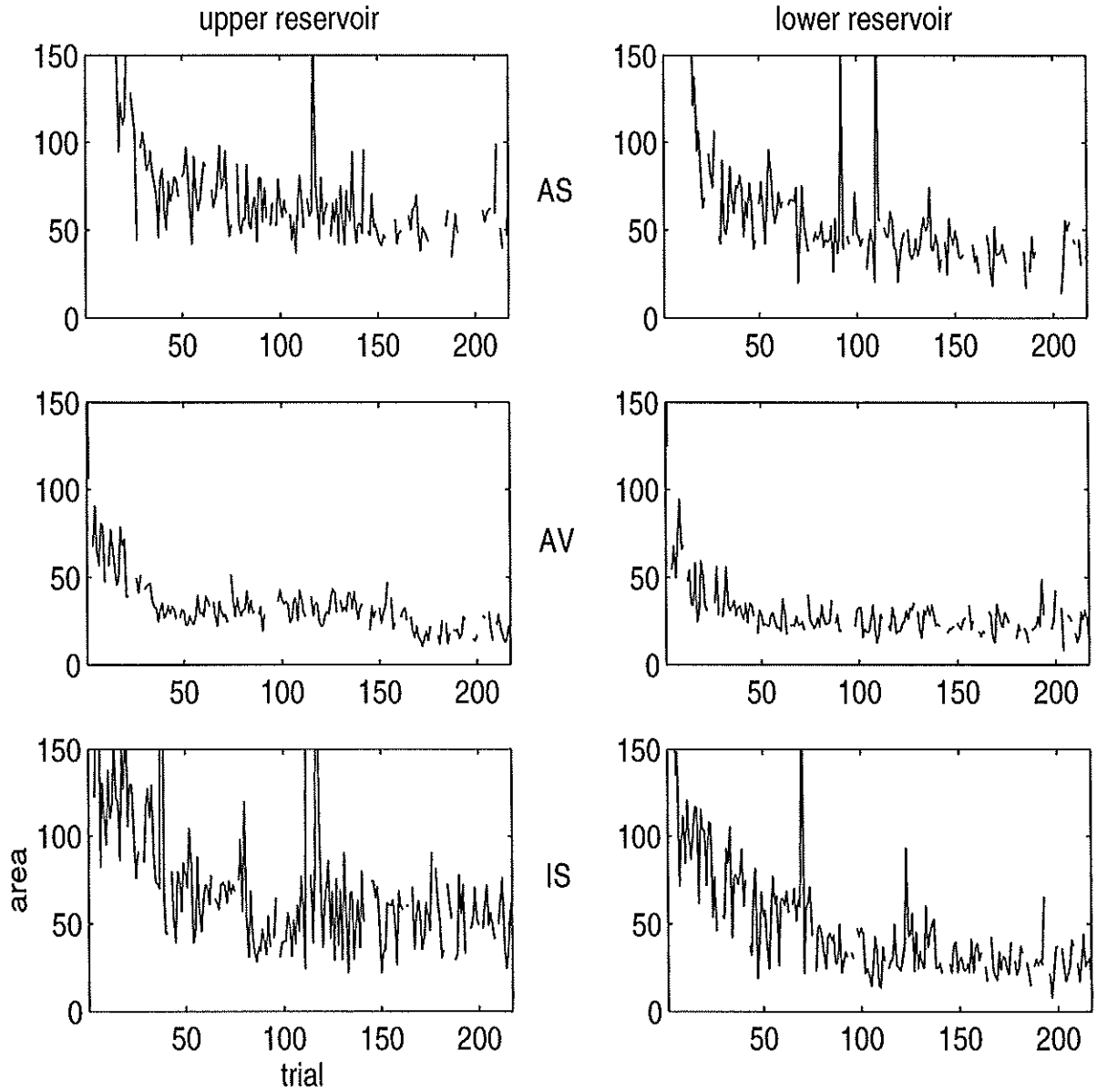

Figure 5: Area under distance to goal of P+F subjects

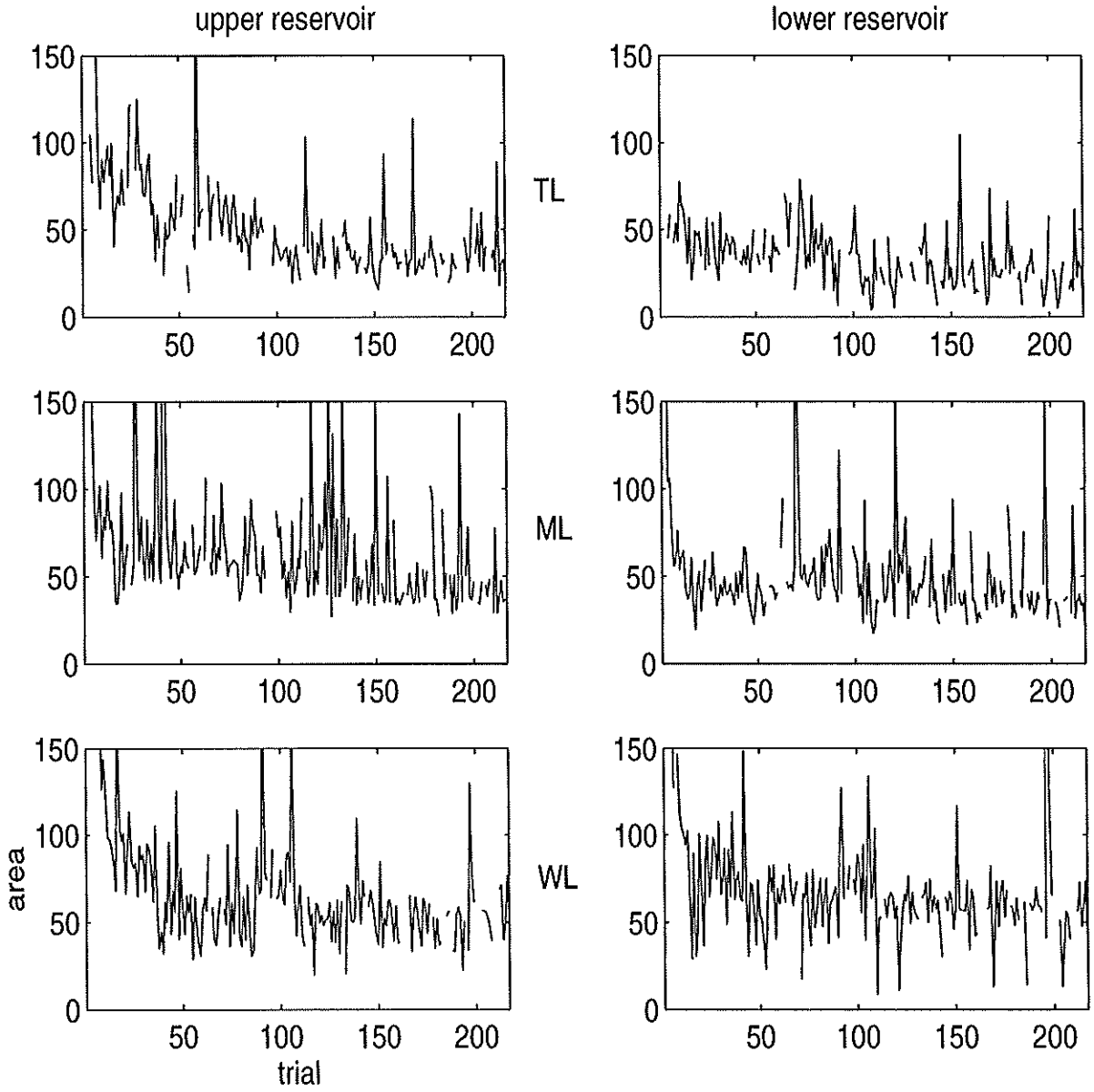

Figure 6: Area under distance to goal of P subjects

- the performance of the P+F group seemed to be more consistent than that of the P

While these trends are interesting, they do not give much insight into the locus of the differences in subjects' performance. For this, we need more fine-grained measures of performance.

Towards More Fine-Grained Measures

Throughout this experiment, subjects faced different challenges during different parts of the trial. For example, at the beginning of a trial, the immediate subgoal is to turn on the required components in order to get the system heading toward the direction of the goal state. The next concern might be to get the system under control and in the goal region. Finally, during the latter part of the trial, the subgoal is to make sure that the goal variables stay within the required tolerances. We found it useful to divide each trial into these three phases. Then, we were able to develop measures that were suited to the characteristics of each phase. Together, these measures provided us with a relatively comprehensive understanding of the differences in performance between subjects, thereby complementing the performance analyses we had conducted earlier (see above and Christoffersen et al., 1994).

Figure 7 shows four graphs, one for each goal variable ($T1$, $T2$, $D1$, $D2$) vs. time. We can use this figure to illustrate how we operationalized each of the three phases just described. Phase 1 begins at $t = 0$s and ends when a goal variable first makes contact with the lower bound of that goal region. Phase 2 begins when phase 1 ends and continues until the goal variable enters and remains within its target boundaries. During this phase, the subject is attempting to stabilize the goal variable and, as a result, the trajectory is marked by oscillations across the goal boundaries, as shown in Figure 7. Phase 3 begins at the end of phase 2 and continues until the end of the trial. Thus, it comprises the time during which steady state has been reached (i.e., the goal variable remains within the target region for at least 300s).

Together, Figure 7 and Table 2 provide an example of how the three phases can be identified. Several points are worth noting. First, the duration of phase 3 will equal exactly 300s for the last goal variable to enter and remain in the goal region, and will probably exceed 300s for the other three variables. Thus, the duration of the phase 2 is constrained by both phases 1 and 3. Second, it is possible for phase 2 to have a duration of zero, as in the case of Figure 7d. This will only occur if there is no oscillation (i.e., if the trajectory goes directly from 0 into the goal region and stays there).

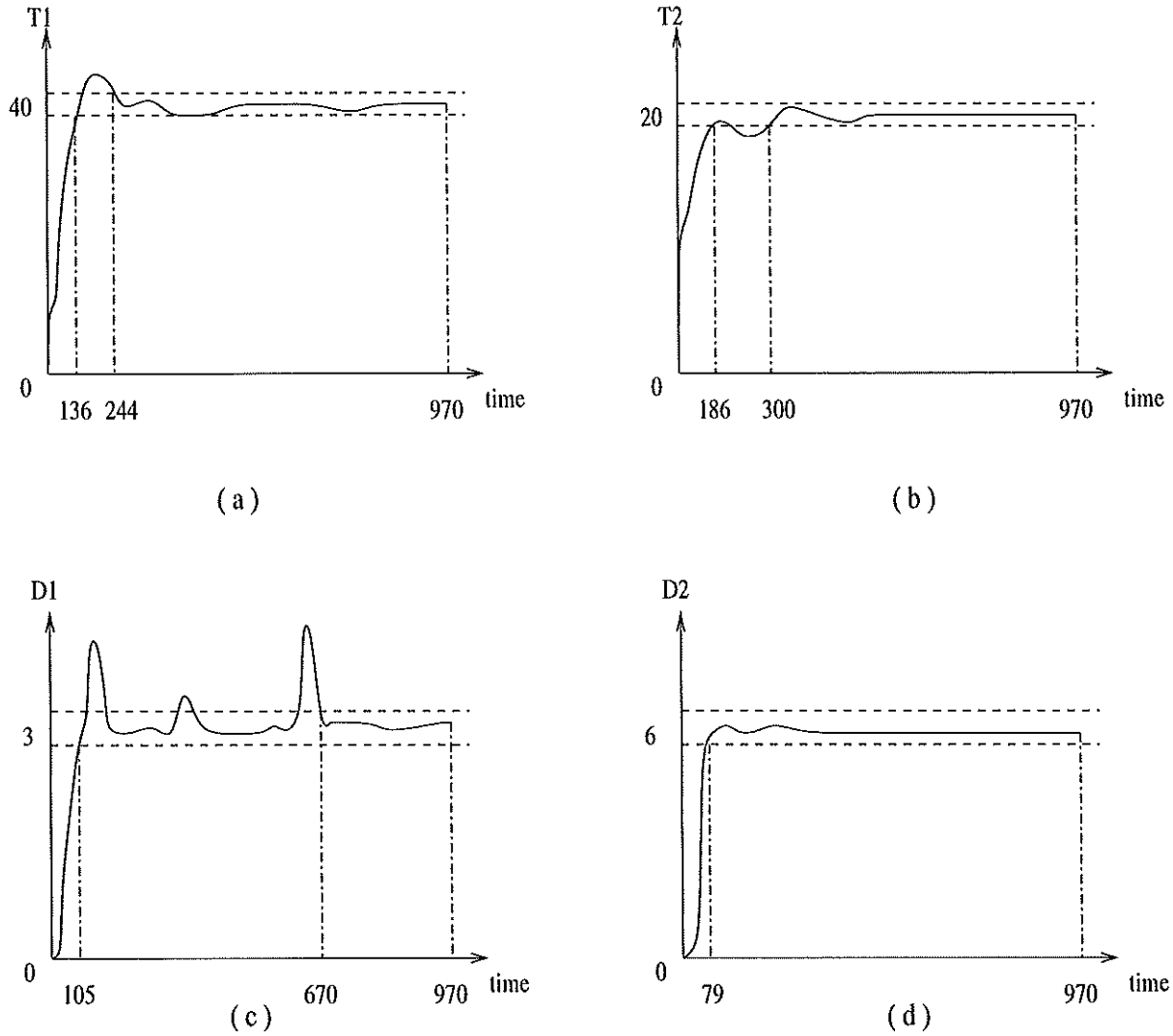

Figure 7: Graphs of goal variable value vs. time for each of four goal variables. Figure 7a is a graph of temperature in reservoir 1 vs. time, Figure 7b is graph of temperature in reservoir 2 vs. time, Figure 7c is a graph of outflow demand in reservoir 1 vs. time and Figure 7d is a graph of outflow demand in reservoir 2 vs. time.

Figure	Graph	PHASE 1	PHASE 2	PHASE 3
7a	Temperature in Reservoir 1 vs. Time	0-136	136-244	244-970
7b	Temperature in Reservoir 2 vs. Time	0-186	186-300	300-970
7c	Outflow demand in Reservoir 1 vs. Time	0-105	105-670	670-970
7d	Outflow demand in Reservoir 2 vs. Time	0-79	does not exist	79-970

Table 2: TTC, OT, and TTGB Regions for Figures 7a, 7b, 7c, and 7d.

The preceding example has used each individual goal variable as the unit of analysis. In this case, the duration of each of the three phases will usually be different for each goal variable. It is also possible to analyze the data in two other, less detailed ways. For example, if we group the temperature and demand goal variables for each reservoir together, then we can generate two graphs instead of four. In this subsystem view, the goal region is defined by a joint function of two variables for each reservoir. For instance, the duration of phase 1 will be determined by the earliest point at which both the temperature and the demand for a particular reservoir cross their lower goal regions. Similarly, we can also create a system view by defining the goal region as a function of all four goal variables. In this case, the goal region is defined by a four-dimensional vector. As a result, we get an even less detailed view, with only one graph for each subject for each trial.

Each of these three units of analysis has its advantages. Higher (i.e., less detailed) levels require less computation time and are easier to interpret because they produce less data as a result of the aggregation of dimensions. However, lower (i.e., more detailed) levels are potentially more sensitive and diagnostic. They can reveal patterns in the data that may be obscured at higher levels of analysis. We analyzed the data primarily at the individual and system levels in order to obtain a detailed understanding of the data set. However, in the analyses reported below, we will usually illustrate only system level analyses since these are easier to interpret. Additional insights obtained from individual level analyses will be described in less detail. The findings obtained for each of the three phases will now be described in turn.

Phase 1

Measures. Three measures were chosen to quantify performance in phase 1: time to contact (TTC), area under the TTC curve (ATTC), and rise time. TTC (Lee, 1976) is a dynamic variable

defined by the time currently needed for a goal variable to move from its current state to its lower boundary. TTC is a function of the distance between the current variable value and the goal state, as well as the rate of change of that distance (see Yu et al., 1997 for the detailed calculations). For the phase 1 of each trial, we can generate one (or more) graphs of TTC vs. time. As an example, Figure 8 shows the TTC curves for each goal variable for trial #1 for AV. Low values of TTC are favorable since they signify the rapid approach of the goal variable to the goal boundaries. When the goal is reached, TTC = 0 and phase 1 is terminated. Thus, we would expect that more proficient performers would have a faster and more direct path to the desired zero point.

This expectation can be examined more directly by looking at a second variable, ATTC, which is derived by calculating the area under the TTC graph for each subject. By computing ATTC, we can obtain one number for each trial. Consequently, we can succinctly summarize a subject's performance over the entire experiment by plotting ATTC vs. trial number. This is clearly more efficient and more informative than plotting TTC vs. time for each and every trial. We would expect that proficient subjects should have smaller ATTC values than less proficient subjects, and that a gradual decrease in ATTC should be exhibited over the six month period by all subjects.

The third variable we used to quantify phase 1 performance was rise time, which is simply the duration of the phase period. There are two advantages to this measure. First, it is simpler than either TTC or ATTC. Second, it adds information that cannot be obtained by looking at ATTC alone. A large ATTC can be obtained in at least one of two ways: by having a large value of TTC over a moderate period of time, or by having a smaller value of TTC over a longer period of time. ATTC alone cannot distinguish between these two cases. By examining rise time, we obtain a more diagnostic view of phase 1 performance. To summarize subjects' performance over the course of the experiment, we plotted the maximum rise time of the four goal variables against trial number. We would expect that better subjects to exhibit faster rise times, and that all subjects would improve with practice.

Results. Figure 9 shows graphs of ATTC vs. trial for each subject. As we would hope, all subjects improved their performance over the course of the experiment. Some subjects also appeared to become more consistent with practice, particularly AV. As with the measures presented earlier, AV and TL are the most proficient in their respective groups, exhibiting the

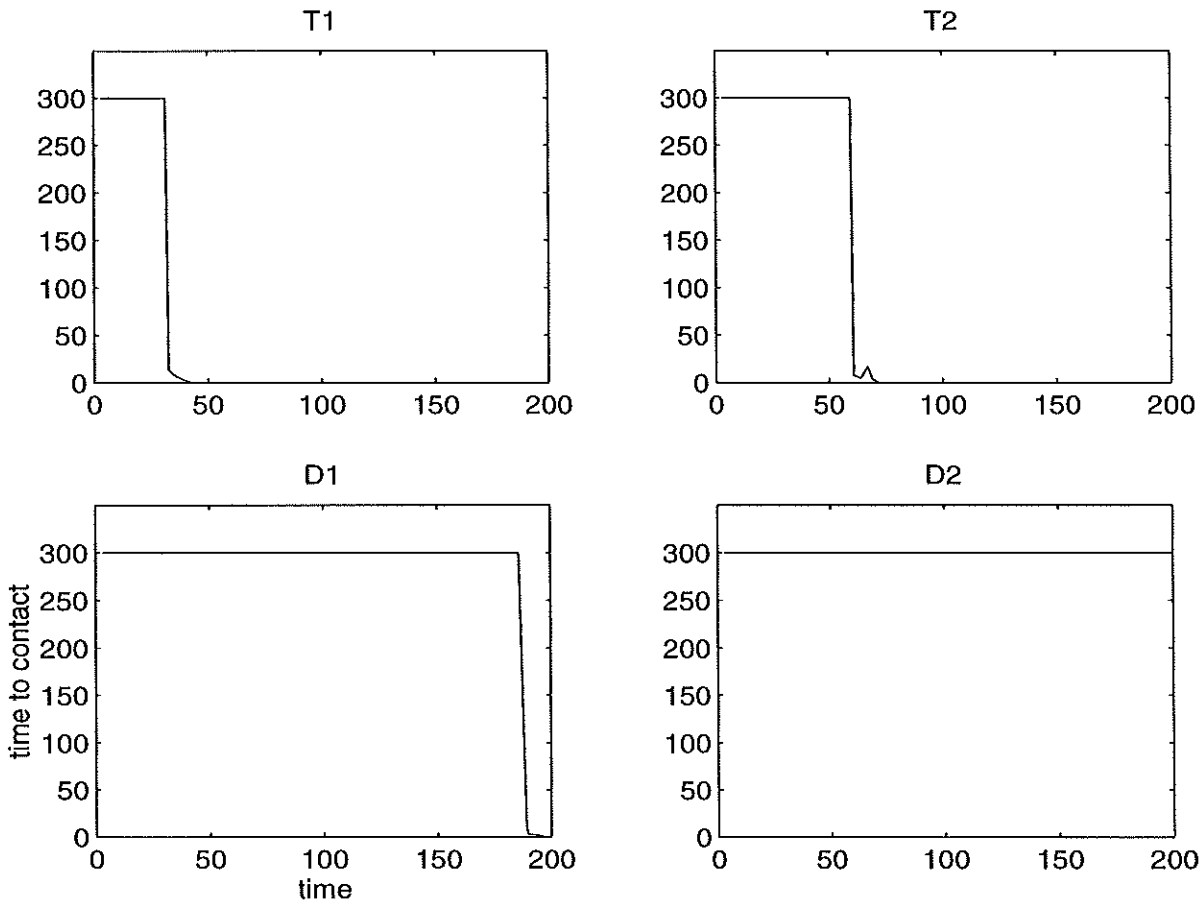

Figure 8: Time to contact of AV at the first successful trial (trial # 1)

lowest values of ATTC. Note, however, that AV is much more consistent than TL. Although less noticeable, it also seems that AV has lower ATTC values than TL, indicating a more efficient approach to the goal values. In general, the relative proficiency of the subjects on this measure is consistent with that observed earlier on other measures. As for group differences, it seems that the P group as a whole is less consistent than the P+F group in the latter part of the experiment.

The more detailed individual level analyses revealed an additional insight. AV was distinguished from all of the other subjects by his consistently low ATTC values for the output demand goals ($D1$, $D2$). This indicates that AV opened the output valves ($VO1$, $VO2$) sooner than the other subjects, which may explain why he consistently had the lowest ATTC values.

Figure 10 shows graphs of maximum rise time vs. trial for each subject. As expected, all subjects improved with experience. Most subjects also became more consistent over time. There do not seem to be any consistent interface differences. However, there are clear differences among subjects. AV and TL are clearly the best in their respective groups. This finding reinforces the results obtained with other measures (see above). AV seems to be the faster of the two. Again, the more detailed individual level analyses suggests that this speed advantage comes from reaching the demand goals ($D1$, $D2$) consistently faster than all other subjects.

Summary. The phase 1 analyses measure how quickly and how efficiently subjects are able to take to DURESS II from a shut down state into the goal regions. The findings confirm that AV and TL are the most proficient subjects in their groups, with AV being the fastest of all in this initial phase of the trial. Detailed analyses indicate that part of this advantage comes from the fact that AV consistently opens his output valves sooner than all other subjects, thereby reaching the demand goal setpoints more quickly. Also, it seems that the P+F group is more consistent than the P group.

Phase 2

Measures. Performance in phase 2 is more difficult to capture, so we used five different measures to characterize performance: oscillation time, number of oscillations, maximum deviation from goal region, maximum overshoot, and error area per unit time. Each measure was applied to each goal variable, resulting in four values. The first measure, oscillation time, is the simplest of all. It simply measures the duration of the phase 2 period. The second measure,

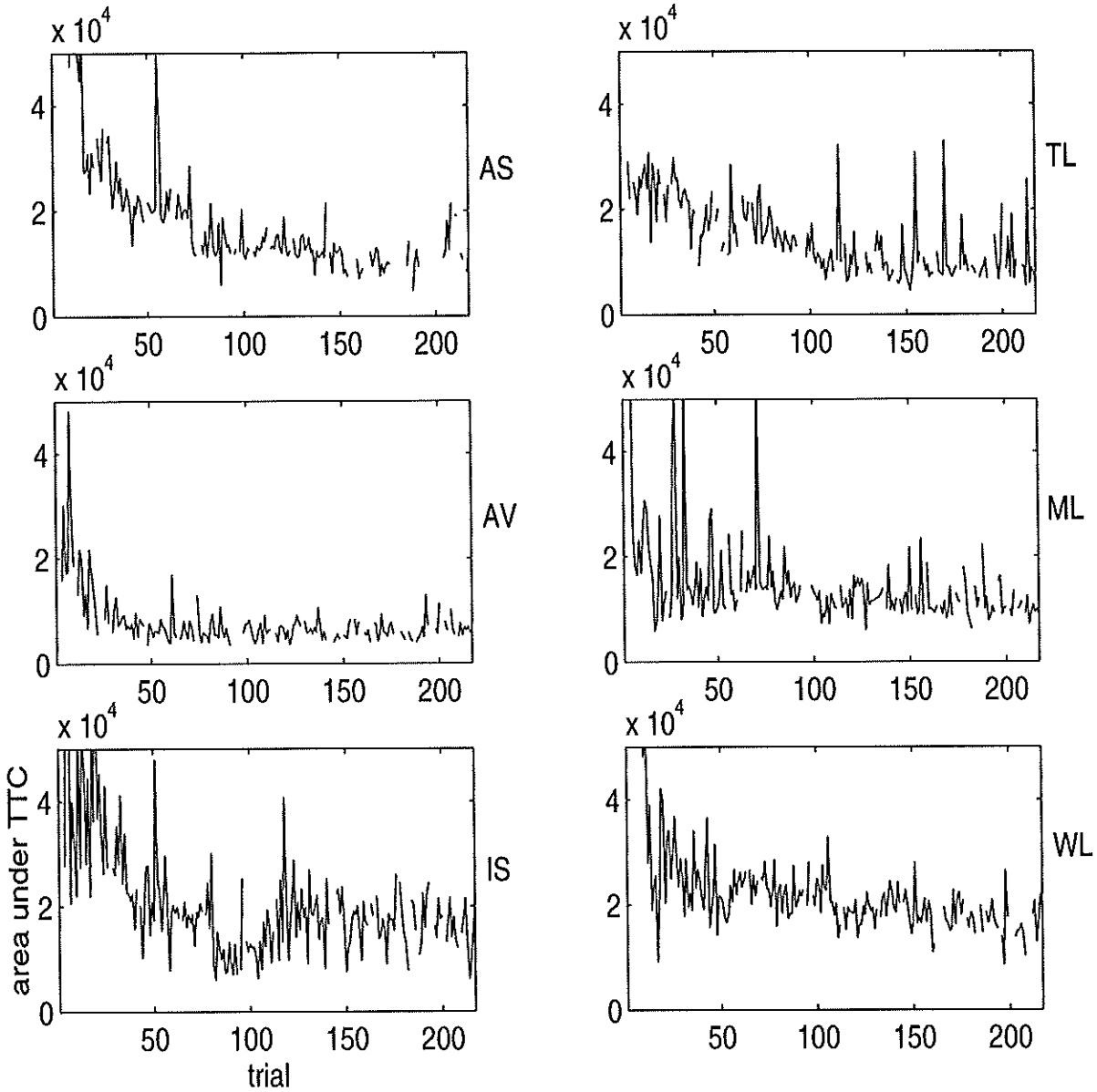

Figure 9: Area under curves of time to contact

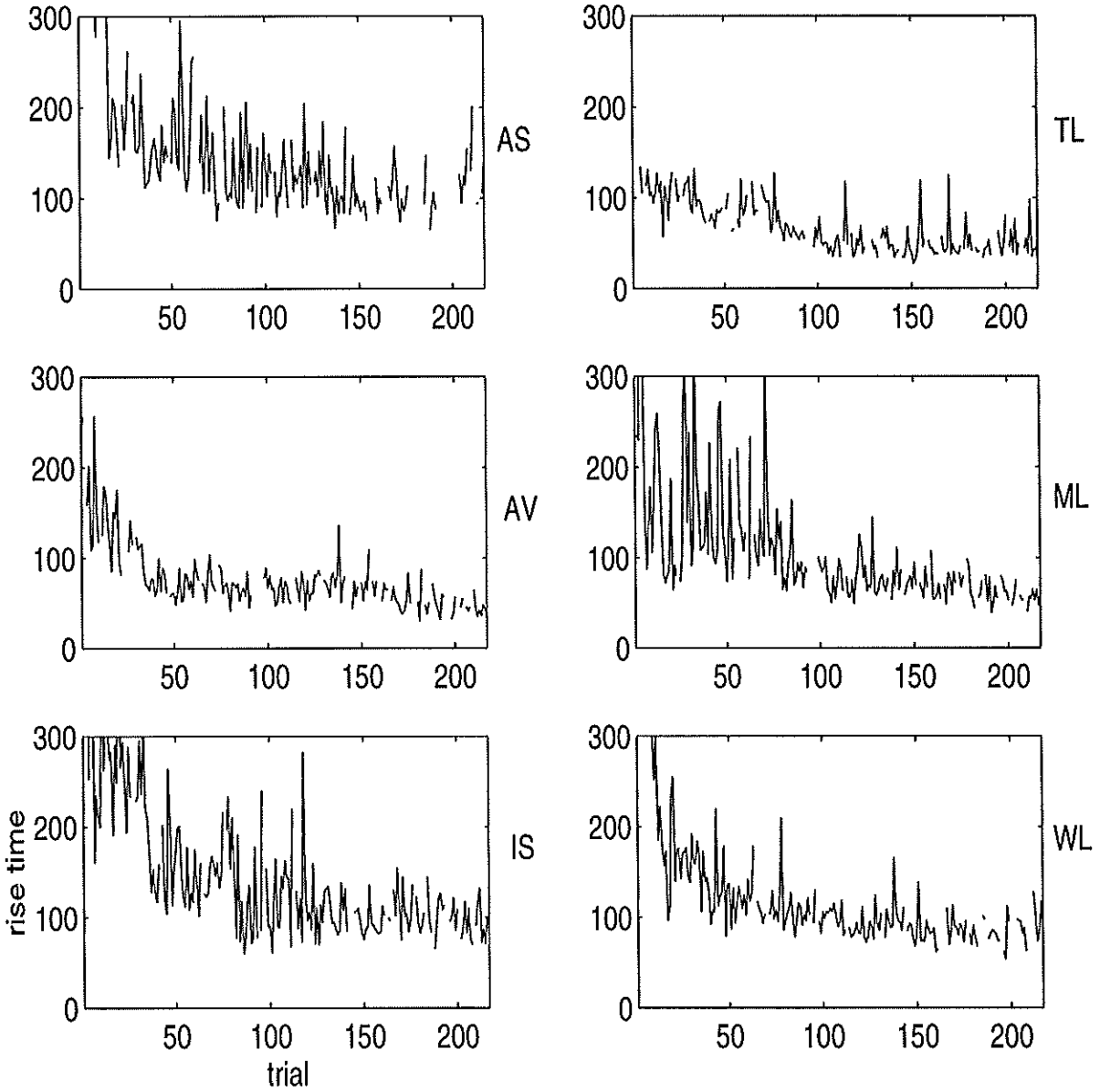

Figure 10: Rise time

number of oscillations, counts the total number of undershoots and overshoots for each goal variable during phase 2. The third measure, maximum deviation from goal region, represents the magnitude of the maximum deviation – whether it be an overshoot or an undershoot -- from the goal region as a percentage of the goal value. The fourth measure, maximum overshoot, is similar except that it only includes overshoots, not undershoots. Finally, the fifth measure, error area per unit time, is calculated by deriving the total area outside of the goal region during phase 2 (a measure of error) and dividing it by oscillation time.

Although these five measures are far from independent, they each emphasize a different aspect of performance. Together, they give us a relatively thorough insight into performance during phase 2. Because there are five different measures, and each applies to each of the four goal variables, we are faced with a very complex set of results. We decided to reduce this complexity by plotting only the worst of the four individual values for each measure. For example, the slowest of the four oscillation times was selected for each trial and then plotted against trial number. This data reduction decision greatly simplified the analysis. Note, however, that as a result of this decision, the data being plotted on each graph are not for the same goal variable. As before, we expect that these measures will show differences between individuals, trials, and perhaps interfaces.

Results. Figure 11 illustrates the results for the maximum oscillation time measure. Several results stand out. First, there is a tremendous amount of variability in the performance of most subjects. Second, there does seem to be an improvement over time, although this trend is more obvious for some subjects (e.g., TL) than for others (e.g., WL). Third, the P+F group seems to outperform the P group, although it is difficult to be sure, given the large variability. Fourth, by the end of the study, TL and IS seem to be the best performers in their respective groups on this measure. Fifth, TL's data undergo a pronounced improvement at around trial #100, suggesting a qualitative change in strategy.

Figure 12 shows the results for the maximum number of oscillations analyses. Again, several interesting results can be observed. First, the P+F group seems to outperform the P group, as evidence by the lower number of trials with 2 and 3 overshoots. Second, there is improvement with experience, but only for some subjects. For instance, IS and TL show clear signs of learning, frequently exhibiting no oscillations by the end of the experiment. In contrast, AV and WL show few signs of learning. Third, TL and IS are the best performers in their

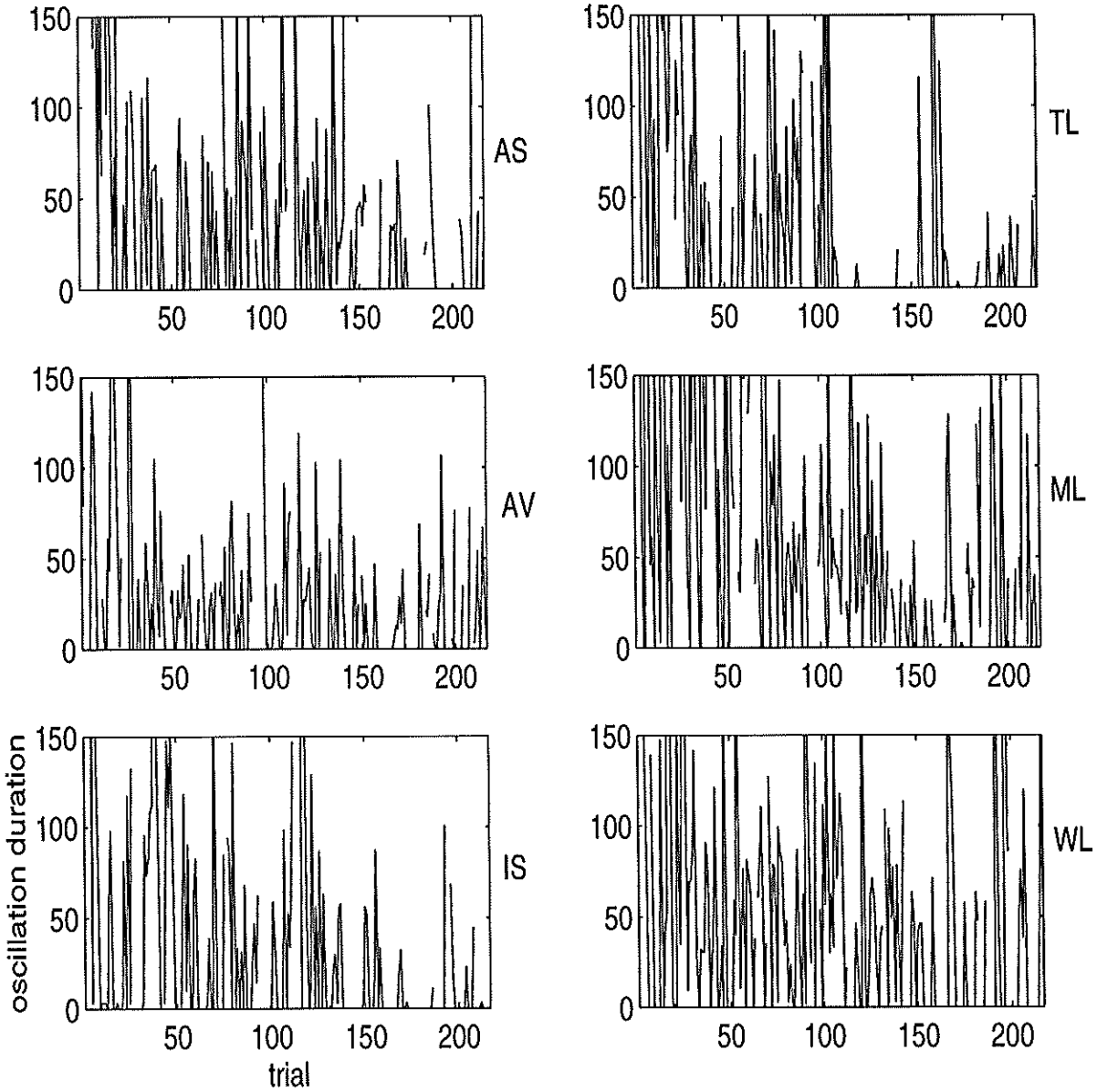

Figure 11: Maximum oscillation time

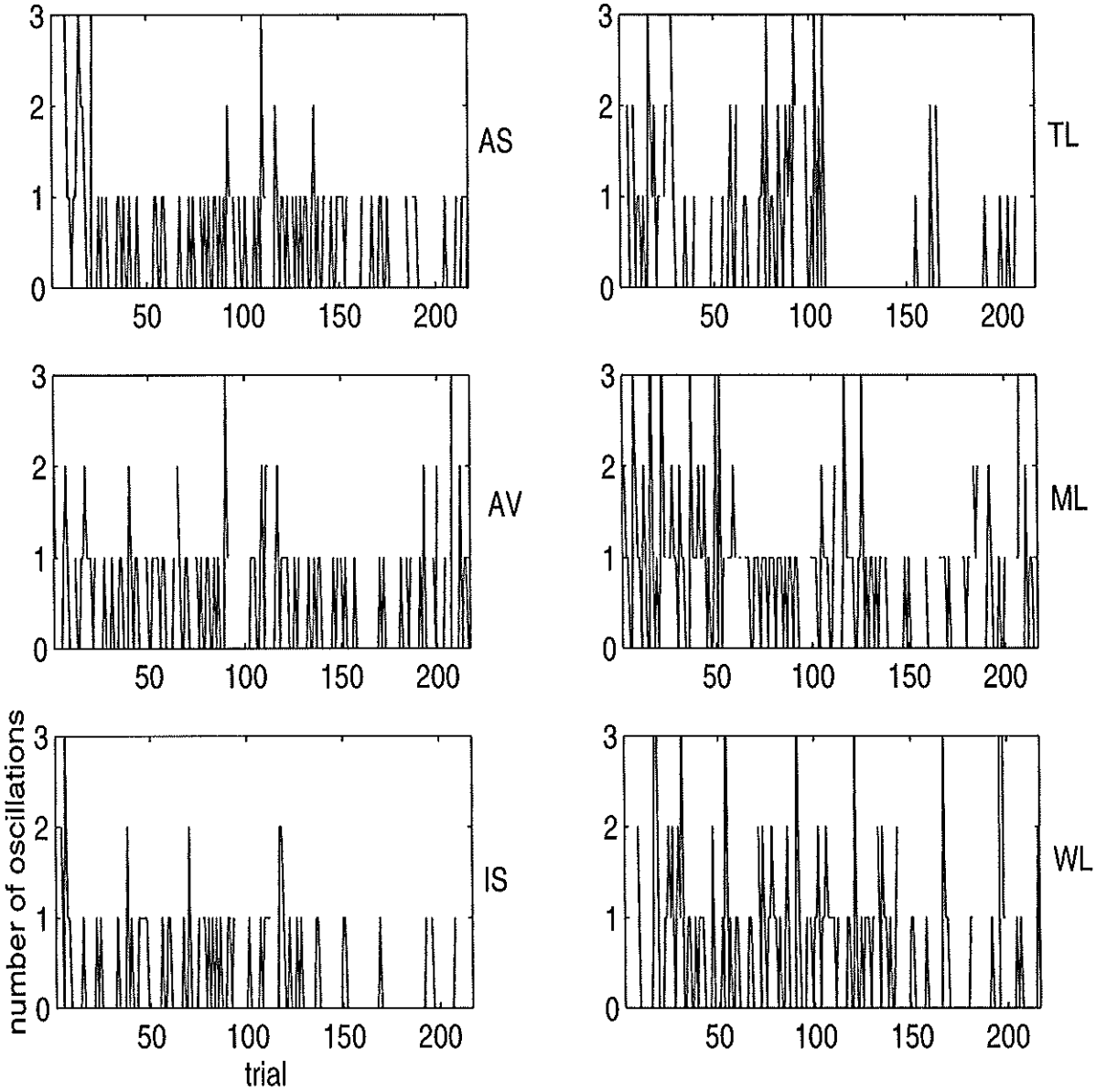

Figure 12: Maximum number of oscillations

respective groups by the end of the study, confirming the findings of the oscillation time analyses (Figure 11). Fourth, again we see that TL exhibits an abrupt improvement in performance starting at about trial #100.

Figure 13 illustrates the results for the maximum deviation from goal region (overshoot or undershoot) analyses. As with the oscillation time measures, there is a great deal of variability in the behaviour of all of the subjects, particularly ML. Because of this variability, it is difficult to discern any interface or practice effects. Some subjects (e.g., AS and TL) seem to improve over time, but others (e.g., ML) seem to get worse. As usual, TL seems to be the best in the P group by the end of the experiment. As with the other phase 2 measures, IS seems to be the best in P+F group although the differences between subjects in that group are not as obvious as in the P.

Figure 14 shows the results for the maximum overshoot measure. This measure is similar to the previous one, the only difference being that undershoots are not included in this case. Figure 14 shows that the overshoot data are much cleaner than the deviation data in Figure 13. Except for ML, the variability is much lower for all subjects. There seems to be a learning effect for most subjects, particularly AS and TL. Interestingly, as with Figures 11 and 12, we again see that TL seems to experience an improvement starting a little bit before trial #100. With the exception of one trial (around 150), his overshoots are consistently very low, barely showing up on the graph. In fact, he is clearly the best subject on this measure, not just in the P group, but overall as well. ML is clearly the worst. In the P+F group, it is difficult to discern differences between subjects. However, all three P+F subjects are much better than either ML or WL in the P group.

Figure 15 illustrates the findings for the error area per unit time analyses. The patterns in the data are similar to those on other measures. There is an improvement with practice for most subjects. TL and IS are clearly the best in their respective groups by the end of the experiment. ML is by far the worst overall, followed by WL. As with three other phase 2 measures, TL shows an abrupt improvement in performance just before trial #100, suggesting that he adopted a qualitatively different strategy for performing the task.

Summary. Phase 2 analyses measure subjects' ability to bring the system quickly and accurately under control after the initial trajectory into the goal region. Perhaps the most intriguing result we obtained was TL's marked performance improvement at around trial #100 that suggests he underwent a qualitative strategy shift at this point. Note that this shift was not

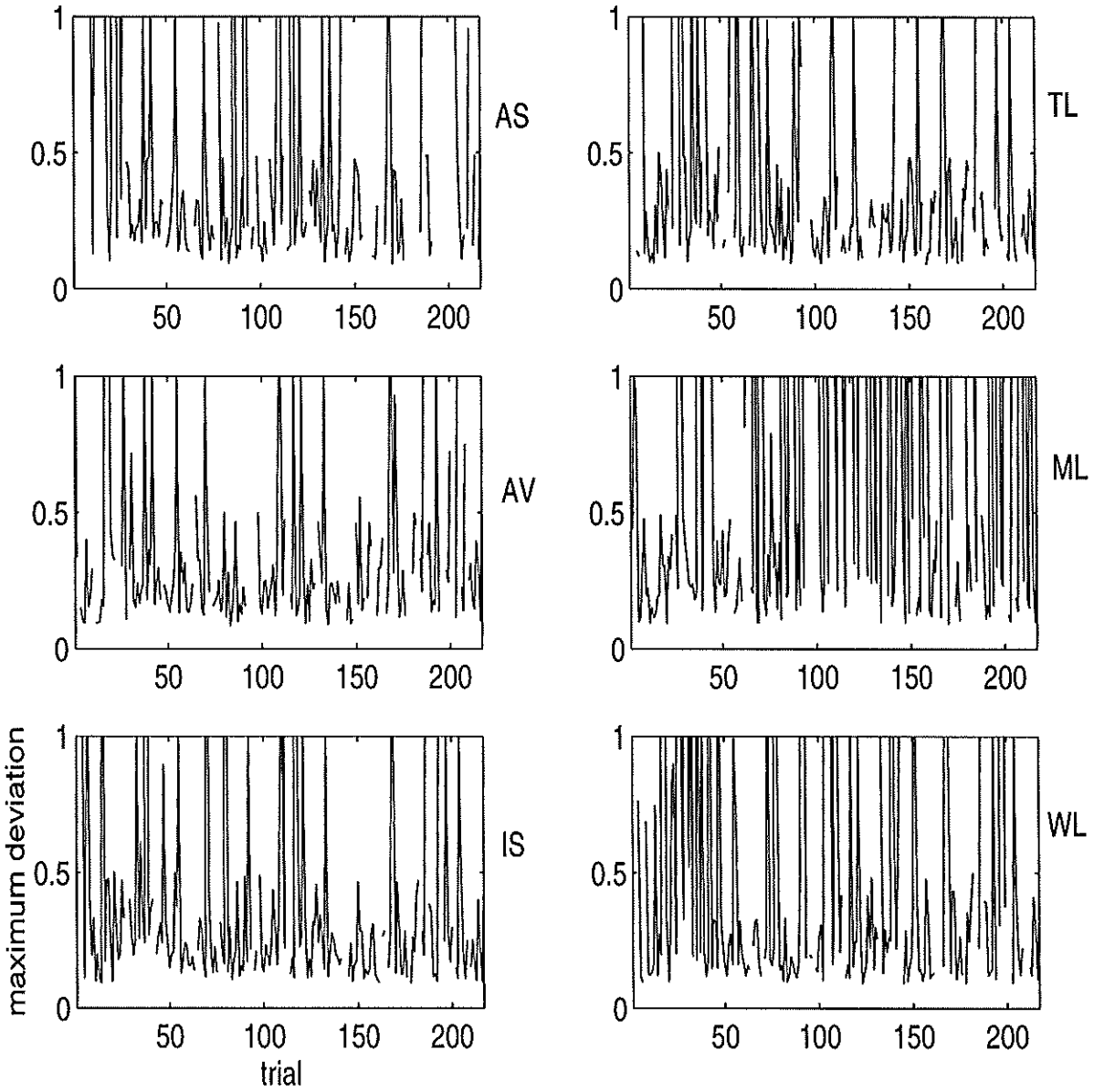

Figure 13: Maximum deviation

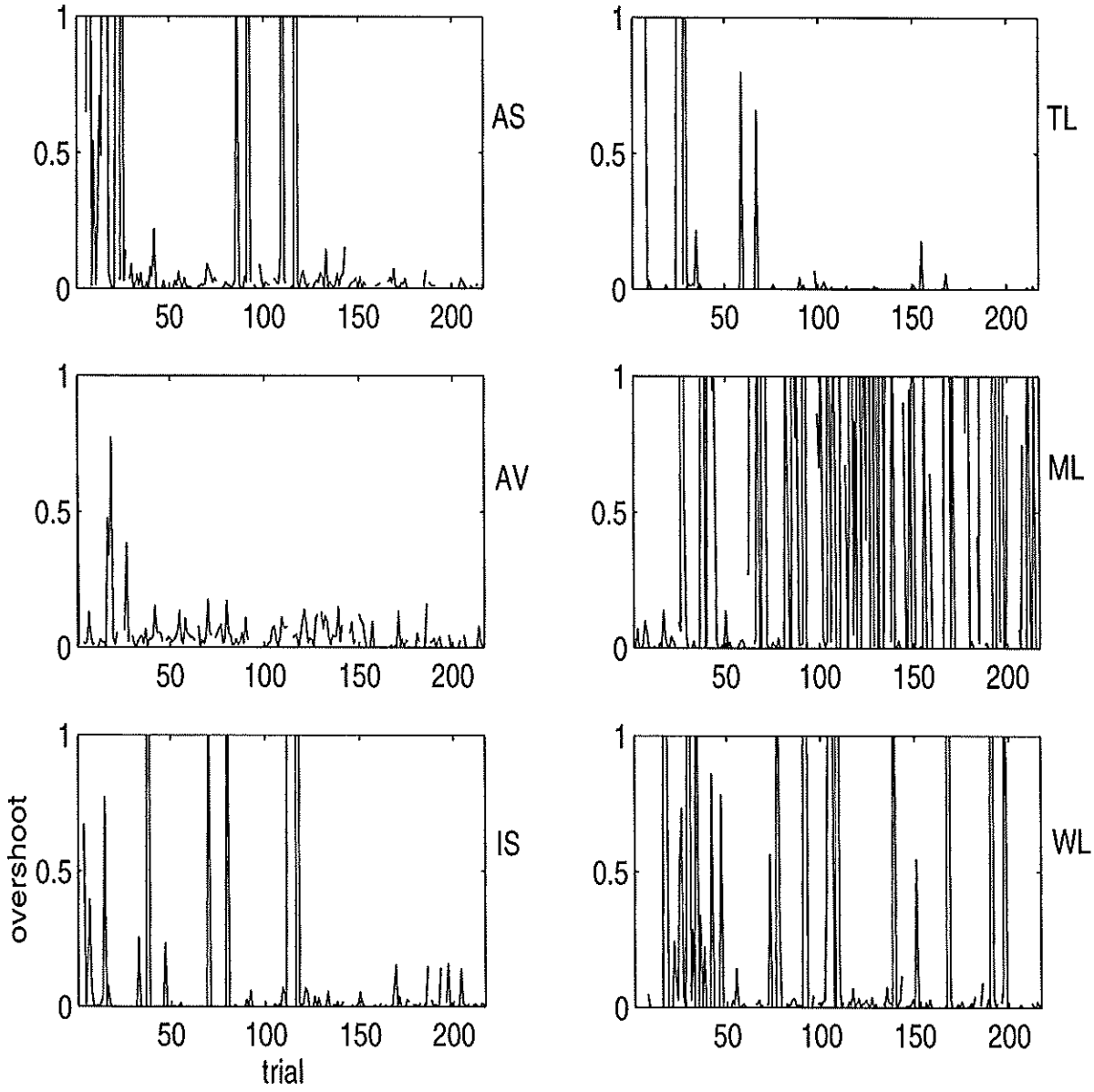

Figure 14: Maximum overshoot

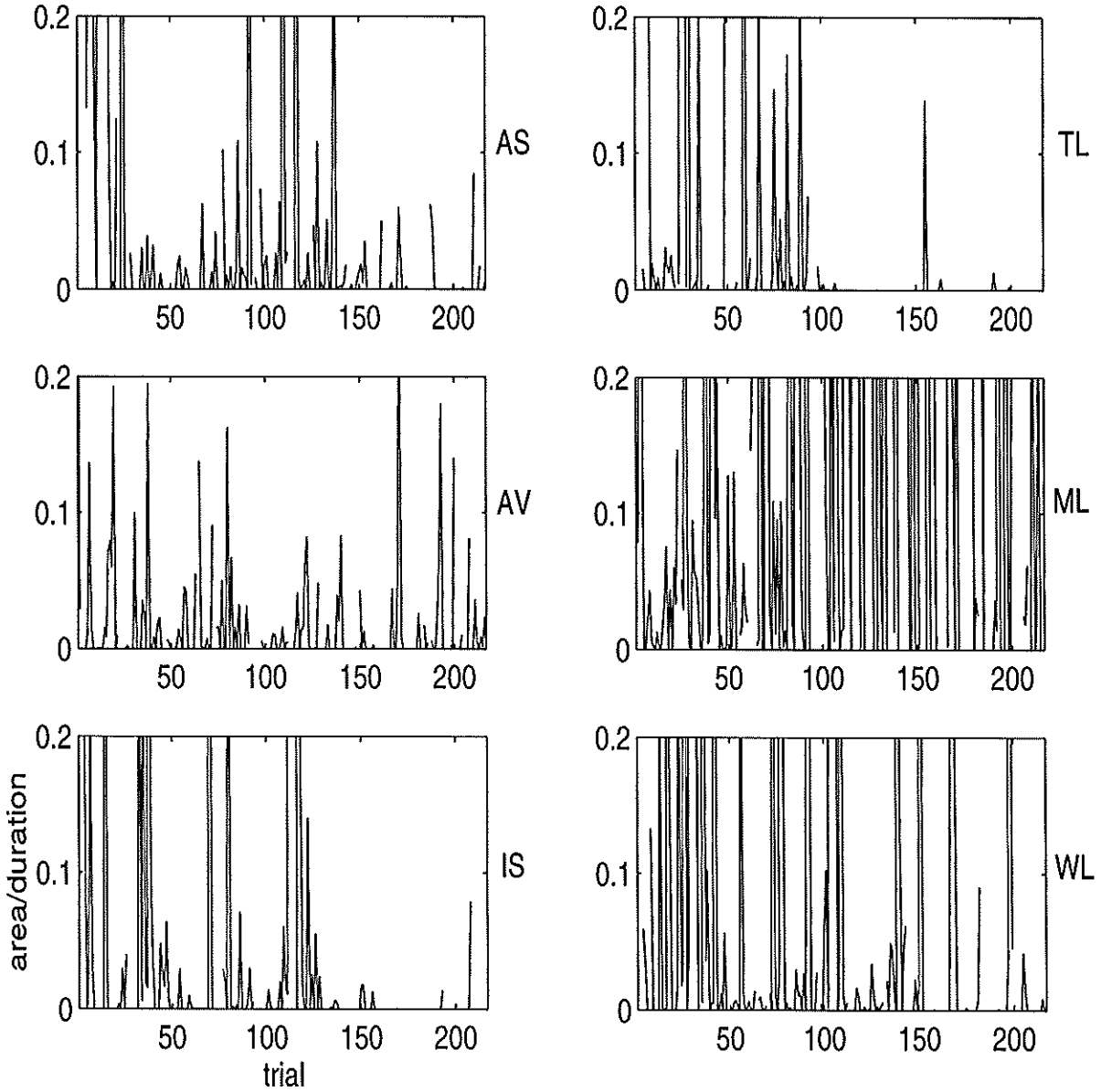

Figure 15: Error area per time unit

observed in phase 1. The significance of this result will be discussed in the following section on analyses of action variability.

The results also show that TL was the best in the P group, reinforcing earlier findings. Interestingly, however, IS replaced AV as the best in the P+F group. Thus, AV's strength seems to be in getting into the goal region quickly and efficiently (phase 1), whereas IS seems to be more proficient in terms of bringing the system under control (i.e., minimizing the number, duration, and size of oscillations in phase 2). This reversal may in fact represent a trade-off. If subjects bring up the system more slowly, as IS seems to do, then it is easier to bring it under control without overshooting. In contrast, if subjects bring the system up very quickly, as AV seems to do, then it is more likely that they will overshoot the target area. AV's trade-off is not necessarily a bad one if he can bring the system under control relatively quickly after overshooting.

To delve into this issue in more detail, we calculated the total transient time (i.e., the duration of phase 1 plus the duration of phase 2) for each subject for each trial. The results are presented in the graphs in Figure 16. AV and TL are clearly the fastest in their respective groups, although AV seems to be more consistent than TL. In fact, the P+F group as a whole seems to be more consistent than the P group. The discrete improvement in performance for TL shows up again, this time shortly after trial #100. Thus, even though AV is less proficient than IS in phase 2, he still comes out ahead when we look at the integration of phase 1 and 2 in terms of completion time.

Finally, there were also several signs in the phase 2 data to suggest that the P+F group was better than the P, especially in terms of consistency.

Phase 3

Measures. The primary concern in phase 3 is to keep the goal variables at least within the setpoint regions and preferably stable. Performance for this subtask can be measured by a variant of TTC. Since the system is already in the goal regions during phase 3, the goal boundaries can be viewed as repellors that must be avoided. Thus, we can quantify performance by measuring the time to reach the goal boundaries (TTGB), which can be considered as a measure of boundary avoidance. Because the boundaries are repellors, higher values of TTGB are preferred because they indicate that the system is temporally far from exceeding the goal boundaries. Since it is a

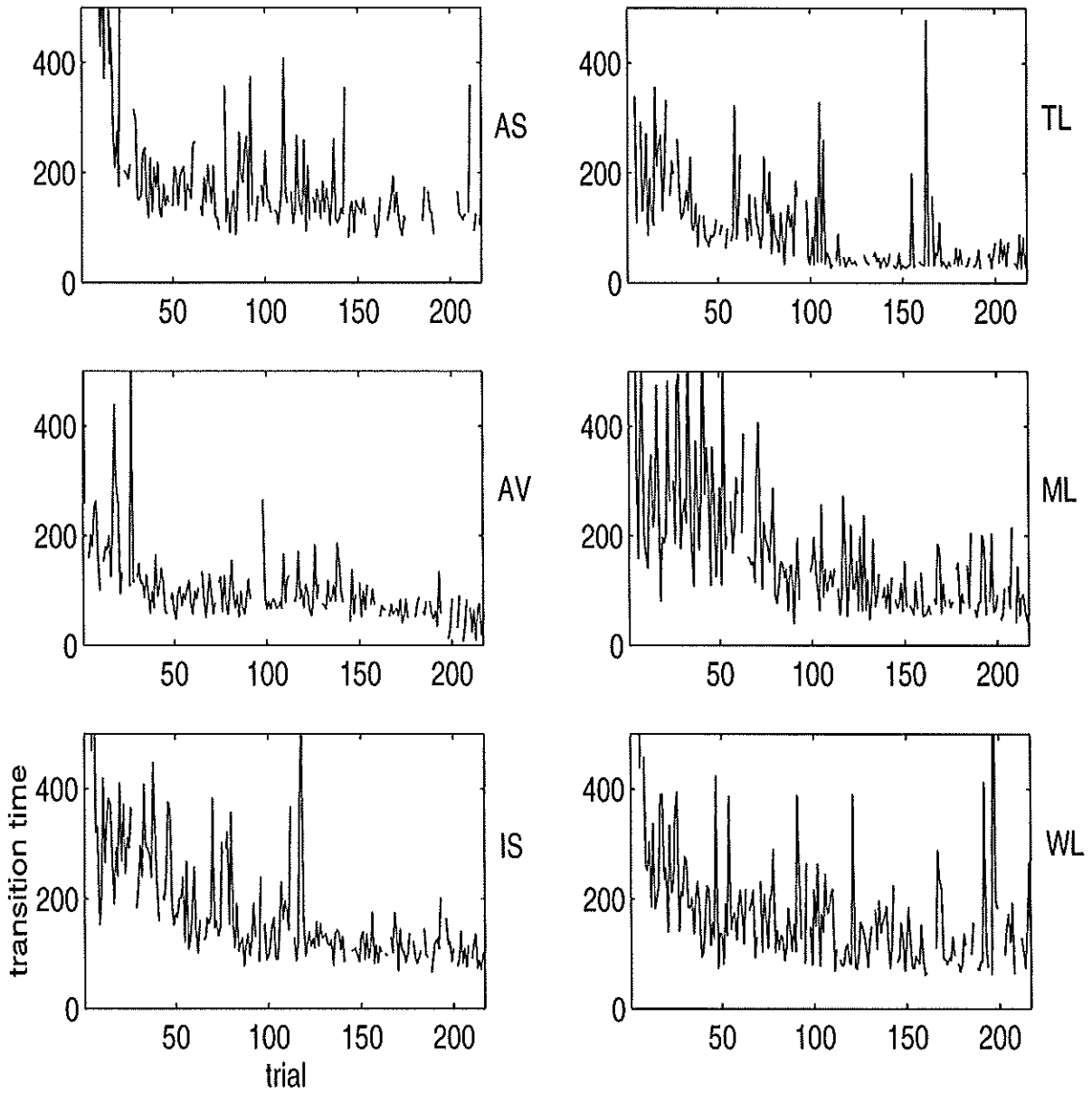

Figure 16: Transition time

dynamic variable, we can plot TTGB vs. time for each goal variable in each trial. An example is given in Figure 17 for trial #217 for AV. Since subjects were only required to keep the system in the goal state for 5 consecutive minutes, we used 300 s as a maximum value of TTGB.

Just as with TTC, we were able to summarize subjects' performance for the entire experiment by taking the area under the TTGB graph for each trial, creating the higher-order variable ATTGB. The most stable result possible would occur in the case where subjects are able to keep the TTGB at 300s for the entire duration of phase 3 (i.e., 300 s). Thus, we can use this limiting case to define a maximal value of ATTGB to normalize the observed values of ATTGB. Thus, a normalized ATTGB value of 1 represents maximally stable performance, whereas any value less than 1 indicates a lower level of stability (i.e., a smaller area under the TTGB graph). We expect subjects' performance to increase in stability with experience, and more proficient subjects to exhibit more stability than less proficient subjects (i.e., ATTGB values closer to 1).

Results. Figure 18 shows the results for the minimum ATTGB value among the four goal variables for each trial. Some subjects improved with experience (e.g., ML), whereas others seem to stay about the same (e.g., WL) or get worse (e.g., AS). TL seems to be the best in the P group, while WL, who had trouble dealing with the heater lags throughout the experiment (Christoffersen et al., 1994), was clearly the worst in the group. The differences in the P+F group were not as pronounced. AS was again the worst in the group, but there was no noticeable difference between AV and IS. Thus, the ATTGB measure does not seem to be as sensitive as some of the others we have examined.

There are at least three possible reasons for this result. First, phase 3 is defined as the time during which subjects keep a goal variable in the target area for 5 consecutive minutes. Thus, if there are notable problems in stability, they will have occurred in phase 2. For example, suppose a subject has a goal variable in its respective target region but has trouble stabilizing it and, after 4.5 minutes, leaves the goal boundary. As we have defined the 3 phases (see above), these 4.5 minutes would actually be part of phase 2, not phase 3. Second, it is possible that the subtask of keeping the goal variables within their target regions is not as difficult as the other subtasks, at least for most subjects. Third, it is also possible that different subjects adopted different performance criteria for phase 3. ATTGB measures stability but subjects need not keep all four goal variables perfectly stable in order to complete the task quickly. For example, in Figure 17

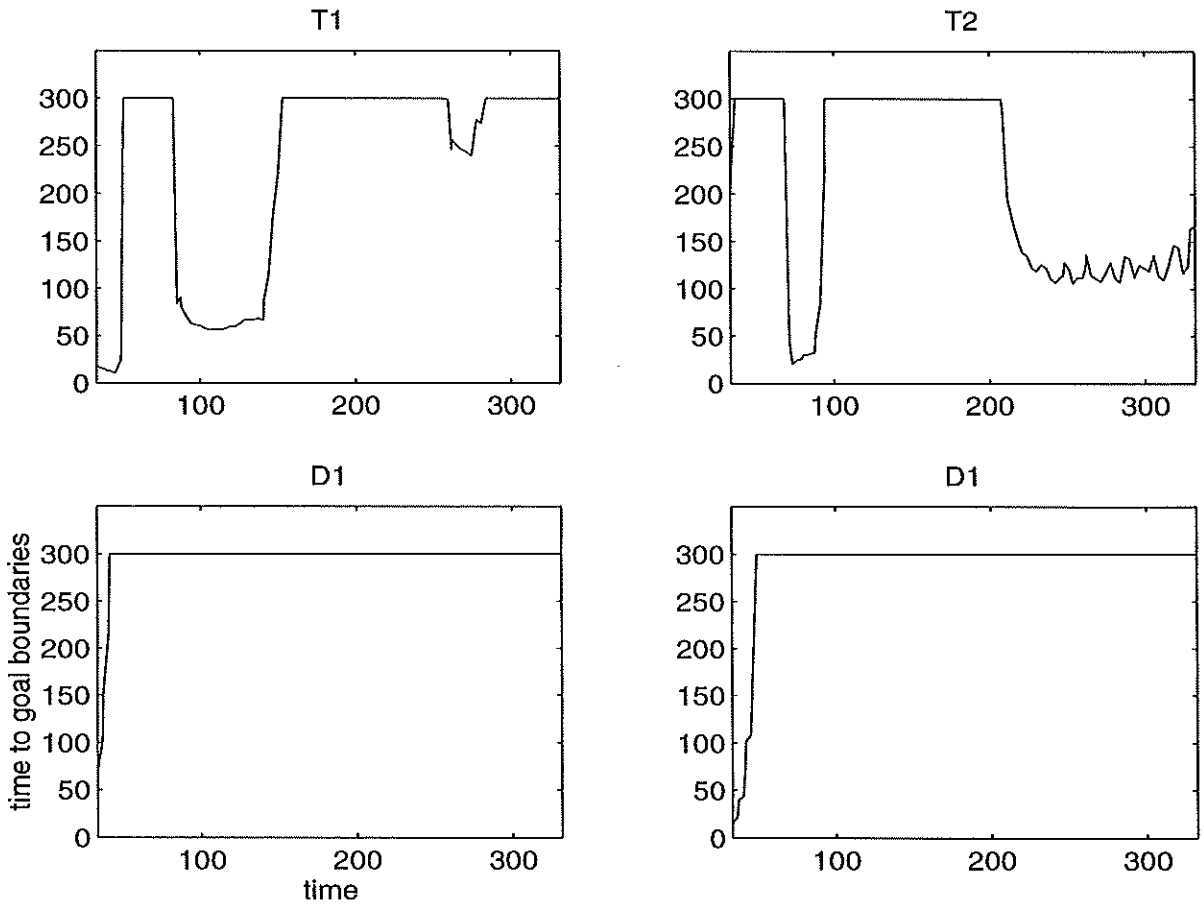

Figure 17: Time to goal boundaries of AV at the last normal trial (trial # 217)

we see that for AV, *T*2 was not as stable as it could be towards the end of this trial. Yet, it was stable enough for him to satisfy the 5 minute steady state criterion. On the one hand, we can say that AV performed the task effectively, but if we look at the ATTGB value for that trial, we would see that it would be far from the ideal. Interestingly, the trial depicted in Figure 17 occurred very late in the experiment, so the lack of perfect stability in *T*2 was not due to a lack of experience.

More detailed analyses of ATTGB for each individual goal variable indicated that stability was better for the output demand goals than the temperature goals. This is not surprising since the mass flow dynamics are faster, and thus easier to stabilize, than the temperature dynamics.

Discussion

The results presented above are the most detailed analyses we have conducted of performance in this six-month longitudinal study. They have confirmed the major findings we had obtained in earlier analyses of this data set (Christoffersen et al., 1994). For example, as we would expect, subjects' performance improved with experience on virtually every measure we examined. Furthermore, the P+F group seems to have an advantage over the P group on some performance measures, particularly in terms of consistency. In addition, these analyses confirm that TL is clearly the best performer in the P group and that AV is generally the best in the P+F group, with IS a close second.

More importantly, these performance analyses have also uncovered new insights that we had not documented before. For example, the superiority of TL can be seen on almost every measure across each phase of the trial. In contrast, AV's strength seems to be in the initial approach phase of the trial (i.e., phase 1). He is the fastest of all subjects in this phase. One reason for this superiority is that AV opens his output valves earlier than other subjects, thereby attaining the output demand goals sooner. Perhaps as a result of this aggressive initial approach, AV does not do as well as IS during phase 2 of the trial where the primary concern is to bring the system under control. IS outperforms AV during this phase of the trial, exhibiting less oscillations. Because IS was less aggressive in phase 1, it was easier for him to gain control of the system once he reached the goal setpoints. Interestingly, these results confirm the subjective impressions we reported in our original contract report, where we stated that AV seemed more aggressive and IS more cautious (Christoffersen et al., 1994). If we look at phases 1 and 2 together, AV seems to be the best subject overall. The other new insight we observed was that

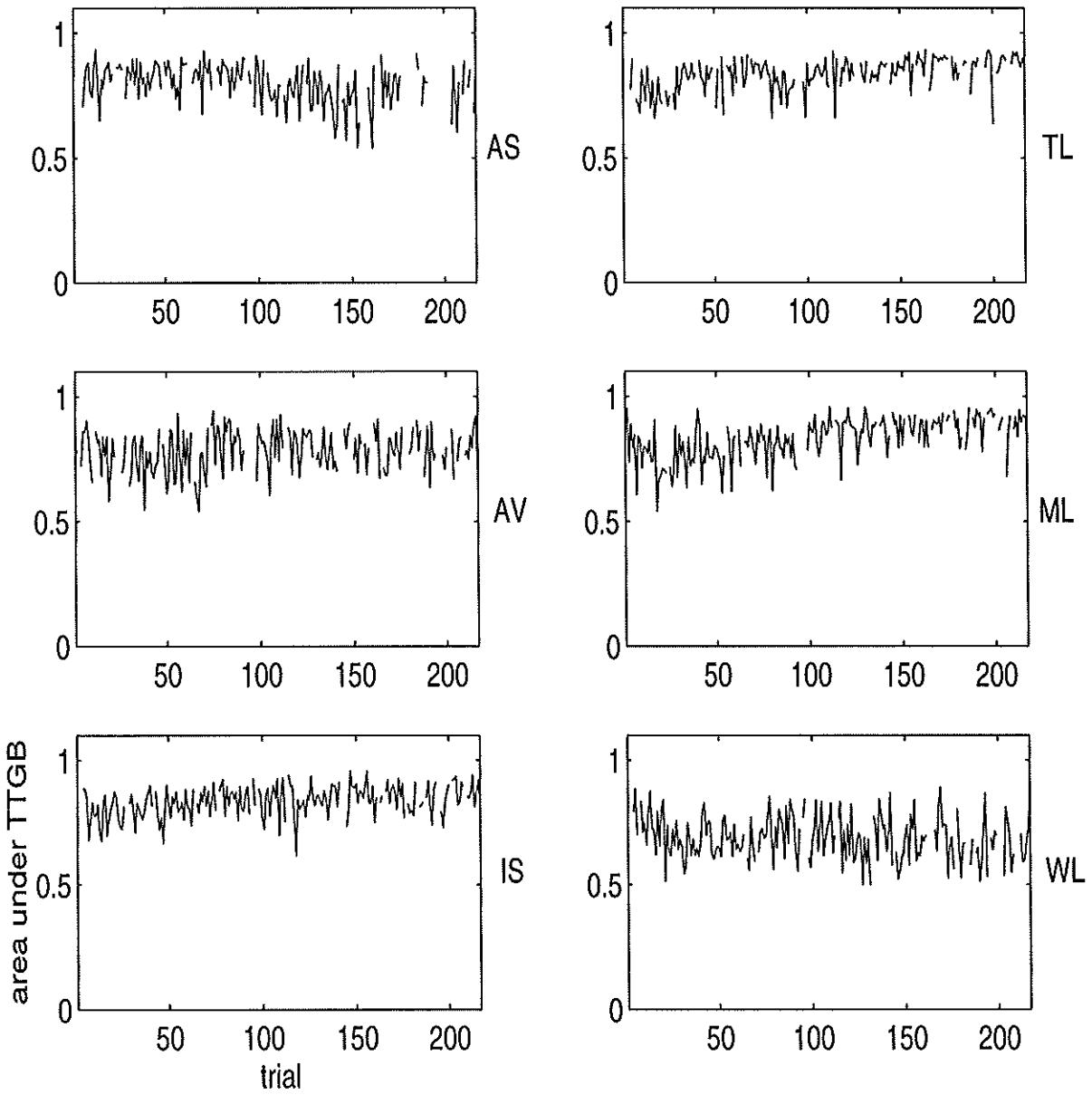

Figure 18: Normalized area under TTGB w.r.t maximum area

TL exhibited a marked improvement in performance on phase 2 measures starting at about trial #100. The change was so noticeable that it is highly likely that TL started using a qualitatively different strategy that allowed him to bring the system under control more efficiently and effectively than before. The significance of this change will be discussed in the following section.

These new insights show that there is value to the novel measures of performance we have adopted for this research. They have provided a more convincing picture of the differences between subjects, as well as a more precise understanding of the nature of those differences. In the remainder of this report, we will present additional insights which try to explain the origin of the results reported in this section.

4. ANALYSES OF ACTION VARIABILITY

In volume 1, we briefly discussed the relevance of measures of operator action variability in acquiring a better understanding of operator adaptation. If participants' behaviour is being driven by some type of procedure (or "script"), then we would expect their actions to be comparatively less variable since they are stereotyped by that procedure. The action-based form of control reported by TL, the best subject in the P group, provides a good example. Because TL reported putting certain components on specific settings (see Table 3), we expect that his actions would be relatively consistent across trials. In contrast, if participants' actions are being driven instead by the goals to be achieved, then we would expect their actions to be comparatively more variable since the demand goals are different for different trials. The function-based form of control reported by AV, the best subject in the P+F group, provides a good example. Because AV reports setting the components to whatever settings were required to achieve the current goals (see Table 4), we expect that his actions will be relatively variable across trials.

While these hypotheses follow directly from the control recipe data in Tables 3 and 4 (described in much more detail in Christoffersen et al., 1994), it is not that obvious how these predictions can be put to an empirical test. In this section, we use three different types of measures in our search for a viable way of quantifying the variability in operators' actions.

Action Frequency Distribution Graphs

A very coarse way of determining whether a subject's actions are being driven by a specific procedure, as TL reported doing, is to examine the frequency with which that subject chooses

specific settings for particular components in DURESS II. Such an analysis can be presented in a graphical format by plotting a distribution of frequency vs. component setting for each component for a particular block of trials. By comparing the procedures subjects reported using with the action frequency distribution (AFD) graphs, we can determine whether what subjects say they did corresponds to what they actually did. For instance, if a subject says that part of his procedure involves putting component *VA* on 7, then we would expect to see a peak around the value 7 in the AFD for that particular component.

The 12 AFD graphs representing the actions during the last block of trials for TL are presented in Figure 19. Table 3 reproduces the control recipe that TL wrote around the same time, explaining the strategy he used for controlling the system during a startup task (AV's recipe is shown in Table 4 for comparison). We see that there is a close correspondence between some of the steps in the control recipe and the peaks in the corresponding AFDs. The most pronounced agreements are found in the AFDs for components *VA*, *VA*1, *VB*, *VB*2, *HTR*1, and *HTR*2. For example, in step 1) TL writes, "open valves *VB*2 and *VB* to their maximum (10) and start *PB*. Set *HTR*2 to 4.5." Examination of the graphs for components *VB*2, *VB*, *PB*, and *HTR*2 reveals noticeably higher action frequencies at the settings stated by TL. Similarly, in step 2) of his control recipe TL states, "set *VA*1 to 7, *VA* to maximum (10)". Examination of the AFD graphs of the *VA*1 and *VA* components reveals peaks at 7 and 10 respectively, again supporting the hypothesis that there is a correspondence between TL's control recipe and his AFD graphs. In step 3), TL states "as the temperature in Res 2 reaches to lower end of the desired range, set *HTR*2 to 3 1/3". As anticipated, a strong peak at 3 1/3 occurs in the AFD graph for *HTR*2.

The AFD graphs also allow us to pursue another hypothesis that had been generated during our previous analyses of subjects' control recipes (Christoffersen et al., 1994). Although TL ended the six-month experiment reporting an action-based strategy for controlling DURESS II, his earlier control recipes suggest that he originally approached the task in quite a different fashion. Like all of the other subjects in this study, TL was asked to provide control recipes at 5 different points in the experiment after he had started controlling DURESS II. Table 5 shows a subset of the analyses of these data. As we can see, there was a distinct change in TL's control recipe responses over the course of the experiment. In the first part of the experiment, TL provided several statements in each control recipe justifying or explaining recipe steps. This is noteworthy because subjects were not instructed to explain their steps, but were merely asked to

Open valves VB2 and VB to their maximum (10) and start PB. Set HTR2 to 4.5. Set VA1 to 7, VA to maximum (10), and start PA. Set HTR1 to 10. As the temperature in Res 2 reaches to lower end of the desired range, set HTR2 to 3 1/3; then set VO2 to desired level. Set VO1 to desired level. As the temperature in Res 1 reaches the lower end of the desired temperature range, set VA1 to level 10. As the resevoirs reaches level 60, adjust VA1 to equal VO1 and VB2 to equal VO2. Modify heater settings as necessary. Hint: HTR1 should be at same numerical setting as VA1, and HTR2 at 1/3 of VB2. (It might not make sense, but it works). IF THE DEMAND ON RES 2 IS GREATER THAN 10. Follow steps 1), 2) and 3) above. 4) Set VA2 such that the total water input into Res 2 is one level above VO2. Set VA1 such that VA1 + VA2 = 10. Set heater levels as defined in the hint section above. 5) As the resevoirs reaches level 60, adjust values so that VA1 = VO1 and VA2 + VB2 = VO2. Modify heater settings as required.

Table 3: Last control recipe for TL on P interface (experimenter notes are in brace brackets).

list those steps. Thus, the fact that TL took the initiative to provide unsolicited justifications and explanations for his recipe steps suggests that he was deliberately and spontaneously trying to understand <u>why</u> he was controlling the system in the way that he was. In the latter part of the experiment, however, we see a very different pattern. The number of justifications of explanations drops to zero, and the number of statements specifying a precise quantitative value for a particular component setting increased dramatically. This marked change suggests that TL gave up trying to understand the basis for his actions, and instead, focused on developing a rote set of specific actions that allow him to control the system efficiently (see Table 3). This interpretation is bolstered by TL's statement, late in the experiment, about one of his quantitative recipe steps: "It might not make sense, but it works"! Table 5 suggests that this change from focusing on understanding to focusing on a very specific web of rote procedures occurred some

Turn PA & PB ON
Turn VA & VA1 ON (MAX)
Turn HTR1 ON (MAX)
Turn VB & VB1 & VA2 (if necessary) {arrow underneath text, from parenthetical phrase to VA2}
Turn HTR2 ON (MAX)
Adjust VO1 & VO2 TO OUTPUT
Adjust VA & VA1, VB & VB1 & VA2 to
necessary input.
Adjust HTR1 & HTR2 For the INPUTS
DO SOME FINE TOUNING.END

Table 4: Last control recipe for AV on P+F interface (experimenter notes are in brace brackets).

Recipe #	Experience	Explanatory	Quantitative
1	21	3	1
2	67	2	2
3	150	0	6
4	217	0	5
5	223	0	5

Table 5: The change in control recipe response pattern for TL with experience. Recipe # indicates each administration of the control recipe knowledge elicitation test. Experience indicates the cumulative number of trials that TL had experienced by each test administration. Explanatory indicates the number of statements in each control recipe justifying or explaining recipe steps. Quantitative indicates the number of statements in each control recipe specifying a precise quantitative value for a particular component setting.

time between trial 67 and 150. Thus, TL seemed to exhibit a qualitative shift in the way in which he controlled the system over the course of the six-month experiment.

By examining the AFDs for TL during this transition period, we can see if there is any behavioral evidence of a qualitative shift in control strategy. We divided the data from the entire experiment into 10 blocks of approximately 22 trials each, and plotted AFD graphs for each block. Figures 20 and 21 show the data from blocks 4 (trial # 67 to 88) and 5 (trial # 89 to 110), respectively. If we compare these two Figures with TL's final control recipe in Table 3, we see

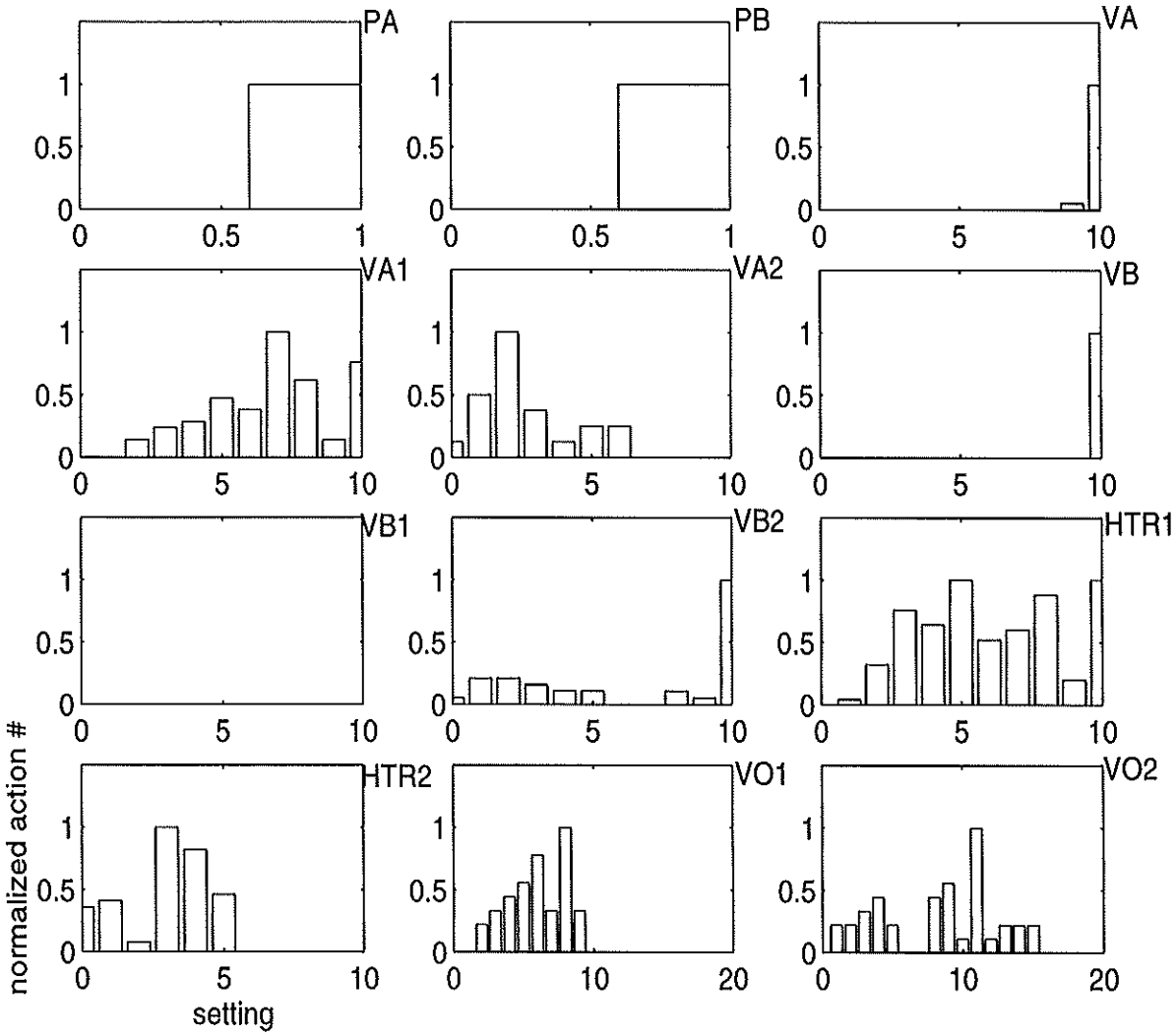

Figure 19: Normalized number of actions vs. settings at the last block of TL

that there are several signs in the data from block 5 that are consistent with this final recipe and that are not observed in block 4, namely:

- *VA* is always set to 10
- we start to see a peak in *VA*1 at a value of 7
- *VB* is always set to 10
- *VB*1 is not used at all
- *HTR*1 has a peak at 10
- the two largest peaks for *HTR*2 are at about 4.5 and 3 ⅓

There are three important features about the behavioral characteristics in this list. First, they are all observed for the first time in block 5 (Figure 21). Not only do they not appear in block 4 (Figure 20), but they also do not appear in any earlier blocks either. Second, almost all of these features are consistently observed in subsequent blocks all the way up until the end of the experiment (Figure 19). In some cases (e.g., the peak in *VA*1 at 7, and the peak in *HTR*2 at 3 ⅓), the characteristics are even more pronounced in later blocks. Third, all of these behavioral characteristics are consistent with the strategy that TL winds up using to control DURESS II by the end of the experiment (see Table 3).

This marked transition between the AFD graphs for blocks 4 and 5 provides behavioral evidence that is consistent with the conjecture derived from the verbal reports summarized in Table 5. The control recipes suggest that TL underwent a qualitative strategy shift, from trying to understand to just developing procedures that work, some time between trials 67 and 150. The data from the AFDs provide objective support for this inference, and in fact, allow us to narrow down the time of the strategy shift. By comparing Figures 20 and 21, we can see that TL probably gave up understanding the basis for his actions somewhere around trial #88, and subsequently focused on developing specific procedures that could allow him to control DURESS II efficiently. Clear evidence of this procedure can be observed from trial #89 onward. The relationship between these findings and the results from the phase 2 performance analyses presented earlier will be discussed at the end of this section.

Information Theory

The previous subsection provided some insight into operator action variability, albeit in a relatively informal manner. The AFD graphs allowed us to make inferences about the extent to

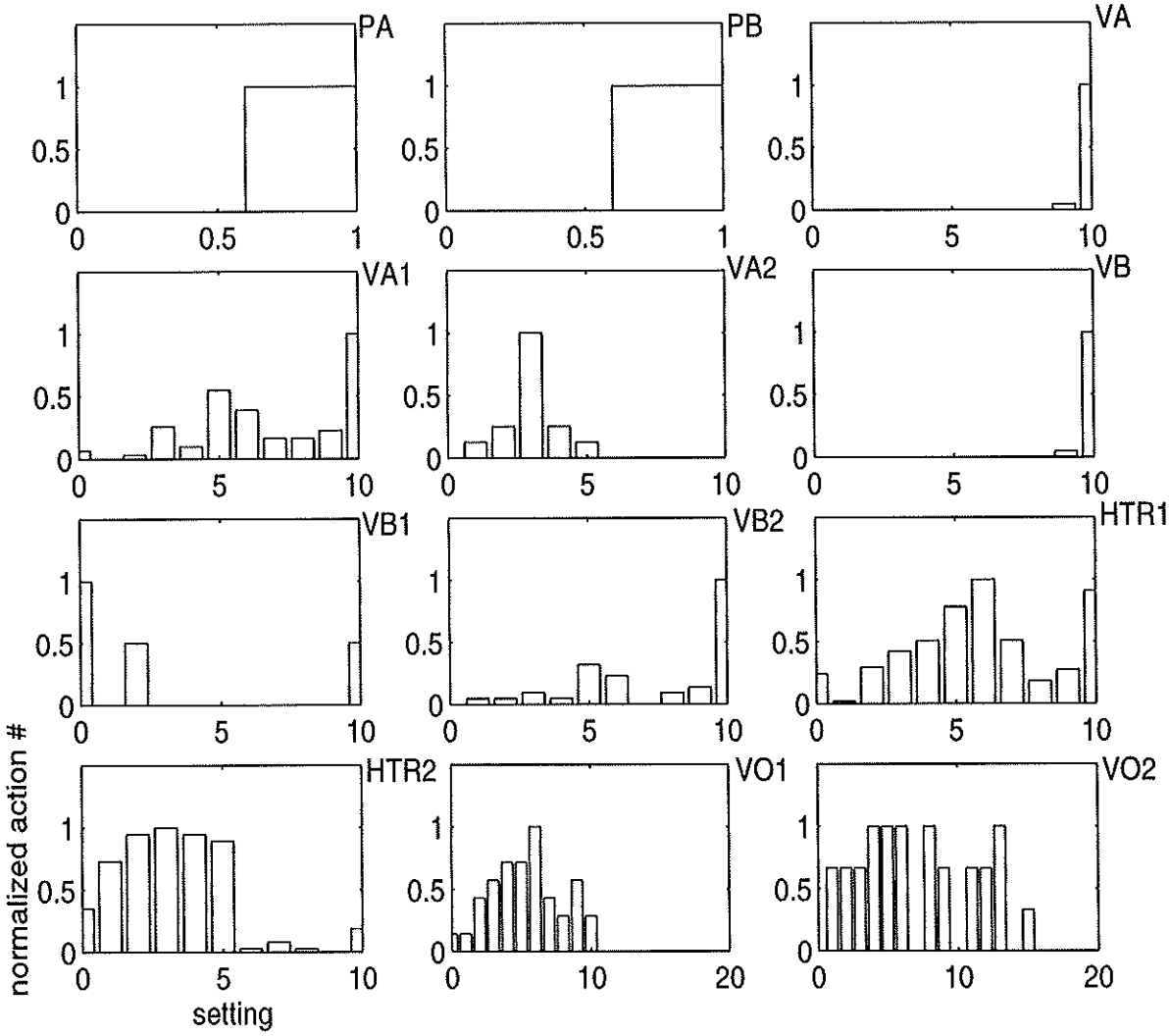

Figure 20: Normalized number of actions vs. settings at block # 4 of TL

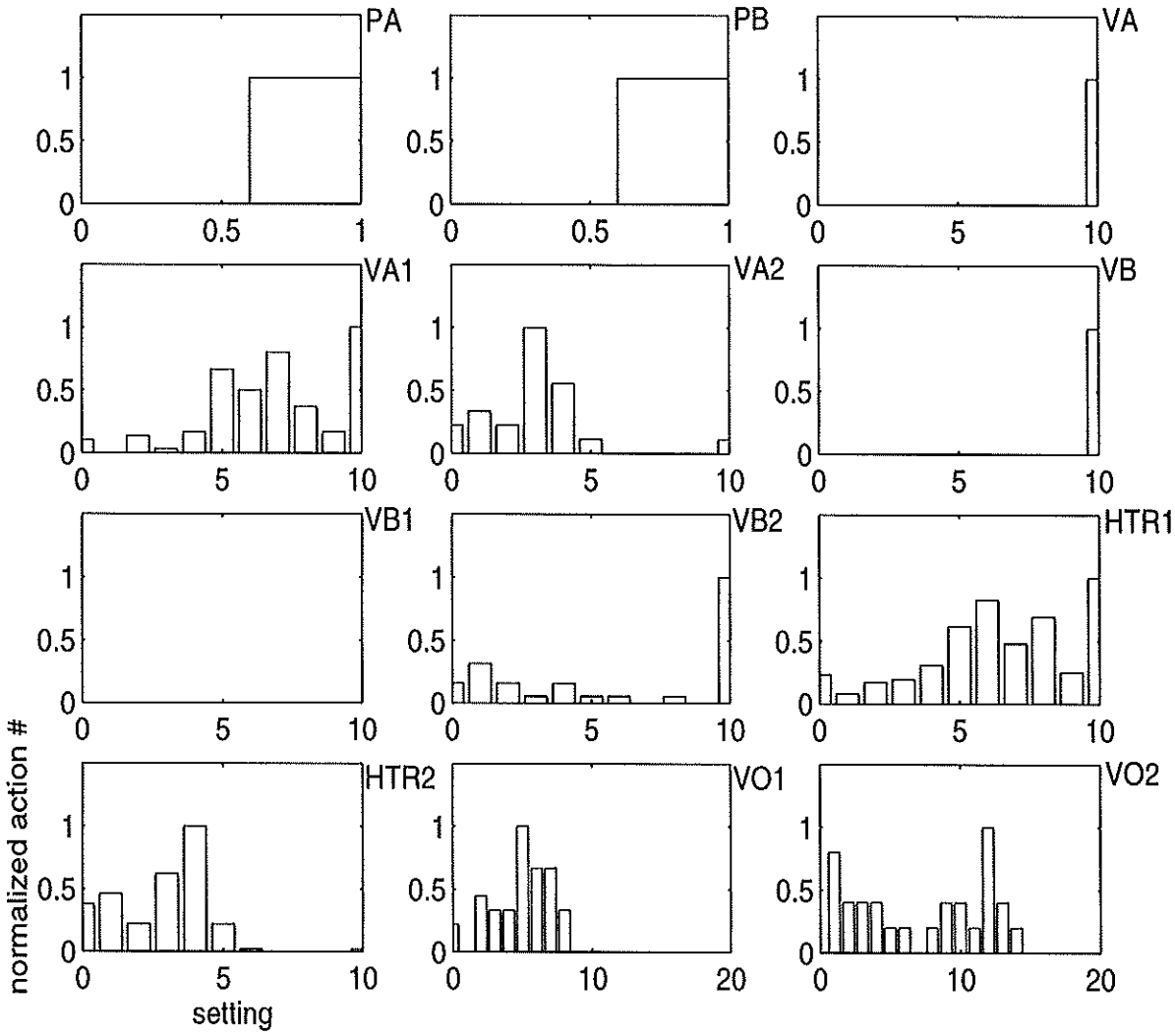

Figure 21: Normalized number of actions vs. settings at block # 5 of TL

which subjects were actually following the specific set of procedures that they report using. However, these inferences are made informally by visual inspection of the AFD graphs. It would be useful to have a more precise and more objective basis for documenting action variability.

In volume 1, we identified information theory (or entropy) as a potentially viable metric for quantifying the variability in behavior. In our case, there are several relevant entropy measures. H_c measures the variability in component usage. If operators act on all 12 components with equal frequency, then H_c is maximized. If only one component (e.g., *VA*) is always used (an impossibility in our case), then H_c is zero. Similarly, H_s measures the variability in setting values (across all components). Thus, a maximum value of H_s is obtained if all setting values are used with equal frequency, whereas a zero value is obtained if only one value setting (e.g., 7) is always used (again, an impossibility in our case). Finally, T_{cs} measures the information transmitted between components and settings, which is basically a non-parametric measure of correlation, or of the degree of interaction, between these two variables. A maximum value of T_{cs} would be obtained if these two variables are perfectly correlated. In such a case, for every component, there would be a unique setting that is consistently used (e.g., *VA* is always set on 7, *VB* is always set on 10, etc.). A zero value of T_{cs} would be obtained if these two variables are independent (i.e., for each component, all possible settings are used with equal frequency). A detailed description of how each of these variables is calculated can be found in Yu et al. (1997).

The most important of these measures is T_{cs} since it may allow us to discriminate between action-based and function-based control. If someone is using an action-based strategy to control DURESS II, we would expect T_{cs} to be high because their procedure would be comprised of precise values for particular components. Based on his control recipe (see Table 3), this is what we would predict in the case of TL by the end of the experiment. In contrast, if someone is using a function-based strategy to control DURESS II, we would expect to be T_{cs} to be low because they would use different values for setting each component, as a function of what the current goals happen to be. This is what we would predict in the case of AV by the end of the experiment, based on his control recipe (see Table 4). No predictions are made regarding H_c and H_s because these quantities are not very meaningful on their own. They are included here because they are required to calculate T_{cs}.

Note that it would also be interesting to look at the interaction between component, setting, and trial (i.e., T_{cst}). This measure would tell us to what extent subjects were becoming more or

less proceduralized over the course of the experiment. Unfortunately, we were not able to calculate this three-way interaction because the number of samples required to get a statistically significant value (see Moray, 1980; Valdes-Perez & Conant, 1983) far exceeds the data we have available from this study. To compensate for this deficiency, we instead calculated T_{cs} for each of 11 successive blocks of 20 trials. The size of the blocks was chosen so that there would be enough observations (i.e., control actions) in each block for each subject to reach statistical significance.

The results from this analysis are presented graphically in Figure 22 for all six subjects. As we would expect from the control recipe data, TL has a more stereotyped form of control than AV by the end of the experiment as indicated by higher values of T_{cs}. Moreover, we can see an increase in TL's values of T_{cs} from the sixth block onward. This point (approximately trial #100) corresponds quite closely with the point at which the aforementioned qualitative strategy shift was observed in the AFD and control recipes data (see Table 5 and Figures 20 & 21).

Another interesting observation is that the weakest performers in the P group (ML and WL) have lower values of T_{cs} than the best performer (TL). This suggests that in order to do well with the P interface, one has to develop a relatively well-defined set of procedures like TL did. Such a strategy would lead to a comparatively stereotyped set of actions, which in turn, would generate a higher value of T_{cs}.

If we perform a similar qualitative comparison of the relationship between T_{cs} and competency in the P+F group, we do not see any consistency, however. In fact, it is very difficult to distinguish the best performer (AV) and the worst performer (AS). They both exhibit lower values of T_{cs} than IS, indicating greater variability in their actions. Therefore, while the information theory approach has provided some encouraging results, it is not as diagnostic as we would like it to be.

The limitations of the information theory metrics can be divided into three categories. First, there is the issue of sample size. Many observations are required to obtain stable and statistically reliable estimates of entropy. In our case, this prevented us from calculating the measure that was actually of the greatest interest and relevance to our goals. Second, it is well known that information theory metrics are syntactic, rather than semantic. That is, they measure consistency, regardless of whether that consistency is functional or dysfunctional. This limitation manifests itself in our study by the fact that there are two very different ways of obtaining lower values of

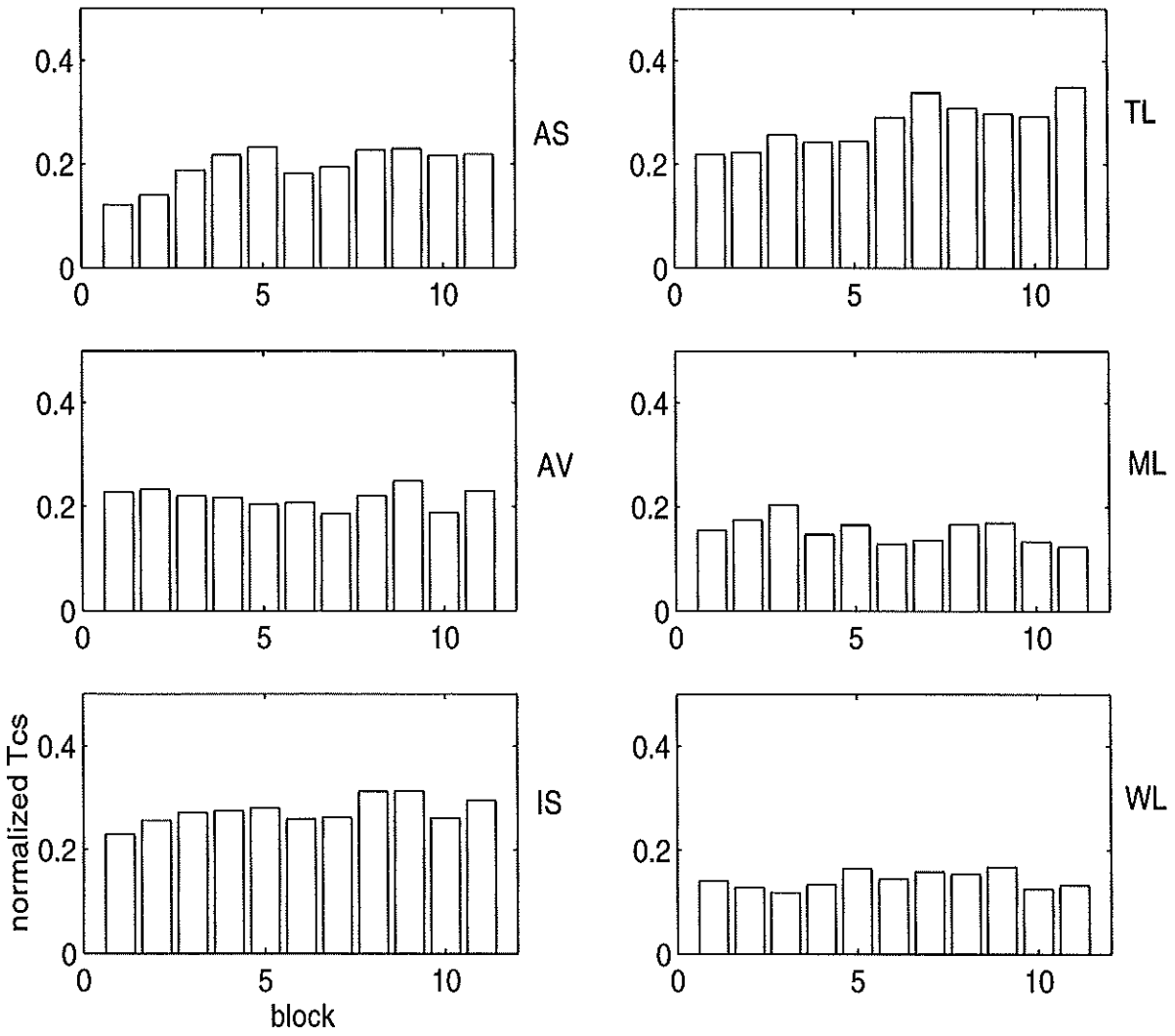

Figure 22: Normalized information transfer Tcs

T_{cs}. One way is to vary one's actions as a function of the current goal state, as AV reports doing. Another is just to act in a trial and error fashion, as AS seems to have done throughout much of the experiment (Christoffersen et al., 1994). Thus, a very good and a very bad performer could exhibit similar values of T_{cs}, but for very different reasons. We suspect that this is why the values for AS and AV in Figure 22 are essentially indistinguishable. Finally, the information theory measures we have used are zero-order Markovian (i.e., they are blind to any sequential dependencies between actions). It is possible to remedy this problem by using information theory to quantify sequential relationships across observations, but this increases the sample size criterion even further (Moray, 1980; Valdes-Perez & Conant, 1983). Thus, we could not conduct such analyses for our data set.

Abstraction Hierarchy

The measures described in this subsection are directed at overcoming some of the limitations associated with the information theory measures just described. As we will see, these analyses will provide a great deal of insight into subjects' variability and adaptation as well. The measures to which we are referring are based on two concepts which, as far as we know, have not been joined together before: the abstraction hierarchy (AH; Rasmussen, 1986) and the dynamical systems approach to complex systems (see Discussion and Bibliography in volume 1).

In the past, the AH has been used for a number of different purposes, including: as a problem space for interpreting verbal protocol data (e.g., Rasmussen, 1986); as a multilevel representation to identify information requirements for interface design (e.g., Vicente & Rasmussen, 1990; Itoh, Sakuma, & Monta, 1995); as a formal work domain representation that can be used as a basis for simulating problem solving trajectories (e.g., Bisantz & Vicente, 1994); as a means for conducting an analytical evaluation of an existing or a proposed interface (e.g., Vicente, 1990); and as part of a framework for conducting a cognitive work analysis (e.g., Rasmussen, Pejtersen, & Goodstein, 1994). However, as far as we know, the various levels of the AH have never been used to define multiple, complementary frames of reference in which changes in operator behavior or work domain state can be plotted as trajectories over time.

This novel usage of the AH as a measurement instrument can be appreciated by examining Figure 23, which provides an outline of the abstraction/decomposition representations that were developed for DURESS II (Vicente & Rasmussen, 1990; Bisantz & Vicente, 1994). There are three levels of resolution in this space connected by part-whole links (system, subsystem, and

component). Also, there are five levels of abstraction connected by means-ends links (physical form, physical function, generalized function, abstract function, and functional purpose). The bottom level of physical form will not be used here since it refers to the physical location and appearance of the work domain, features which are not particularly meaningful in a microworld simulation like DURESS II.

Figure 23 shows that the abstraction and part-whole dimensions, while conceptually orthogonal, are coupled in practice. At higher levels of abstraction (e.g., functional purpose), operators tend to think of the system at a coarse level of resolution (e.g., system), whereas at lower levels of abstraction (e.g., physical function), operators tend to think of the system at a detailed level of resolution (e.g., component). Therefore, certain cells in the space are not very meaningful (e.g., functional purpose/component). In the specific case of DURESS II shown in Figure 23, four cells have been identified as being useful for the present purposes:

- Functional Purpose/System – outputs to the environment
- Abstract Function/Subsystem – mass and energy topologies
- Generalized Function/Component – liquid flow and heat transfer rates
- Physical Function/Component – component settings

Each of these cells contains a different representation of the very same work domain.

The basic insight behind the analyses that follow is that each of these four representations of DURESS II provides a different frame of reference for measurement. Each frame of reference can then be used to conduct a different dynamic systems analysis. For example, at the top level of Functional Purpose/System, the entire system can be described in a four dimensional space defined by the outputs to the environment (i.e., the goal variables) $T1$, $MO1$, $T2$, $MO2$. The behavior of the system during one trial can be plotted as a function of time as a trajectory in this space. For a successfully completed startup trial, this trajectory would start at the origin of the space (because the system is shutdown) and would end at the small area defined by the particular goal values (and tolerances) for that trial. For this reason, we refer to this level as a goal space. This representation provides one frame of reference for conducting a dynamical systems analysis.

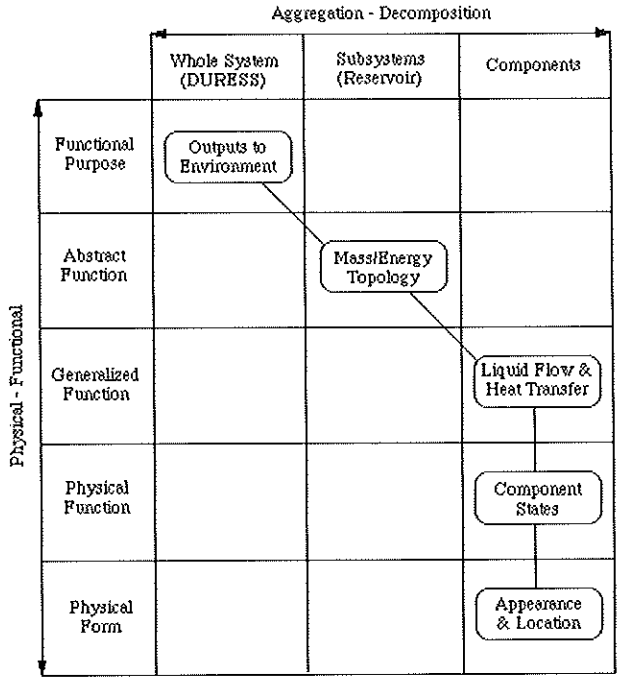

Figure 23: Representation of DURESS II in abstraction/decomposition space (from Bisantz & Vicente, 1994).

The other three representations described above can be used in an analogous fashion. For example, the next level (Abstract Function/Subsystem) describes DURESS II in terms of a 12 dimensional space consisting of mass and energy inputs, levels, and outputs: *MI*1, *M*1, *MO*1, *EI*1, *E*1, *EO*1, *MI*2, *M*2, *MO*2, *EI*2, *E*2, and *EO*2. Again, the behavior of the system for one trial can be plotted as a function of time as a trajectory in this space. Note that we are describing the very same system, but from a different perspective. Similarly, the third representation (Generalized Function/Component) describes DURESS II in terms of a different 10 dimensional space consisting of the liquid flowrates and heat transfer rates being produced by each of the components: *FA*, *FA*1, *FA*2, *FO*1, *FH*1, *FB*, *FB*1, *FB*2, *FO*2, and *FH*2 (the flows through the pumps are the same as those through valves *VA* and *VB* and thus are not included). This multidimensional space provides a third frame of reference that can be used to conduct a dynamical systems analysis.

The fourth and final representation (Physical Function/Component) is somewhat different from the previous ones because it describes the state of variables that operators can control directly, i.e., the settings of the components. This space consists of 12 dimensions, one for each of the components in DURESS II: *PA*, *VA*, *VA*1, *VA*2, *VO*1, *HTR*1, *PB*, *VB*, *VB*1, *VB*2, *VO*2, and

*HTR*2. In this case, a trajectory in the space is a record of the subject's actions for a particular trial. Thus, whereas the previous trajectories evolve as a function of system dynamics, at this level the trajectory does not evolve until a subject changes the setting of one of the components. It is for this reason that we refer to this frame of reference as an action space.

In summary, the AH provides a systematic basis for identifying frames of references for conducting dynamical systems analyses. Each cell in the abstraction/decomposition space provides a different representation of the same system, and thus a different frame of reference for measurement. As we will see, each of these frames of reference provides complementary information, and as a whole, they provide a great deal of insight about operator adaptation that we have been unable to obtain using any other measures. Note, however, that the frames of reference we have described above are not mutually exclusive. For example, the variable *MO*1 appears in more than one level because it is a goal variable as well as a mass variable. Note also that some discretion is involved in choosing the specific variables to be included at each level. For example, we decided to omit the reservoir volumes from the Physical Function/Component description. Although a reservoir is a component, operators have no direct control over the state of the reservoir. By omitting the volumes, we preserved this frame of reference as an action space

Now that the four frames of reference have been described, we can make some predictions based on the control recipes data (see Tables 3 and 4). If TL, the best subject in the P group, is following a detailed set of procedures, as he suggests, then we would expect that the variability in his trajectories at the action space level would be comparatively low. Trial after trial, he should be performing very similar actions, based on the detailed quantitative settings specified in Table 3. In contrast, AV, the best subject in the P+F group, should have a higher level of variability at the action level because he does not seem to have a stereotyped set of actions for controlling the system. Instead, he reports adapting his actions to the specific goals for each trial (see Table 4). Thus, we would expect the action space trajectories of TL to be less variable than those for AV, at least towards the end of the experiment.

If AV's control recipe is accurate, then he controls the system by focusing on the goals or functions to be achieved rather than the specific actions that are used to achieve those goals. In fact, the actions that are required to do the task should change from trial to trial, whereas the task goal remains invariant (although the specific setpoint values change from trial to trial). If this is

the case, then we would expect that AV's trajectories should be relatively consistent at a functional level of abstraction above that of actions (i.e., functional purpose, abstract function, or generalized function). There is not enough information available to predict which of the remaining three frames of reference AV is focusing on. Nevertheless, we would expect that an inversion of results when we adopt a different frame of reference. AV's trajectories at a higher level of abstraction should be less variable than those for TL, at least towards the end of the experiment.

The variability in the trajectories for each subject were calculated at each level of the AH described above, by block (see Yu et al., 1997 for a detailed account of the mathematics). We will begin by discussing the results from the goal space level, Functional Purpose/System, illustrated in Figure 24. Note that these trajectories were normalized with respect to the setpoint values for each trial, thereby allowing us to meaningfully compare trajectories across trials. The graphs in Figure 24 show the variability in trajectories for each subject over the course of the entire experiment, as a function of 11 blocks of approximately 20 trials each. Perhaps surprisingly, the data for TL and AV are quite similar. After the initial part of the experiment, both of these subjects exhibit very consistent trajectories at the goal space level. Thus, as with many of the measures of performance discussed earlier, this frame of reference does not allow us to differentiate TL and AV. According to this measure, they are behaving in the same fashion. Furthermore, we again see a small but noticeable, permanent improvement in the performance of TL starting between trials 80-100 (i.e., block 5), just as with previous analyses. Figure 24 also shows that, with the exception of ML, all of the other subjects were also able to achieve consistently low goal space variances by the latter part of the study. However, we know from previous analyses (see above) that there are substantial differences in skill between these subjects, particularly AS and AV. Thus, the goal space variability measure is not a very diagnostic one.

Figure 25 shows the results of the variability analysis at the action space level. Note that, at this level, the trajectories are not normalized with respect to the goal values for each trial. Such a normalization is not possible because there is no direct relationship between goal values and component settings. Thus, the variability analysis at this level is based on absolute setting values (with a compensation for the fact that different components have different scale values; see Yu et al., 1997). The results indicate that the least proficient subjects in each group, AS and ML, have

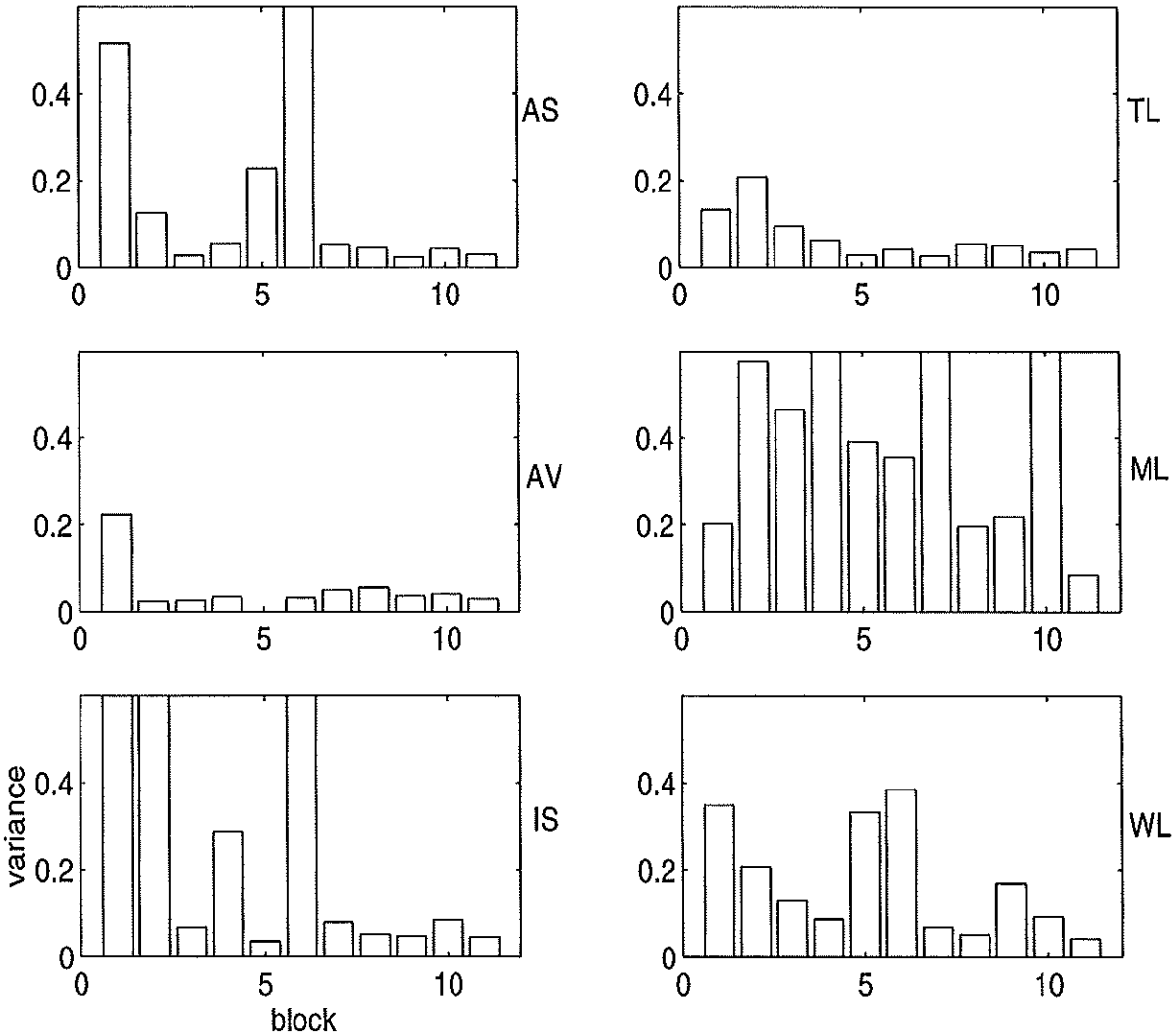

Figure 24: Variance of outputs

the highest action variability, presumably because they engage in iterative, trial and error behavior. The primary contrast of interest, between AV and TL, is difficult to discern in Figure 25 because of the relatively small differences compared to the scale size. Figure 26 shows a direct comparison of the action variability for AV and TL during the last few blocks of the experiment. These data clearly show that TL's behavior is consistently less variable than AV's. Although the difference is not a large one, it is consistent with the observation that TL's behavior is driven more by a fixed set of specific actions than AV's. Thus, this finding provides support for the hypotheses generated by the control recipe data for these two subjects.

Figure 27 shows the results of the variability analysis at the Generalized Function/ Component level. The trajectories in this frame of reference were also not normalized with respect to the goal values for each trial for the same reasons stated above. The results from this frame of reference are actually very similar to those in Figure 25, and thus add little to our understanding. In retrospect, the reason for this similarity is straightforward. These two levels, component settings and components flows, are actually strongly correlated. By inspecting the equations describing the system dynamics (see Yu et al., 1997), we see that there is a direct correspondence between these two sets of variables after the transient produced by a change in component setting. In other words, if we are given the component settings we can usually uniquely derive the liquid flow rates and heat transfer rates (for normal trials). The only times during which this relationship is weakened is during the short period after an operator action. For example, if *VB* is closed and a subject sets it to 10, it will take approximately 15 seconds for the valve to actually reach that state because of the first order lag dynamics. After that lag, there will be a fixed relationship between valve settings and flow rates. However, during the preceding transient period while the valve is opening from its closed state to its final state, this relationship is not fixed. It is for these reasons that the results in Figure 27 are so similar to those in Figure 25.

The final set of AH variance analyses were conducted at the Abstract Function/Subsystem level. There are two interesting differences between this frame of reference and the last two we have just described. First, the measurement is taking place at an aggregate level. We are now examining variables at the subsystem level, which are aggregates of the variables that we examined at the part-whole level of components. Second, measurement at this level is in terms

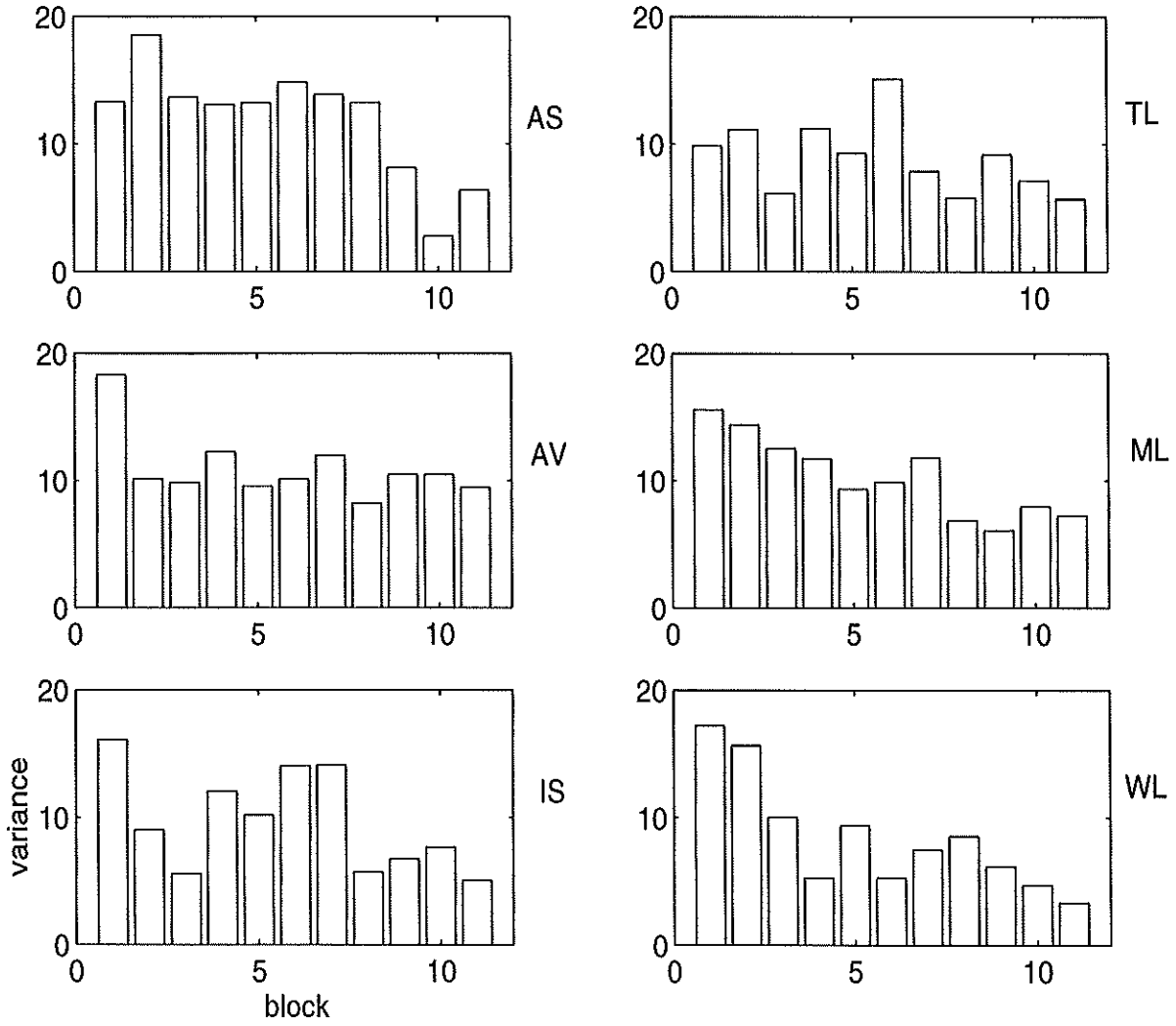

Figure 25: Variance of component settings (actions)

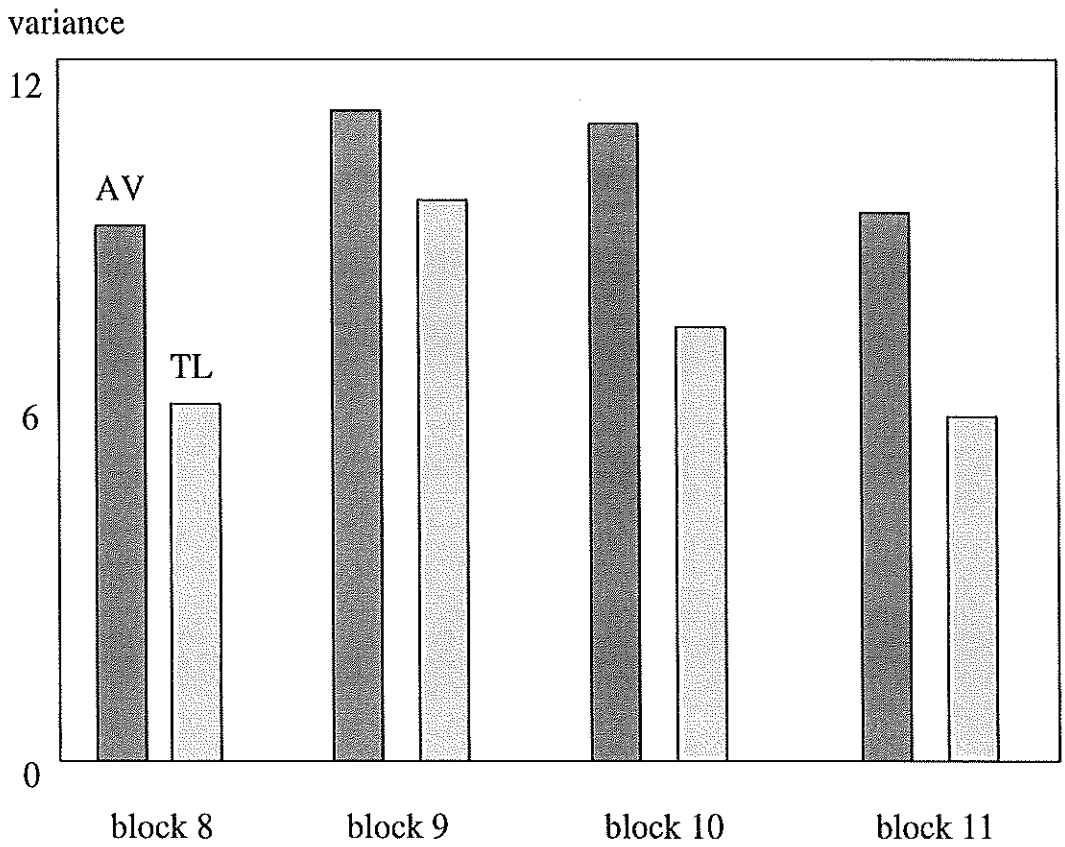

Figure 26: Comparision of variance of component settings at the last four blocks

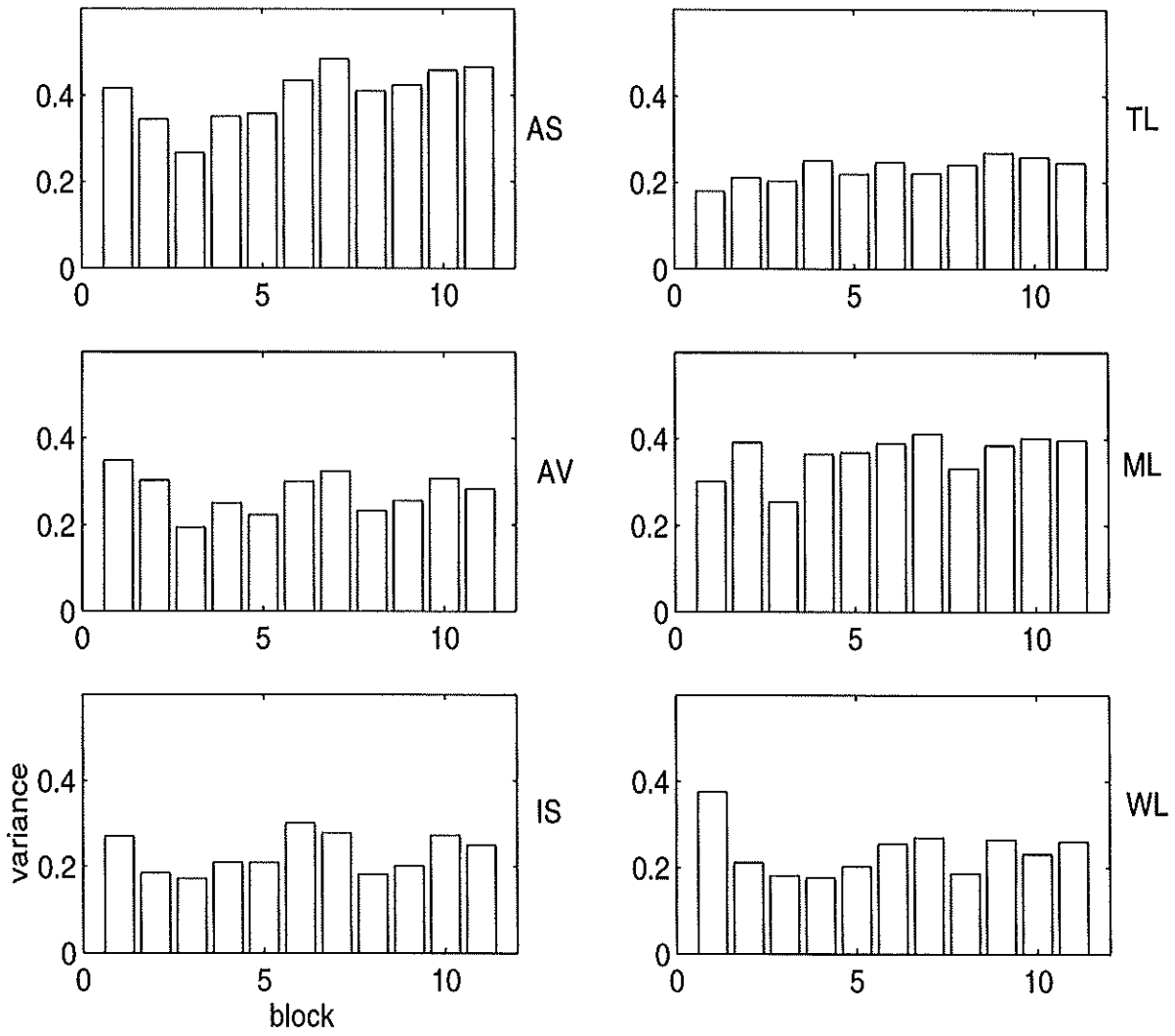

Figure 27: Variance of flows and heat transfer

of variables that describe the system in terms of first principles (i.e., mass and energy conservation laws). In this sense, this frame of reference is a privileged level of description.

As before, we examined the variance of each subject's trajectories as a function of block number. The first analysis we conducted at this level was based on trajectories that were not normalized for the different goal values for different trials. In this case, the calculations are based on absolute data values (except for a compensation for the fact that different components have different scale values; see Yu et al., 1997). The results from this analysis are presented in Figure 28. It is difficult to discern any patterns in the data. In particular, TL and AV do not seem to be differentiated in any consistent, meaningful fashion.

There is another way to look at these data, however. Because each trial has a different set of setpoints for the four goal variables, we would expect there to be variance in the trajectories for this reason alone. Although the trajectory for each trial begins at the origin (because the system is shutdown), the end point for each trajectory (and presumably the trajectory itself) will be different for different trials as a function of the setpoints for that trials. If we assume that subjects try to stabilize both volume and temperature for each reservoir, then it is possible to correct the trajectories for differences in setpoint values across trials. This is accomplished by dividing the mass input and output flowrates (i.e., $MI1$, $MO1$, $MI2$, $MO2$) by the demand setpoints ($D1$, $D2$), and dividing the energy input and output flowrates (i.e., $EI1$, $EO1$, $EI2$, $EO2$) by the product of the demand setpoints ($D1$, $D2$) and the temperature setpoints ($T1$, $T2$). The details of these calculations are provided in Yu et al. (1997). Normalizing the trajectories in this fashion eliminates any variability caused solely by differences in setpoints across trials.

The results from this second, more informative analysis are presented in Figure 29. There are several interesting findings emerging from this alternative way of looking at the data. The most important of all is the large difference between the variances for TL and AV. From the beginning of the experiment, but especially in the second half, the trajectory variance for AV is much smaller than that for TL. This strong result is consistent with the predictions we made from the control recipe data. AV is thinking about, and controlling, the system at a high level of abstraction, focusing on the mass and energy level. Moreover, he contextualizes his control at this level based on the setpoint values for each trial. This can be observed by the noticeable difference in the data in Figures 28 and 29 for AV. It is only when we compensate for differences in setpoint values that we see that, at high level of abstraction, AV is acting in a

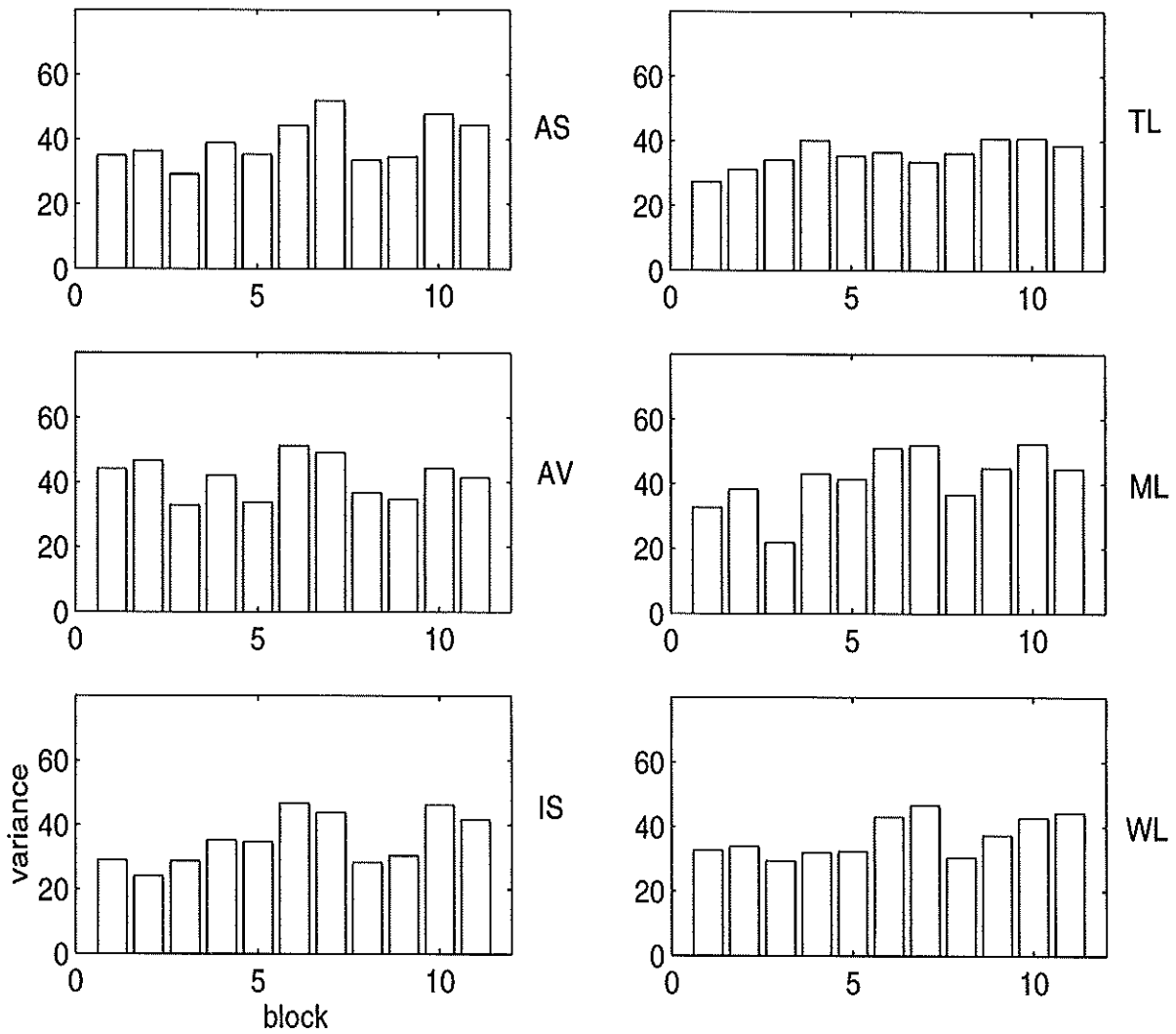

Figure 28: Variance of mass and energy (normalized by scale only)

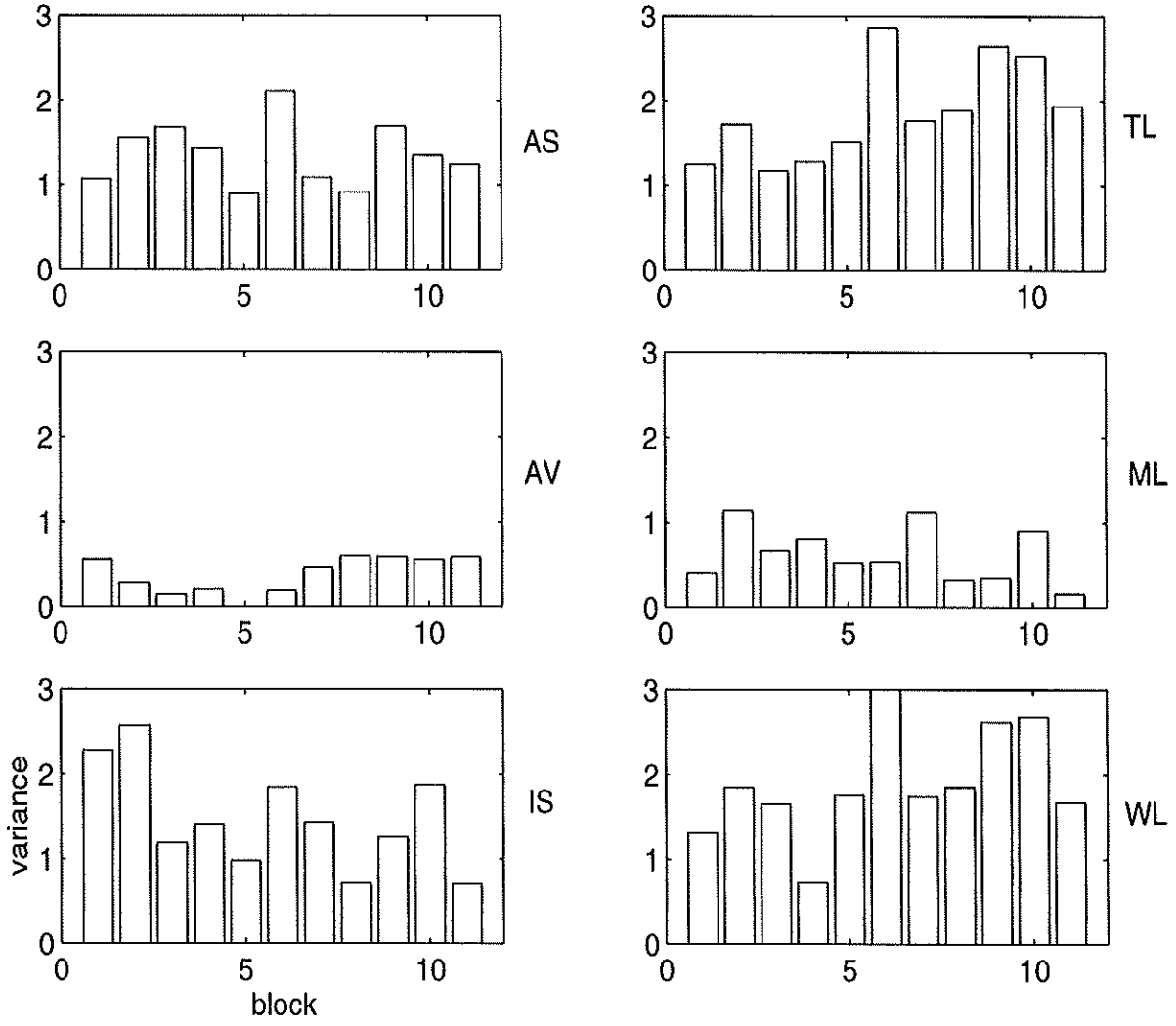

Figure 29: Variance of mass and energy (normalized by both Scale and goals)

consistent fashion across trials. In contrast, the regularities in TL's behavior are more at the action level (Figure 26) where he exhibited a lower variance than AV. Because TL's actions are relatively similar for trials with different setpoints, his behavior is not as contextualized (or situated) as AV's. Thus, when we examine TL's data at a contextualized, functional level of abstraction, he exhibits less structure than does AV. This set of results has important theoretical implications that we will discuss below.

Another important result to emerge from this analysis is the noticeably abrupt increase in the variability for TL. Yet again, we see a distinct change in his data, in this case after trial #100 (i.e., block 5). The other interesting finding that can be seen in Figure 29 is that ML, perhaps the worst subject in the P group, also exhibited a very low variance at this level of abstraction. In fact, towards the end of the experiment, his variance was sometimes even lower than that of AV. This suggests that he too was controlling the system at a high level of abstraction.

Discussion

Motivated by the control recipe results, the analyses in this section were geared towards measuring differences in the variability in operator action. Three different types of measures were considered: AFD graphs, information theory, and dynamical systems analyses at different levels of the AH. What insights have we gained from these analyses? Three points seem to stand out, two of theoretical interest and one of methodological interest.

The most important theoretical insight stems from the behavioral evidence showing that the best subjects in each interface group adopted very different ways of controlling DURESS II. The previous section showed that both TL and AV attained a very proficient level of performance that was superior to that obtained by the other subjects in their respective groups. However, the various analyses of action variability show convincingly that they were using qualitatively different strategies to control the system. TL adopted a rote set of procedures that allowed him to control the system efficiently, at least under normal conditions, but without any deep understanding of the system. Thus, the regularities in his behavior across trials can be found primarily at the level of control actions. In contrast, AV focused on the functions that had to be satisfied rather than the specific actions that were required to achieve those goals. Thus, the regularities in AV's behaviour could be best revealed at the level of first principles. The AH analyses in particular show that AV controlled the system by focusing on higher-order functional variables and used whatever actions were required to satisfy the current goals. Because he was

not following a script, AV exhibited a greater level of variability in his actions than did TL. However, this variability was not chaotic. Instead, it was situated to the current context.

There are several reasons why this is an important result. First, it is indirectly supported by converging evidence from a number of different measures. The control recipes, the AFD graphs, the information theory analyses, and particularly the AH dynamical systems analyses all provide evidence that is consistent with this interpretation.

Second, these differences can be interpreted in terms of the different interfaces that TL and AV had to control DURESS II. AV used the P+F interface which presented him with both physical and functional information. Because he could see the state and structure of the system, he did not have to memorize a set of procedures. Instead, he could use the information in the P+F interface as an error signal to generate actions that were appropriate to the current context. Thus, there was a stronger coupling between AV's actions and DURESS II, as shown by the AH analysis at the level of first principles. This stronger coupling also led to a larger degree of <u>context-conditioned variability</u> (Turvey, Fitch, & Tuller, 1982). Because different trials had different goal setpoint values, AV's actions were more variable across trials. Thus, unlike other subjects (e.g., AS), this variability was functional not dysfunctional. It was merely reflecting changes in the demands imposed by the system.

TL used the P interface which presented him only with physical information. Although it provided him with enough feedback to control the system efficiently, the P interface does not reveal all of the interactions that govern DURESS II. As a result, TL could not rely primarily on the feedback in the interface to generate his actions. Instead, he had to acquire a rote set of detailed actions that he used as a script for each trial. Thus, TL's actions were less variable across trials because they seemed to be governed more by the steps in his procedure than by what was presently going on in the system. Consequently, TL exhibited a weaker coupling to the first principles of the system. The regularities in his control were at the action level.

Interestingly, these two forms of control are virtually identical to concepts that have been discussed for years by human movement scientists (see Bibliography in volume 1 for a detailed list of references). The motor control literature distinguishes between two different types of behavior, execution-driven and program-driven (Turvey, 1988). In <u>execution-driven</u> control, behavior is governed by a product (a goal to be achieved). In contrast, in <u>program-driven</u> control, behavior is governed by a process (actions to be performed). Behavior can be driven by a

combination of these two modes, but it is easier to illustrate the differences by comparing the two pure forms which define the continuum.

These two forms of control serve as the basis for two, competing accounts of human motor control. The traditional program-driven account is that motor control is governed by a motor program (much like a computer program) residing in human memory, which specifies a set of actions that need to be performed to perform a task (Keele & Summers, 1976). Thus, it is the actions that are performed which remain invariant since these are specified by the program. The alternative execution-driven account advocated by ecological psychologists is that motor control exploits goal-relevant information from the environment to directly specify how the goal can be achieved (Turvey, 1990). This information is mapped onto action using a flexible and adaptive coordinative structure (Tuller, Turvey, & Fitch, 1982). In this case, it is the outcome that remains invariant because the focus is on the goal to be achieved, not the specific actions to be performed.

The role of information is a key differentiator of these two theoretical accounts. The traditional account assumes that the stimuli that are perceptually available from the environment are low-level and not very meaningful. As a result, a motor program stored in memory must make up for this deficient input if the task is to be performed successfully. In contrast, the ecological account tries to identify situations where there is rich information in the environment to support action. The ecological account of human motor control has several advantages, including: less burden on memory; less processing effort; greater stability in the face of disturbances; and more flexible, adaptive control (Turvey, 1988). Therefore, it provides a good role model for how to support coordination.

Our results can be readily interpreted in terms of these motor control concepts. TL used the P interface which provided an impoverished set of stimuli. As a result, he had to resort to developing the equivalent of a motor program to guide his actions. AV, on the other hand, used the P+F interface which provided a rich source of information about the system. Consequently, he was able to develop the equivalent of a coordinative structure that mapped information onto action in a functional, contextualized fashion. These results are fascinating because the differences we have observed between AV and TL provide a prototypical example of the contrast between execution-driven and program-driven control, respectively. As far as we know, this is the first time that empirical evidence for something like a coordinative structure has been

obtained in the cognitive engineering literature on process control. If generalizable, this finding is potentially of great theoretical significance. It suggests that the problem of how to coordinate a complex system with many degrees of freedom in a goal-directed fashion may have a generic solution, whether it is implemented biologically (as in the case of motor control) or in a well-designed human-machine system (as in the case of AV and the P+F interface for DURESS II). Therefore, it seems worthwhile to explore the connections between theories of motor control and long-term operator adaptation. The result may very well be progress towards a general theoretical account of the generic systems problem of coordination.

The other important theoretical insight obtained from these analyses stems from the evidence indicating that TL experienced a qualitative strategy shift during the course of the experiment. The control recipe data suggested that TL started off by trying to understand the basis for his control actions. However, about mid-way through the experiment he seemed to give up this quest for understanding, and instead resorted to developing a rote set of detailed steps that seemed to get the job done efficiently. The AFD graphs provide objective, empirical evidence of a strategy shift occurring around trial #88. Similarly, the information theory analyses showed an increase in action consistency around trial #100, a symptom which is also consistent with the development of a procedure. The AH dynamical systems analyses also showed discrete changes at about the same point in the study. Recall that several of the performance analyses reported in the previous section indicated that it was at this time that TL exhibited a quantum improvement in performance. He was clearly the best performer in the P group. Together, these facts lead to the following provocative hypothesis: to control a system efficiently with an impoverished interface, it does not pay off to try to understand the basis for one's actions; instead, one is better off following a "script" that seems to work most of the time. This hypothesis should be tested in future research with a larger sample size. If evidence is found in favour of this conjecture, then the limitations of traditional interface design practices will be put in a new theoretical light.

Finally, the methodological novelty generated by these analyses is that the AH and the dynamical systems approach can be profitably combined to develop novel and informative measurement tools. The different levels in the AH can be used to define frames of references for the measurement of operator actions, or process behaviour, over time. As far as we know, this is the first time that the AH has been used for this purpose. There are several advantages to adopting this technique. First, the AH provides a relatively principled basis for identifying

frames of reference for dynamical systems analysis. This suggests that it may have applicability in different domains beyond DURESS II. Although the content of the levels of the AH will differ for various work domains, the relationship between levels will be the same. Thus, the combination of the dynamic systems approach and the AH as a measurement tool is a potentially widely generalizable technique. If the results obtained here are any indication, then this tool should allow researchers to obtain important insights into their data. Having said this, it should be pointed out that our experience suggests that the AH cannot be used in an algorithmic fashion. Some discretion is required to identify the variables that should be included at each level.

5. ANALYSES OF ADAPTATION TO STATE

The AH analyses described in the previous section allowed us to measure the variability in operator action (in the case of the action space) and the extent to which process trajectories were consistently adapted to the goal values for each trial (in the case of the goal and first principles spaces). However, none of the measures we have considered so far have focused directly on the relationship between operator actions and system state. In this section, we consider several measure of adaptation to state.

Taxonomy

To conduct such an analysis, we must have an idea of which system state properties that operator actions might be adapted to. One very simple taxonomy is as follows:

1. **Goal oriented actions**. The purpose of these actions is to get the goal variable values towards the goal boundaries, in an effort to achieve steady state. These actions are expected to occur in phases 1 and 2 of a trial, as defined earlier (i.e., not during the last 300 s).
2. **Failure avoiding actions**. These actions are geared towards keeping the system within safe levels of operation. Blow-ups can be caused by reaching failure boundaries, which are hard constraints on control. Failure avoiding actions are also expected to occur in phases 1 and 2.
3. **Fine-tuning actions**. These actions are directed towards stabilizing the goal variable within the goal boundaries. As a result, these actions are predominantly those which occur during phase 3 of the trial (i.e., during the last 300s).

4. **Exploratory actions**. Sometimes, subjects seem to experiment with certain system components to learn more about the effects of their actions on the system or to test the limits of the system (Christoffersen et al., 1994). These actions may appear irrational and inappropriate, but they actually play a very valuable role during the process of adaptation.
5. **Miscellaneous actions**. All remaining actions are included in this miscellaneous category.

To put this taxonomy to use, we need some way of empirically operationalizing the different types of actions. This is a challenging task because we cannot rely on subjects to explain verbally the rationale for each and every action. Thus, we must see if there is a way of operationalizing these categories from the data contained in the log files and our knowledge of DURESS II. So far, we have not been able to discover a way of empirically identifying exploratory actions, so we will not consider this category any further. The fifth category is viewed as noise from the perspective of this taxonomy, and so it too will not be considered any further.

Using a state space approach based on the dynamic systems perspective (see volume 1), we have developed preliminary ways of empirically identifying the three remaining categories of actions (i.e., goal oriented, failure avoiding, fine tuning). Goal oriented actions are geared towards getting the system into the goal state, which serves as an attractor. Thus, one might expect that these actions would be driven by current values of TTC (see above). The longer the TTC, the more likely the operator should be to act to get the system into the goal state. This conjecture can serve as the basis for empirical measurement by plotting the frequency of actions as a function of the longest of the TTC values for the four goal variables. Only actions that occurred during phases 1 and 2 of a trial were considered.

Failure avoiding actions are geared towards keeping the system away from failure boundaries, which serve as repellors. Thus, one might expect that these actions would be driven by current values of the time to failure boundaries (TTFB). The shorter the TTFB, the more likely the operator should be to act to get the system away from the failure boundaries. This conjecture can also serve as the basis for empirical measurement by plotting the frequency of actions as a function of the shortest of the TTFB values for the five failure boundaries (see Yu et

al., 1997 for the calculation details). Only actions that occurred during phases 1 and 2 of a trial were considered.

Finally, fine tuning actions are directed at stabilizing goal variables and keeping them within the goal regions. In this case, the goal boundaries serve as repellors to be avoided. Thus, we might expect that these actions would be driven by the current values of the time to goal boundaries (TTGB). The shorter the TTG, the more likely operators should be to act on the system to keep it from exceeding the goal boundaries. This conjecture was evaluated by plotting the frequency of actions as a function of the shortest of the TTGB for the four goal variables. Only actions that occurred during the last 300s of a trial were considered.

Results

For each subject, we plotted graphs of action frequency vs. TTGB, TTC, and TTFB for the first and last block of 22 trials. A bin size of 15s was used for all analyses, and times that were greater than 300s were lumped together into one category. We expected that more proficient subjects, and perhaps the P+F group as a whole, would exhibit signs of a tighter coupling between actions and system state. This better adaptation would be observed as a peak in action frequency at low values of TTGB and TTFB and at high values of TTC because it is at these times that urgent action is required. Lack of adaptation would be evidenced by a uniform distribution of action frequency as a function of TTGB, TTC, and TTFB, representing a lack of sensitivity to system state.

There were no noticeable differences between interface groups, so in the interests of brevity we will only present the results from the P+F group. Figure 30 shows the data for the first block of trials and Figure 31 shows the data for the last block of trials at the end of the six month study. Several patterns can be observed in the graphs. With respect to TTGB, we see that the distribution for all subjects became more skewed and moved to the left with experience. Over time, subjects learned to act more frequently when the TTGB was small. This trend is particularly clear in the case of IS. This change is evidence of a greater sensitivity to the goal boundaries, and shows that there was an increase in subjects' adaptation to system state over time. Notice that AS' distribution in the last block of trials (Figure 31) is more spread out and farther to the right than those of the other two subjects, indicating that he was not as well adapted to system state. This finding is consistent with our earlier observations that AS was the least

proficient subject in the P+F group. Similar differences were observed in the data for the P group with TL showing the strongest adaptation.

As for TTC, there was little change with experience. In both blocks, we see that all subjects acted when TTC was very high. It is in these situations that there is a strong need to act to bring the system towards the goal state. In this case, there do not seem to be any obvious differences between subjects in terms of how well adapted their actions are to system state. The findings for the P group showed the same pattern.

As for TTFB, there is perhaps a slight shift to the left in the distributions of the subjects, suggesting an increase in the adaptation to failure boundaries. Again, there do not seem to be any strong differences across subjects. The data for the P group were not noticeably different.

Discussion

The state space analyses described in this section were not as informative as we had expected. There does seem to be some evidence of an increase in the extent to which actions are adapted to the TTGB with experience. Actions seem to be adapted to TTC right from the start of the experiment for all subjects. There is a hint of an increase in the adaptation to the TTFB with experience. However, there were not noticeable differences between interface groups, and very few differences between proficient and weaker subjects.

In retrospect, these findings are not surprising given the bluntness of the measurements. There are several reasons to suggest that there may be a substantial amount of noise in these graphs. First, an action motivated by one reason (e.g., high TTC) can also show up on another graph (e.g., TTFB). This will add noise to the latter graph. Second, the three types of actions we have examined are surely incomplete. There may very well be other reasons behind a subjects' actions (e.g., exploratory behaviour), and again, this will be seen as noise in these graphs. Third, these analyses used a single action as the unit of analysis. However, as the AH analyses showed, the regularities in at least some subjects' actions are at a more molar level of analysis. Such patterns cannot be easily observed in the graphs we have presented in this section.

Having said this, it is worthwhile pointing out that the analyses described in this section can be made more rigorous by quantifying the statistical properties of the distributions in the graphs. Newell and Hancock (1984) have pointed out that distributions can be profitably compared by calculating, not just the mean and the variance as is usually done, but also higher moments of the distribution, such as skewness and kurtosis. Future research might benefit from applying Newell

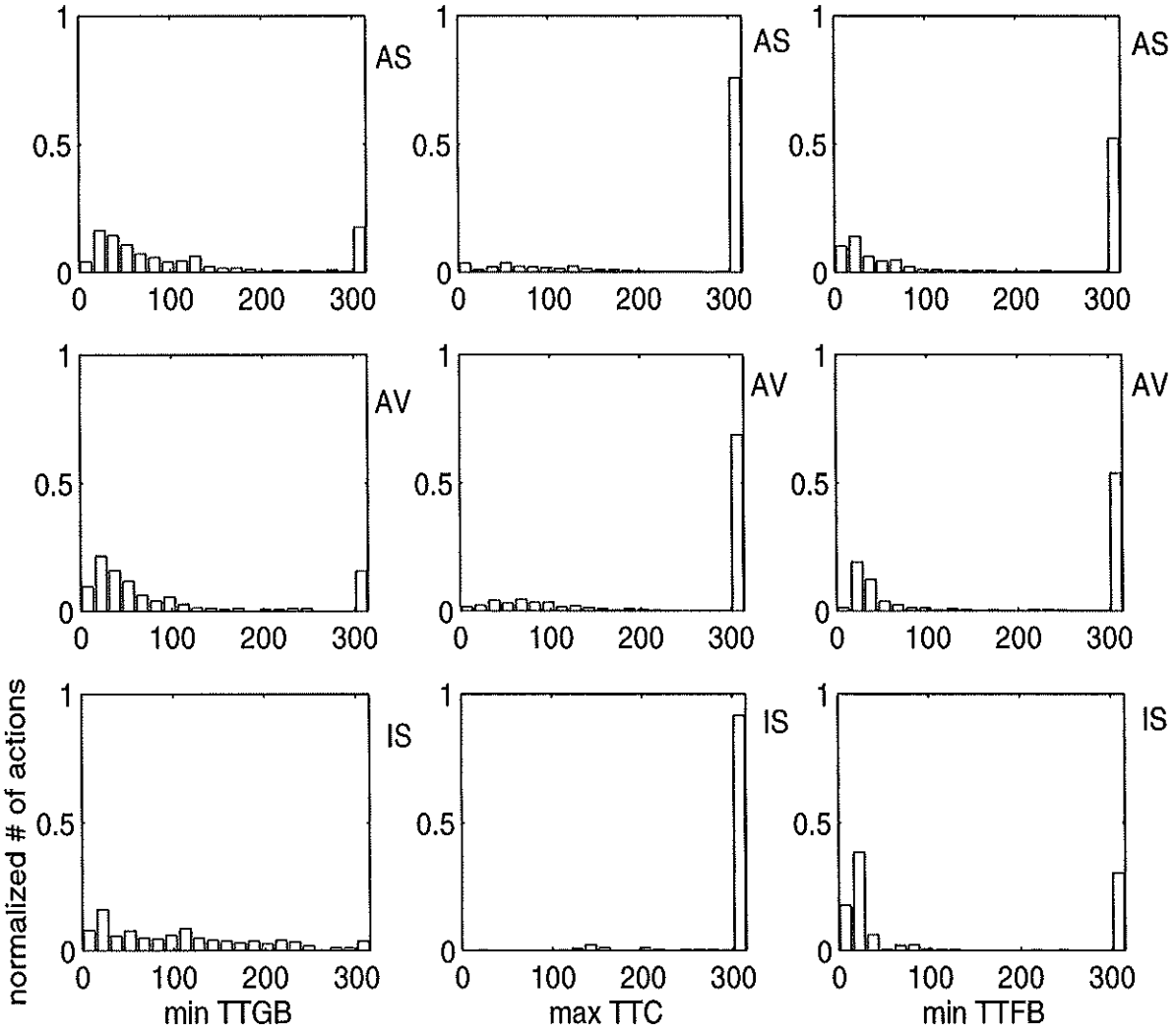

Figure 30: Normalized number of actions vs. TTGB, TTC and TTFB of P+F subjects at the first block

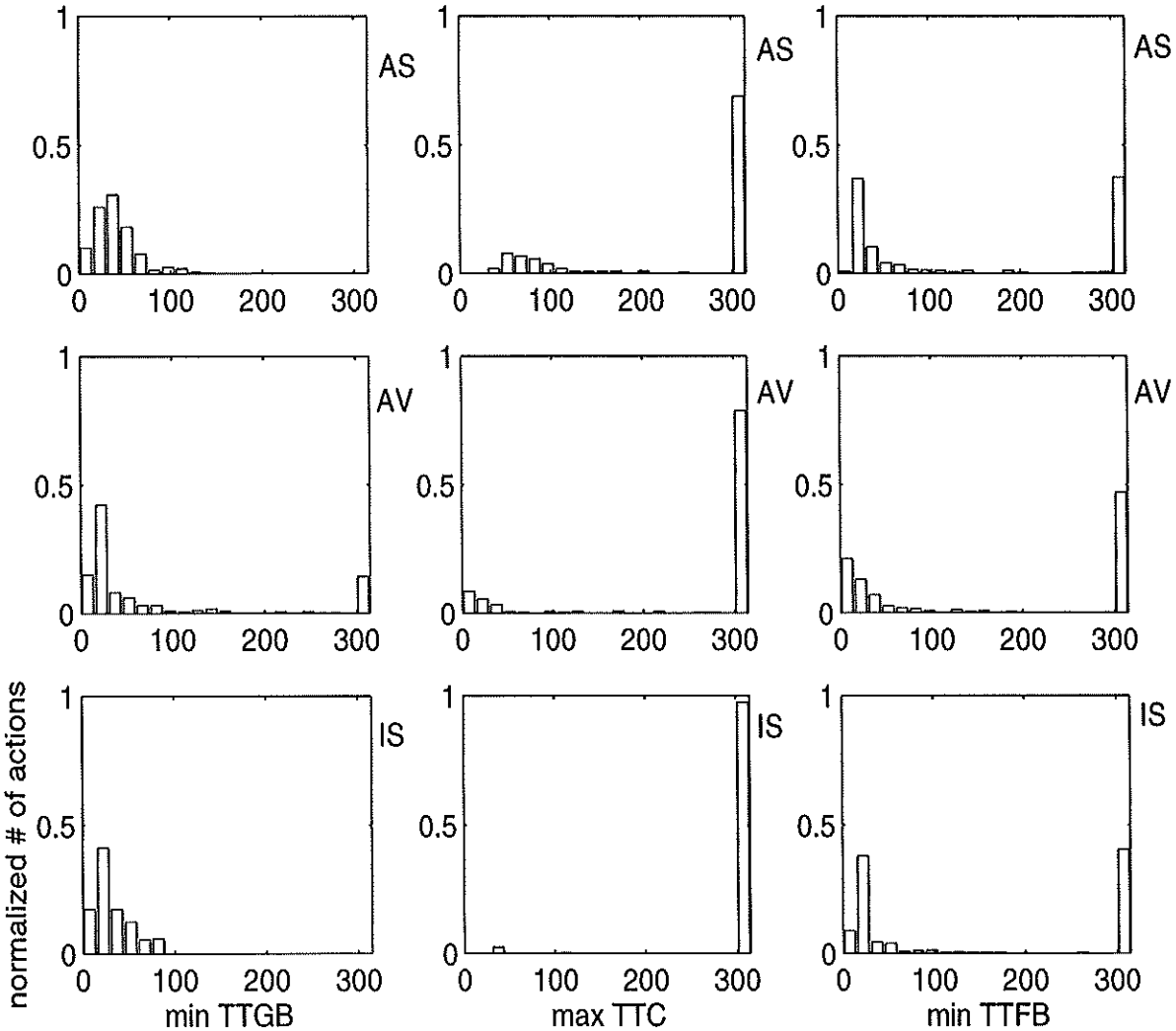

Figure 31: Normalized number of actions vs. TTGB, TTC and TTFB of P+F subjects at the last block

and Hancock's insights to the state space measures described in this section. Such work may make it is easier to objectively document differences in adaptation to system state between subjects and within subjects with experience.

6. ANALYSES OF ADAPTATION TO STRUCTURE

The discussion of adaptive landscapes in volume 1 showed us that any investigation of adaptation requires an environmental referent. We cannot judge behaviour as being adaptive unless we have an independent way of defining adaptation a priori. In the case of the DURESS II microworld, we are fortunate because we have a cognitive work analysis of the environment that we can use as a referent (Vicente & Pawlak, 1994). Using that analysis, we can identify several structural features of DURESS II which can serve as selection constraints on adaptive behaviour. In this section, we discuss three data analyses, each based on a different structural constraint.

Coordinating Mass & Energy Inventories

For each reservoir in DURESS II, there is a temperature goal that must be achieved and maintained to successfully complete a startup task. Temperature, however, is defined as energy per unit mass. Consequently, there are two ways of changing reservoir temperature, by changing the mass level or by changing the energy level in the reservoir. We can use this axiomatic relationship between these three variables to define a frame of reference to measure operator adaptation to structure (Howie & Vicente, in press). Specifically, the total mass can be plotted against the total energy in each reservoir. Because temperature is proportional to energy per unit mass, a straight line in the energy vs. mass space represents an isotherm (i.e., a constant temperature). Thus, the goal temperature for each reservoir can be plotted as a line in this graph. As with the other state space measures we have developed, the behavior of the reservoir during a particular trial can be plotted as a trajectory in this space. For a successfully completed startup trial, the trajectory would start at the origin of the graph (because the reservoirs are empty) and end somewhere on, or very close to, the goal isotherm. Note that the goal state is a line, and not a point, in this space because subjects are free to choose whatever volume (i.e., mass level) they wish for each reservoir.

By examining the trajectories in this space, we can gain some insight into the degree of coordination exercised by each subject in controlling the mass and energy levels. Comparison across trials should show the development of coordination strategies. We would expect that

more proficient and experienced subjects should exhibit more direct trajectories in this space that are tailored to the goal isotherm. In contrast, less proficient and experiences subjects should exhibit trajectories that are more meandering and not as adapted to the relationship between temperature, energy level, and mass level.

Figures 32 and 33 show the energy vs. mass graphs for the first trial completed by each of the P+F and P subjects, respectively. The dotted lines represent the goal isotherms for each reservoir. We can see from the graphs that most subjects took a wandering path through the space until they eventually settled on the goal isotherm. The trajectories taken by AS are particularly unstructured and inefficient.

Figures 34 and 35 show the energy vs. mass graphs for the last trial completed by each of the P+F and P subjects, respectively. There are several noteworthy changes. First, the trajectories of all of the subjects are more direct. There is less meandering through the space. Second, the trajectories are not only more direct but they are also better adapted to the relationship between temperature, energy, and mass, as evidenced by the fact that they tightly follow the isotherm lines. Thus, with experience, subjects seem to have learned to control energy and mass in a more coordinated, goal-oriented fashion. Third, AV and TL exhibit the most adapted behaviour of all of the subjects, once again showing their superiority. Within the resolution limits of the graph, AV and TL are able to track the isotherm perfectly from the beginning to the end of the trial. Thus, there is compelling visual evidence to indicate that the behaviour of these two subjects is exquisitely adapted to this structural constraint. AV and TL are able to coordinate changes in energy and mass levels in a way that is very sensitive to the structural constraints of the system and the goals of the task.

Steady State Heater Settings

A second structural feature of DURESS II to which subjects can become adapted is the relationship between demand setpoints ($D1$, $D2$), steady state heater settings ($HTR1$, $HTR2$), and the goal temperatures for each reservoir ($T1$, $T2$). Because of the particular value that was chosen for the maximum heat transfer rate produced by the heaters, there is a very simple quantitative relationship between these three variables for each reservoir (assuming a stable volume). In the case of reservoir 1, if the heater is set to a value that is numerically equal to the target output demand value (i.e., if $HTR1 = D1$), then after the lags in the system have dissipated, the temperature in reservoir 1 will settle at a steady state value that is exactly equal to the target

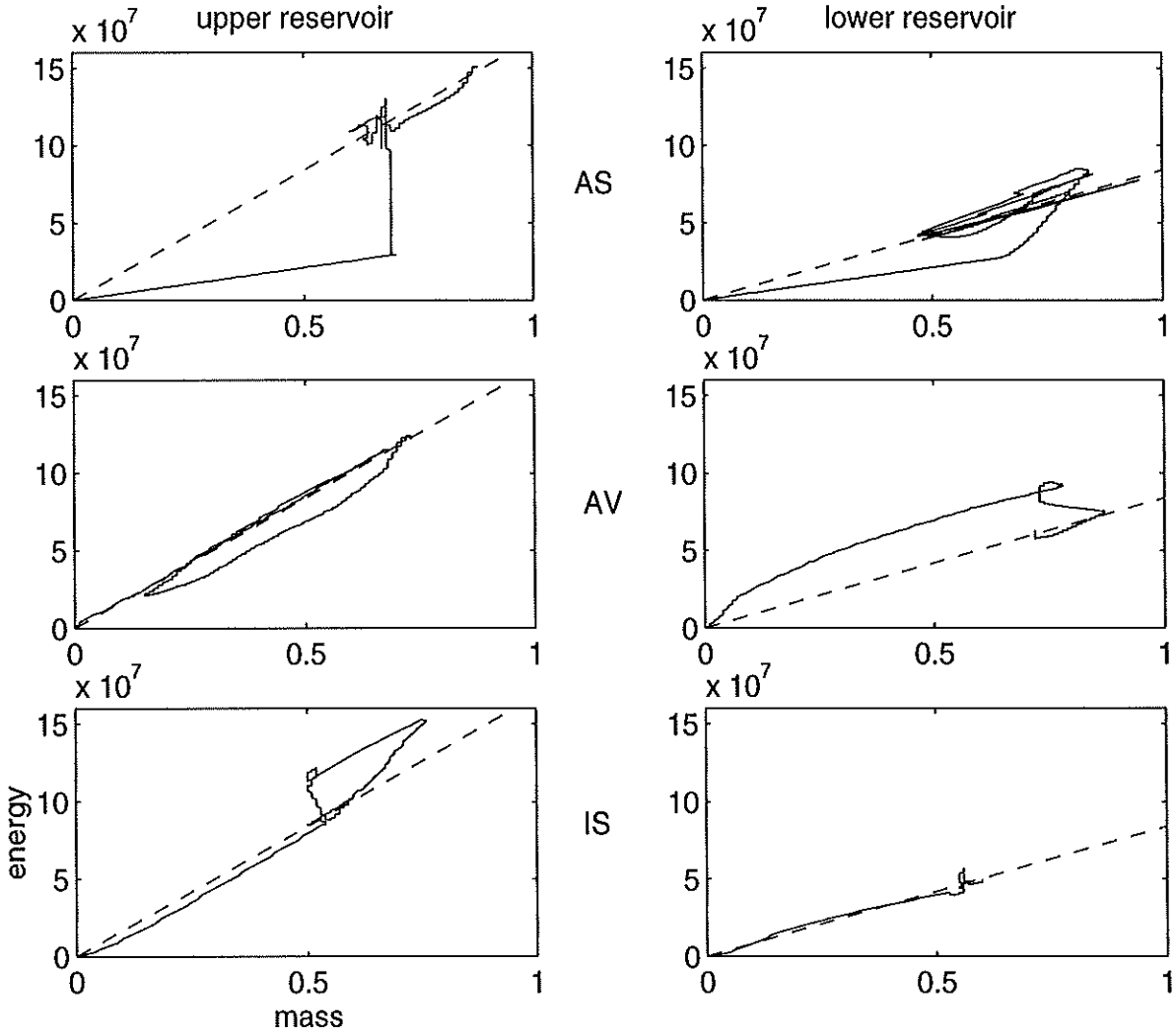

Figure 32: Graphs of mass vs. energy of P+F subjects at the first successful trial

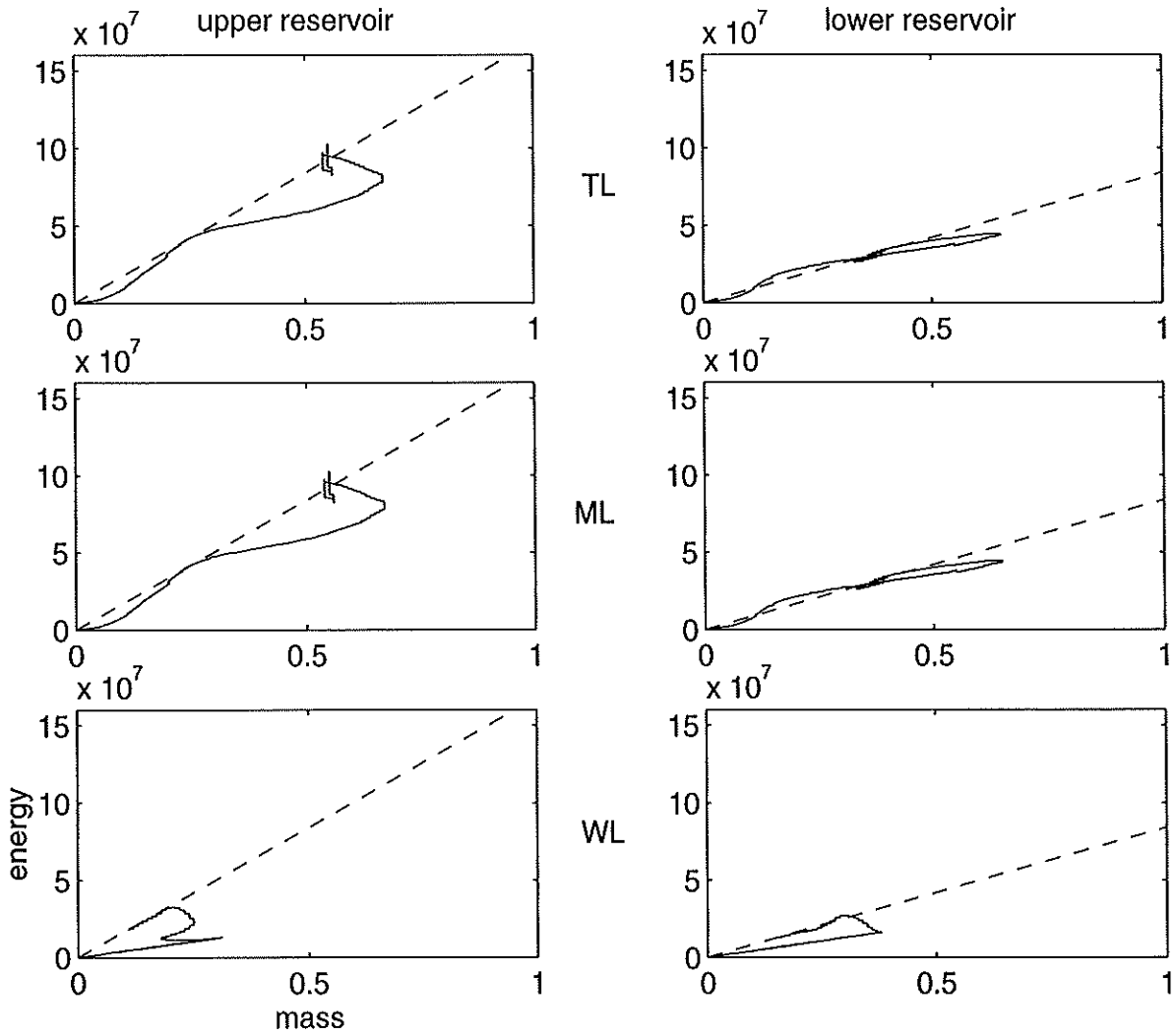

Figure 33: Graphs of mass vs. energy of P subjects at the first successful trial

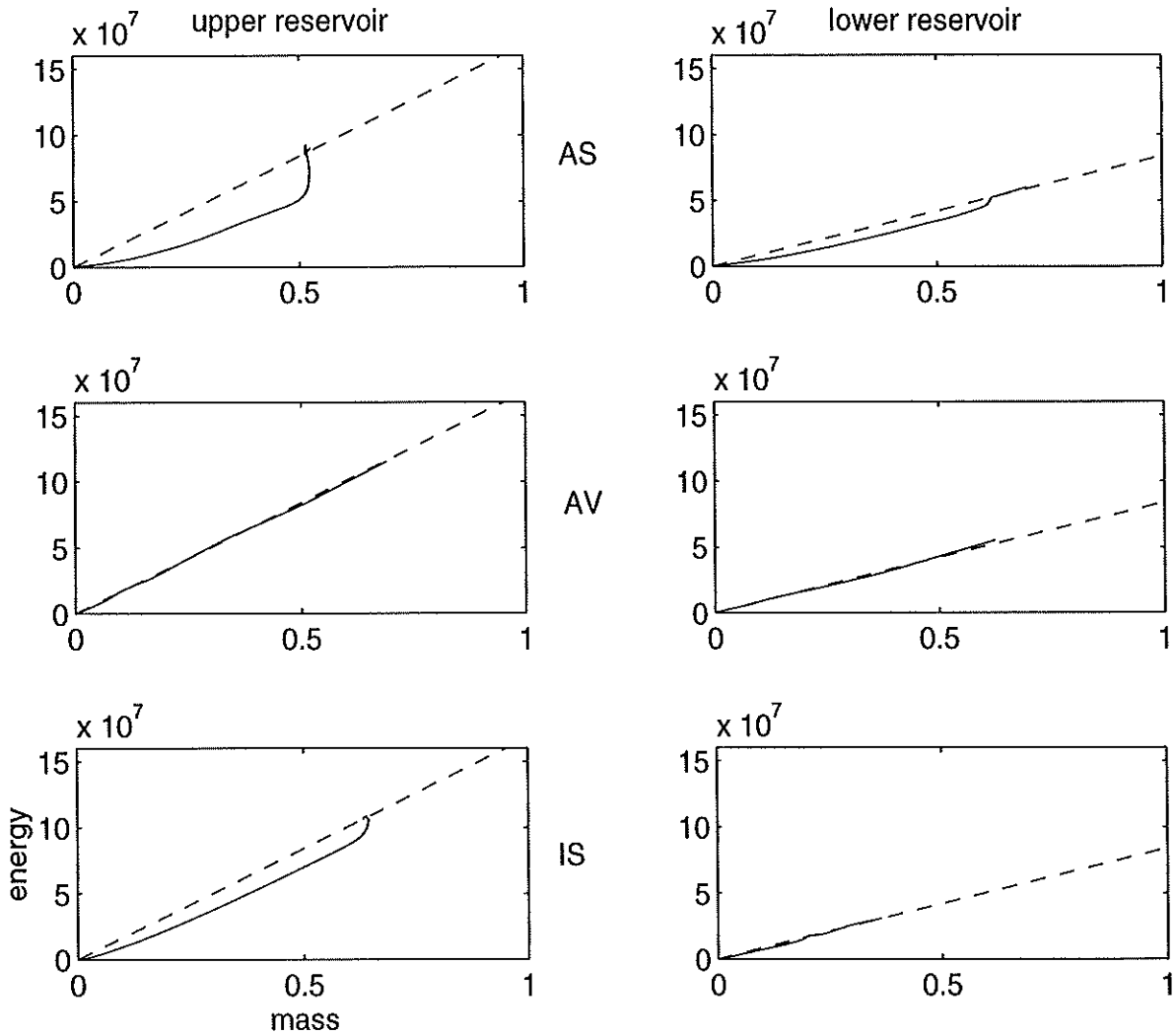

Figure 34: Graphs of mass vs. energy of P+F subjects at the last successful trial

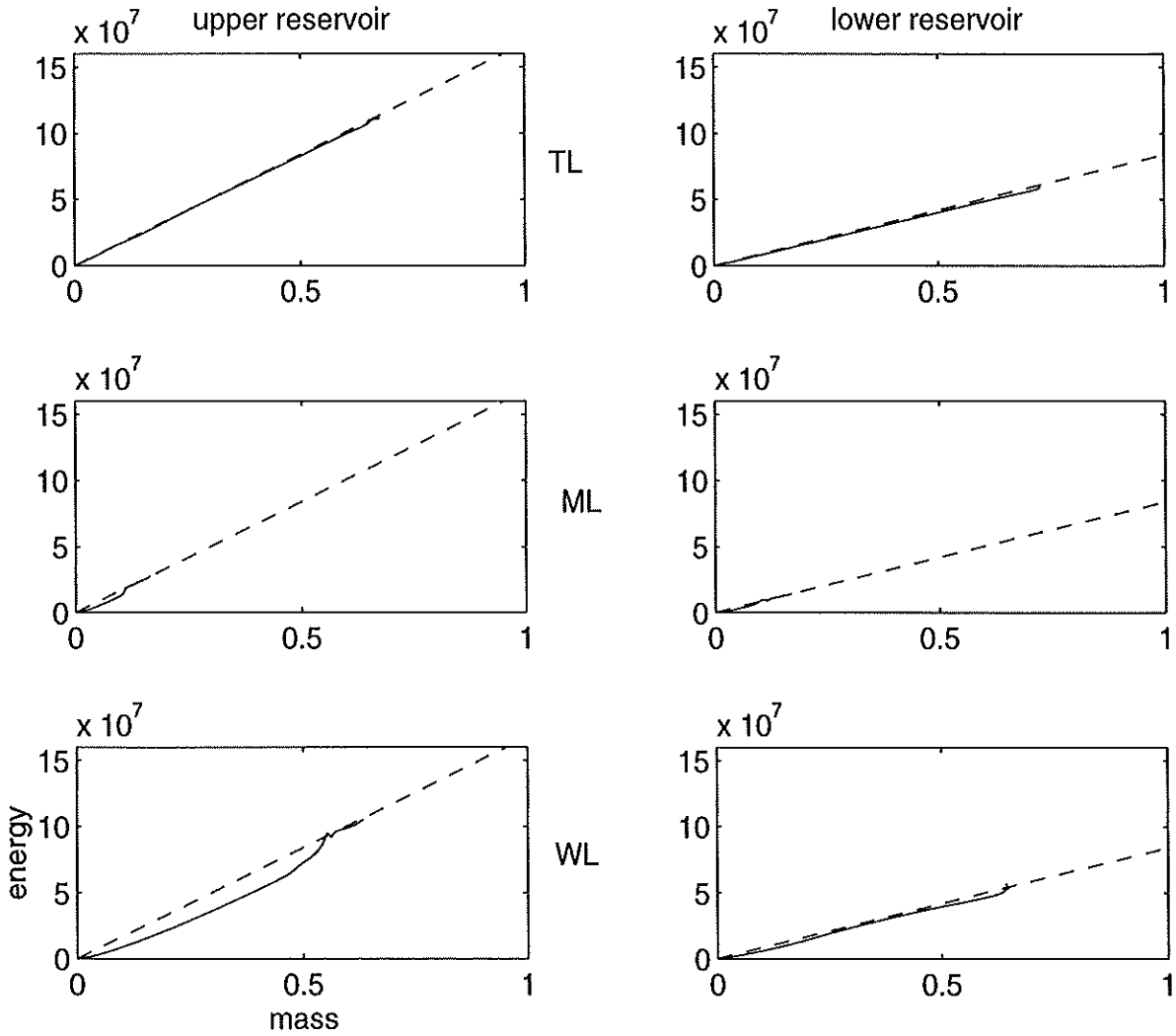

Figure 35: Graphs of mass vs. energy of P subjects at the last successful trial

temperature for that reservoir. In the case of reservoir 2, if the heater is set to a value that is numerically equal to one third of the target output demand value (i.e., if $HTR2 = D2/3$), then the temperature in reservoir 2 will settle at a steady state value that is exactly equal to the target temperature for that reservoir. These structural constraints can be used as referents to evaluate operators' actions. If the actions are tailored to these constraints, then we can say that operators are adapted to the structural properties of the system. Note that this information is not provided to subjects at any point during the experiment, although we know from the control recipe data that some subjects discovered these rules (Christoffersen et al., 1994).

To illustrate the degree of adaptation in subjects' actions, we plotted action frequency vs. heater to demand ratio for each reservoir for a particular block trials. Figures 36 and 37 illustrate the graphs for the P+F and P groups, respectively, for the first block of 22 trials. All subjects usually stabilized volume by the end of a trial, so their final heater settings would have to correspond to the ratios mentioned above if they were to complete the trial successfully. Thus, whether subjects were aware of these ratios or not, we would expect to see a large proportion of actions at the 1 and $\frac{1}{3}$ ratios for $HTR1$ and $HTR2$, respectively. Figures 36 and 37 show that the peak of the distribution for each subject is precisely at these values. There do not seem to be any consistent differences between the two groups of subjects early on in the experiment.

Figures 38 and 39 illustrate the action frequency vs. heater to demand ratio data for the last block of trials for the P+F and P groups, respectively. If we compare these graphs to the previous two, there are a number of interesting and informative differences. First, for all but one of the subjects (i.e., AV), we see that the peaks of the distributions remain at the expected ratios of 1 and $\frac{1}{3}$. However, in this last block, these peaks are higher than before, indicating that a larger proportion of subjects' actions are at the heater settings required to satisfy the temperature goals. Thus, as we would expect, subjects' behaviour became more adaptive with experience. Second, the peaks seem to be particularly prominent for IS, TL, and ML. Interestingly, these same three subjects were the only ones to report use of the "ratio rule" in their control recipes (Christoffersen et al., 1994). Thus, these subjects seem to be particularly attuned to the structural constraint between heater setting, demand value, and goal temperature.

Third, the behaviour of AV in the final block of trials (see Figure 38) is different from that of all of the other subjects. The peaks in his distributions are at ratio values that are greater than those that are required to achieve the temperature goals (i.e., at $1\frac{1}{3}$ and $\frac{1}{2}$, instead of at 1

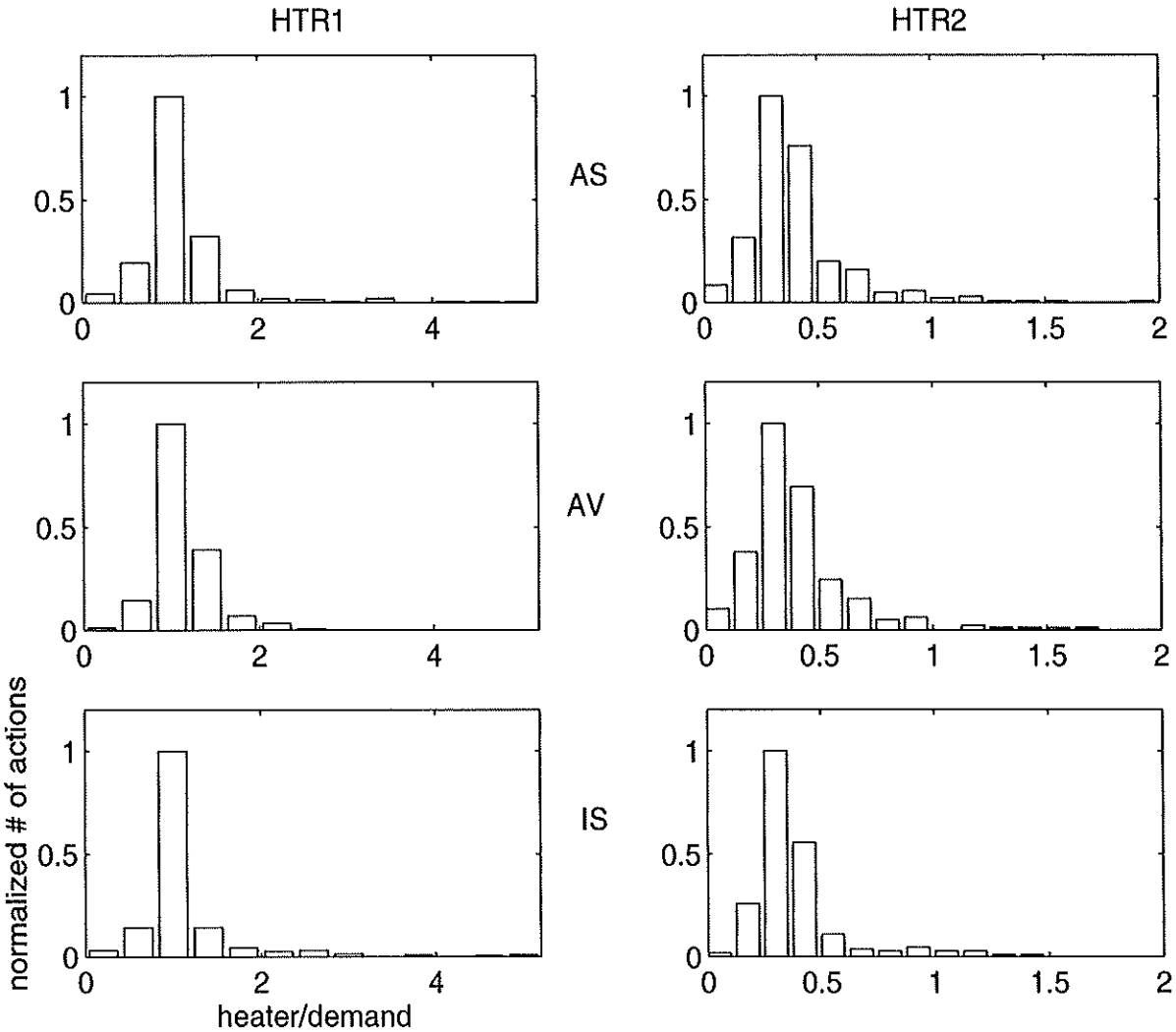

Figure 36: Distributions of actions on heaters of P+F subjects at the first block

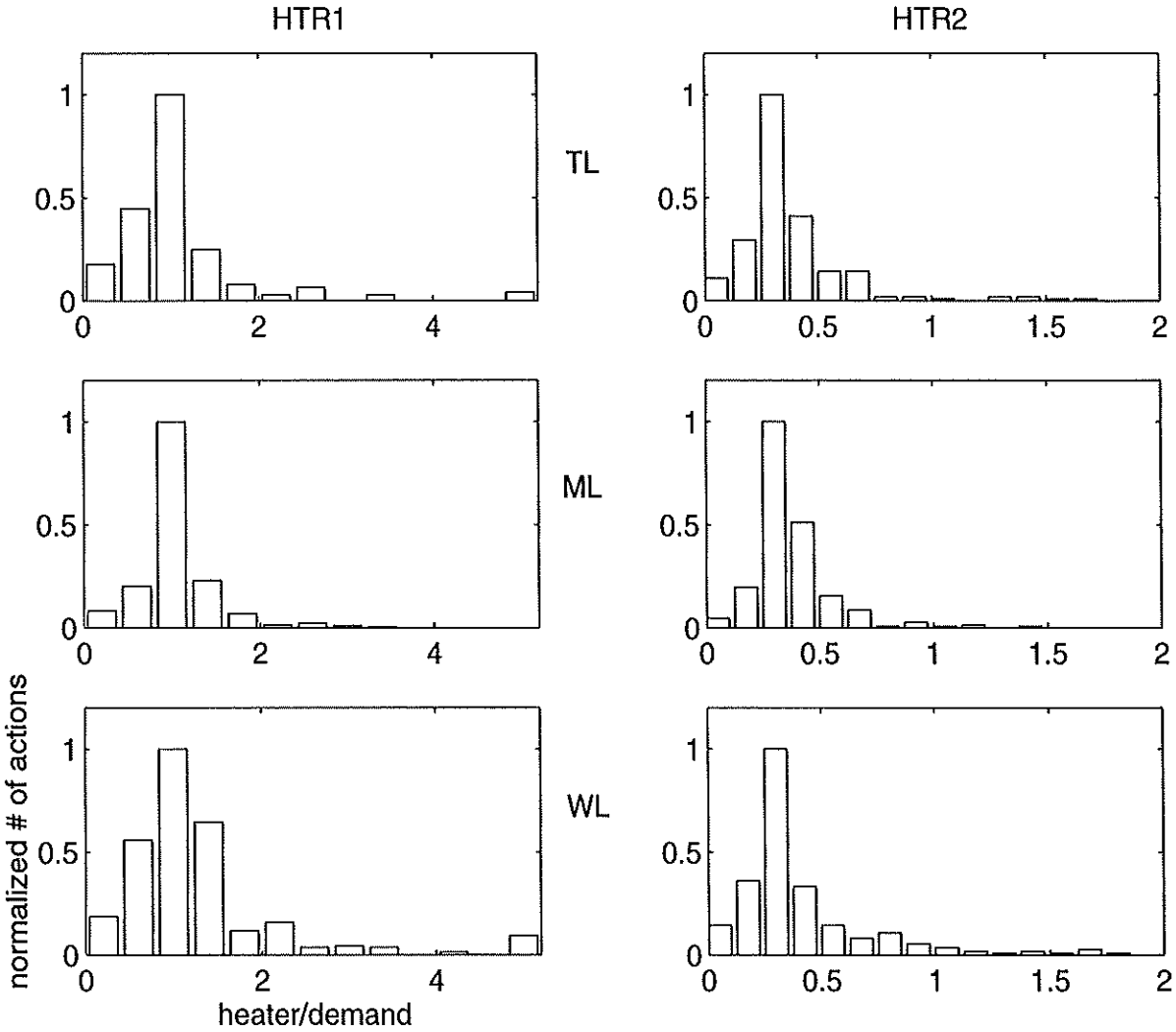

Figure 37: Distributions of actions on heaters of P subjects at the first block

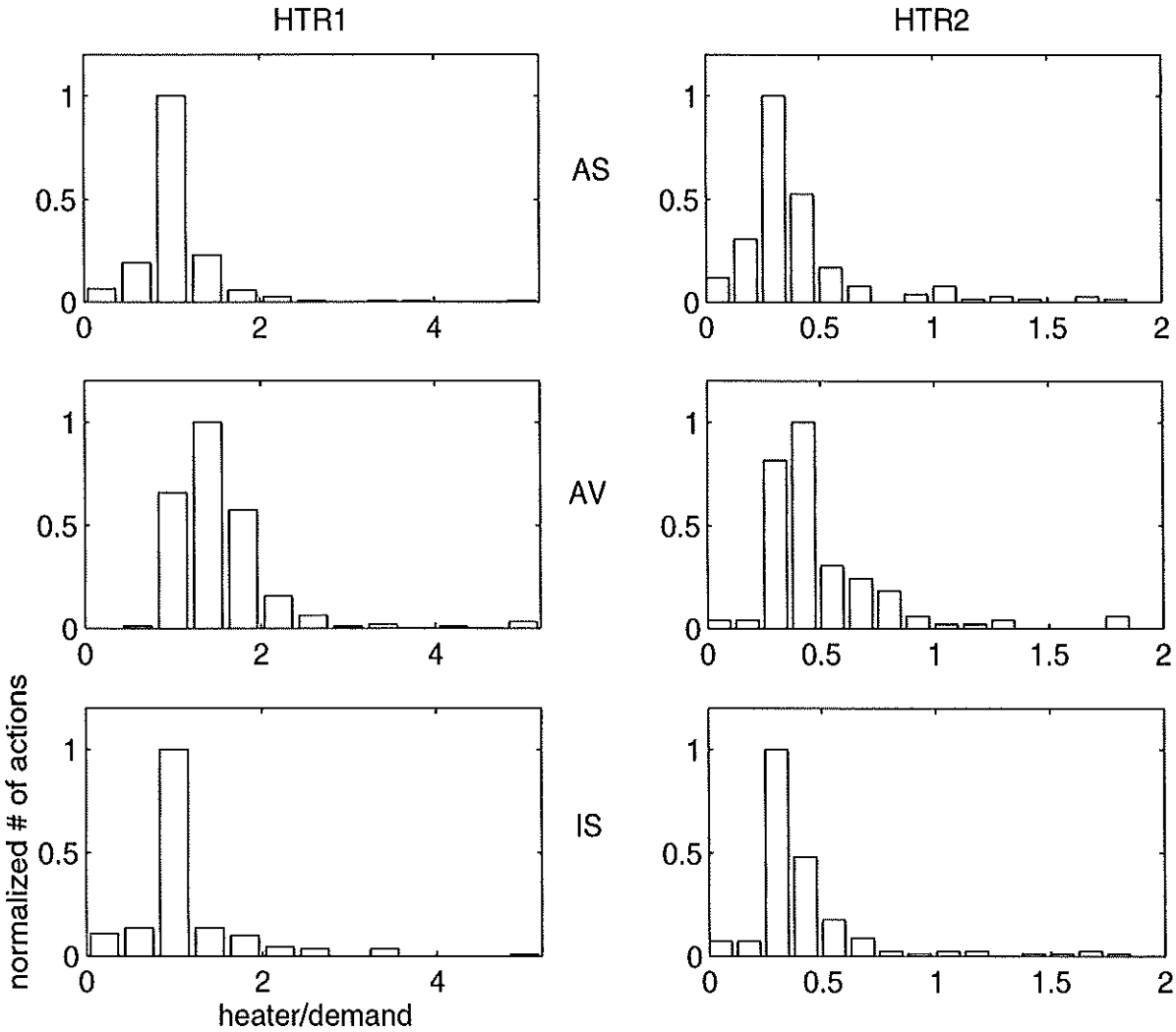

Figure 38: Distributions of actions on heaters of P+F subjects at the last block

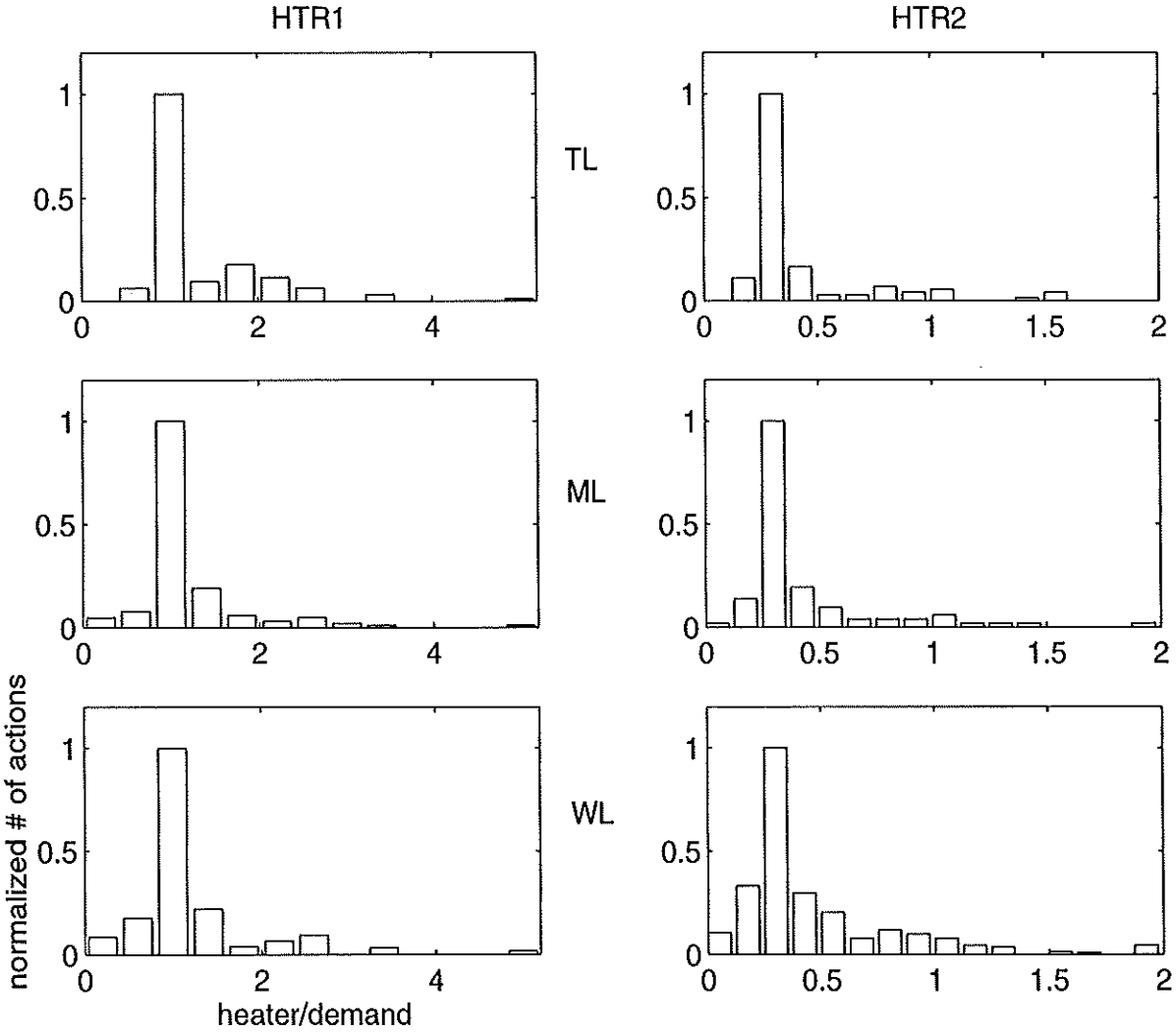

Figure 39: Distributions of actions on heaters of P subjects at the last block

and ⅓). Not only are the modes of AV's distributions greater than those of other subjects, but his distributions are generally more skewed to the right as well. Together, these patterns in the data indicate that AV very frequently sets the heaters to a higher value than they need to be to achieve the temperature goals. In fact, inspection of Figure 38 reveals that AV very rarely sets the heaters to a value that is below that required to achieve the goals. These results are consistent with a very aggressive control strategy, whereby the heaters are frequently set to a high level to get the reservoir temperatures rising more quickly than they otherwise would. Thus, AV seems to be exploiting the first order lag dynamics in the system to reach the goal states quickly. As we mentioned, no other subject exhibits the same type of behaviour (see Figures 38 and 39).

Choice of Steady State Volumes

Another structural feature of DURESS II that can serve as a referent for evaluating adaptation is the impact of steady state volume on the ease with which the temperature can be controlled. With very low volumes, it is more difficult to stabilize the temperature because the heater is very sensitive (i.e., a small change in setting causes a large change in temperature) and because the tolerance on the energy level is very narrow (see Pawlak & Vicente, 1996 for details). Larger volumes thus make it easier for subjects to maintain temperature in the goal region. This relationship is particularly strong for lower output demand values. We would expect that subjects who recognize this structural constraint could have an advantage over those who do not, especially in terms of shorter startup times and ease of control.

To determine how well adapted subjects were to this constraint, we plotted the final reservoir volume for each trial vs. trial number based on the assumption that this final value was indicative of the steady state volume. The results for the P+F and P groups are shown in Figures 40 and 41, respectively. Interestingly, most subjects seemed to have settled into a preferred range of steady state volumes. For example, ML kept his volumes quite low throughout the experiment. This pattern may help explain why he was frequently the worst subject in the P group. In contrast, TL who was the best performer in the P group, kept his volumes above the half-way point, particularly in the latter half of the study. As a result, TL was able to avoid the heater control problem we described above. WL also had relatively high volumes, but despite that, he was considerably less proficient than TL, as our previous analyses have indicated.

The most interesting pattern in the data for the P+F group was the change in strategy by the most proficient subject. As shown in Figure 40, AV started off the experiment with low volumes but, starting at about trial #100, he started using increasingly greater volumes. By the end of the experiment, he had settled on volumes of about ¾, thereby steering well clear of the heater control problem described earlier. Thus, AV seems to have learned to adapt to this structural constraint over the course of the experiment.

Discussion

These analyses of operator adaptation to structure have led to a number of interesting findings. First, as subjects gained more control experience, their behaviour tended to become increasingly tailored to the structural constraints governing the system. Second, the more proficient subjects (i.e., AV and TL) usually exhibited the strongest degree of adaptation across all subjects. Together, these findings support the view of expertise as adaptation to the goal-relevant constraints in the environment. Although this view of expertise has been held in the ecological psychology community for decades (cf. Gibson, 1991), it has only been gaining attention in mainstream cognitive psychology in recent years (Ericsson & Lehmann, 1996).

Third, the heater setting analyses showed that AV exhibited a more aggressive heater control strategy than any of the other subjects. This strategy may provide another reason why AV was so adept at quickly bringing the system up to the goal regions at the start of the trail (see the phase 1 analyses presented earlier). Although we have no direct evidence to support his claim, it seems likely that AV was able to do this because of the rich feedback offered by the P+F interface. Because he had higher-order functional information (e.g., mass and energy balances), AV could see, not just where the system was, but also where it was going. As a result, he could afford to start up the system aggressively while still being able to bring it back under control in a relatively efficient manner. In contrast, the best subject in the P group, TL, probably could not afford to adopt such an aggressive strategy because he did not have this functional information. Since he could not directly see where the system was going, he probably had to adopt a more conservative form of control if he was to bring the system under control efficiently and avoid oscillating in and out of the goal boundaries.

Fourth, the other two adaptation analyses presented in this section failed to show a substantial difference between the top performers in each group. For example, on the mass vs.

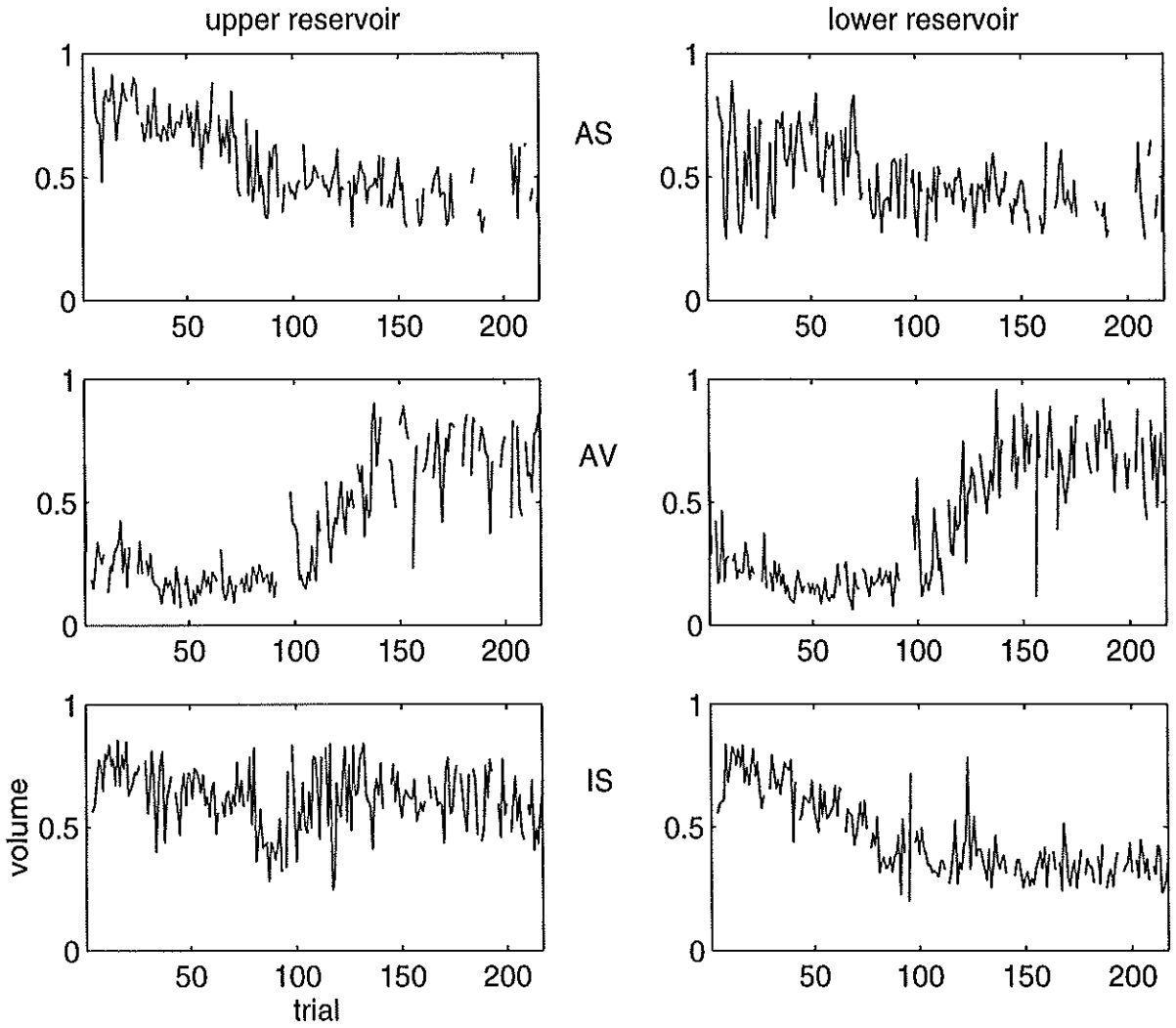

Figure 40: Reservoir volumes at steady state of P+F subjects

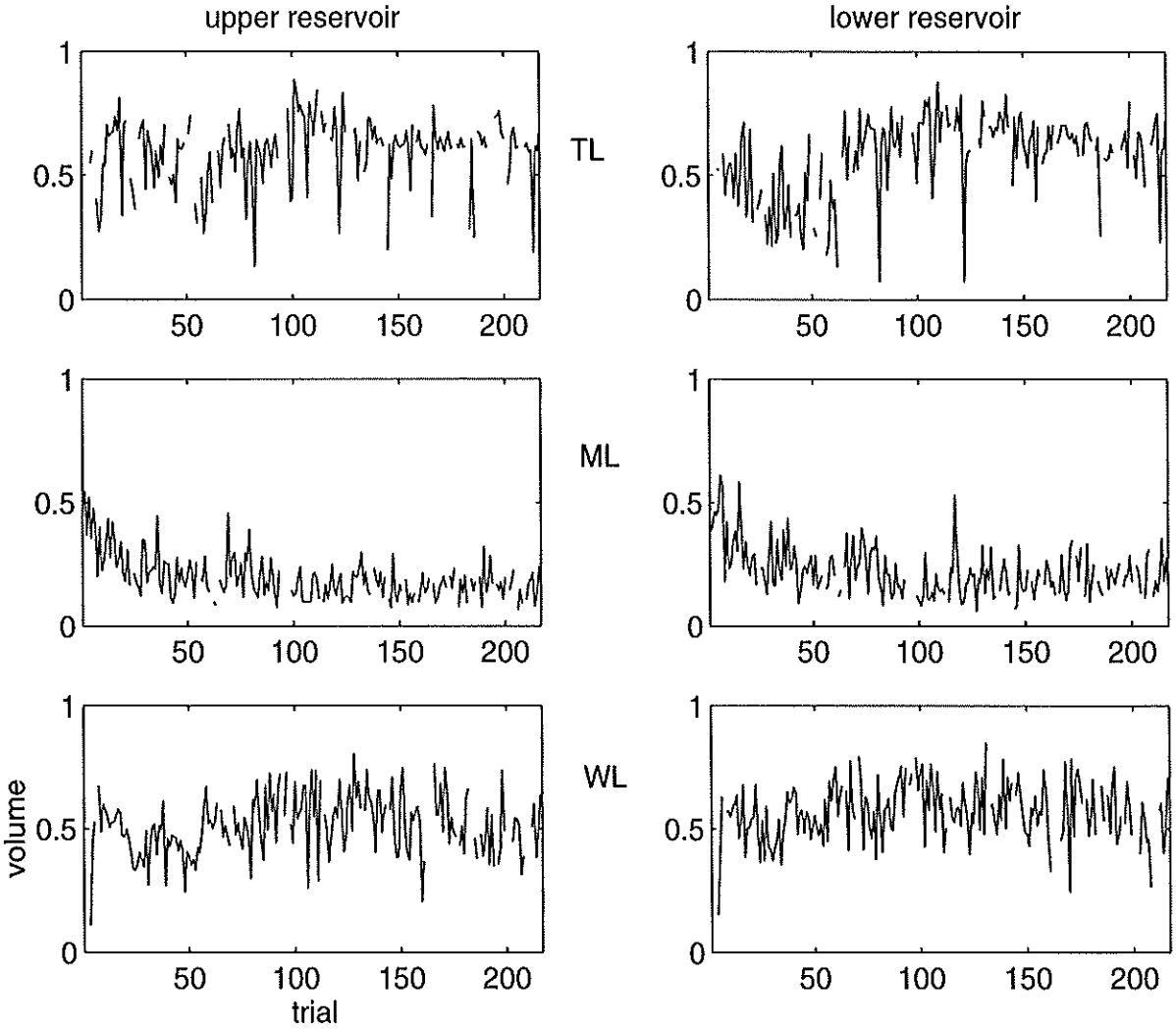

Figure 41: Reservoir volumes at steady state of P subjects

energy graphs, AV and TL were indistinguishable. Similarly, there was no noticeable difference between them in terms of steady state volumes by the end of the experiment. But as our previous analyses have shown, these two subjects were not controlling the system in the same way. Therefore, we can infer that it is possible to achieve a very high level of adaptation using qualitatively different control strategies. This is an important point that has not received sufficient attention in the cognitive engineering literature.

7. ANALYSES OF ADAPTATION TO INFORMATION PROCESSING LIMITATIONS

In the last two sections, we have presented analyses of operator adaptation to system state and structure. In this section, we will present analyses of the extent to which subjects also adapt their behaviour to their very own information processing limitations. The approach we take has been influenced by the work of Payne, Bettman, & Johnson (1993) on adaptive decision making. The fundamental insight behind this work is that people select their strategies based on an demand-resource tradeoff (see also Rasmussen, 1986). In many cases, this choice is adaptive because it allows people to deal with complex tasks despite the fact that they have a limited capacity for information processing. We will extend Payne et al.'s work by showing how their ideas can be applied to study expertise.

The Feedwater Stream Configuration Subtask

Based on a cognitive work analysis of DURESS II (Vicente & Pawlak, 1994), we have chosen to focus our analysis on a specific decision task, namely feedwater stream (FWS) configuration. In order to successfully complete a startup task, subjects must configure the FWS so that it supplies enough water to satisfy the output demand goals for both reservoirs ($D1$, $D2$). In the following analysis, we will assume that the goal of this task is to produce mass input flowrates that match the demands for each reservoir (i.e., $MI1 = D1$, $MI2 = D2$), thereby resulting in stable volumes. Although subjects need not stabilize volume to satisfy the demand goals (as long as they do not let the reservoir levels empty or spill over), most subjects seem to formulate the task in this way.

The means that subjects have available to them to perform this subtask are the two pumps (PA, PB) and the six input valves (VA, $VA1$, $VA2$, VB, $VB1$, $VB2$). For the most part, the use of the pumps is non-discretionary (they must be on for there to be any flow at all), so we will focus our analysis on the use of the valves. In particular, we would like to predict the final valve

settings that subjects settle on by the end of a trial, and to explain the criteria behind their choice. Note that we are not concerned with predicting the intermediate values that the valve settings take on during the trial.

The problem we are dealing with can be formulated as follows:

$$D1, D2 \rightarrow f(?) \rightarrow VA, VA1, VA2, VB, VB1, VB2$$

That is, given the output demand values for a particular trial, what function describes how subjects choose to configure the FWS components? The interesting feature of this problem is that it is underspecified. There are six unknowns, but only two givens. And as we will see, there are not enough system constraints to identify a unique solution. Thus, there are an infinite number of solutions to the problem. From a normative perspective, we can ask: what criteria should operators use to resolve the remaining degrees of freedom? From a descriptive perspective, we can ask: what criteria do operators use to resolve the remaining degrees of freedom?

The FWS configuration subtask can be formulated more precisely by breaking it down into three, nested steps:

1. How many valves will be used?
2. Which valves will be used?
3. At what quantitative value will each valve be set by the end of the trial?

We postulate that the way in which subjects answer each of these steps is constrained by two factors: first, the characteristics of DURESS II; and second, psychological criteria.

Normative Analysis

Based on Payne et al.'s (1993) work, we can conduct a normative analysis of this problem that is sensitive to psychological constraints. Our guiding assumption will be that adaptive operators would try to perform the FWS configuration task effectively while simultaneously trying to minimize the information processing demands they experience. Such a solution would require a sensitivity to both system constraints and psychological constraints. Since subjects cannot change the way in which DURESS II is built, they have no choice but to obey the system constraints if they are to be able to satisfy the task goals. Thus, we will describe these constraints first.

System constraints. What constraints does the system itself impose on the resolution of the FWS configuration subtask? We can answer this question in three steps, as a function of the three issues identified above. Answers to the first issue — how many valves will be used? — are constrained by the capacity of the two FWSs. Each pump can produce a maximum of 10 units/s of flow. As a result, each FWS can only supply 10 units/s of flow to the two reservoirs. Because of this structural system constraint, we can determine the minimum number of input valves (out of 6) that can be used to perform the task, as a function of the demand values (Vicente & Pawlak, 1994). Three, mutually exclusive and exhaustive cases can be distinguished:

- Mode 1 — If $(D1 + D2) \leq 10$, then min # of valves = 3
- Mode 2 — If $(D1 + D2) > 10$ AND $(D1 \leq 10)$ AND $(D2 \leq 10)$, then min # of valves = 4
- Mode 3 — If (D1+D2>10) AND (D2>10), then min # of valves = 5

Of course, subjects are free to use more than the minimum number of valves in any mode. Thus, there is no unique solution to this problem.

As for the second issue — which valves will be used? — the system imposes an additional set of structural constraints because of the topological connections between valves and reservoirs. More specifically:

- If $D1 > 0$, then (VA AND $VA1$) OR (VB AND $VB1$) must be used
- If $D2 > 0$, then (VA AND $VA2$) OR (VB AND $VB2$) must be used.

Once again, we see that the system constraints limit, but do not uniquely specify, what the operator should do.

As for the third and final issue — what value should each valve be set at? — there are several quantitative system constraints that must be taken into account:

- $MI1 = FA1 + FB1$ – conservation of mass into reservoir 1
- $MI2 = FA2 + FB2$ – conservation of mass into reservoir 2
- $FA = FA1 + FA2$ – conservation of mass in FWS A
- $FB = FB1 + FB2$ – conservation of mass in FWS B
- If $(VA1 + VA2) > VA$ then $FA = VA$ – flow is constrained upstream in FWS A
  else $FA = VA1 + VA2$ – flow is constrained downstream FWS A
- $FA1 = \dfrac{FA * VA1}{VA1 + VA2}$ – flow split relation in FWS B

- If $(VB1 + VB2) > VB$ then $FB = VB$ – flow is constrained upstream in FWS B
  else $FB = VB1 + VB2$ – flow is constrained downstream FWS B
- $FB1 = \dfrac{FB * VB1}{VB1 + VB2}$ – flow split relation in FWS B

These relationships perhaps appear to be more complicated than they really are. The first four are merely additive constraints that represent the fact that water is neither created nor destroyed in the system. The next four constraints describe the relationships between the valve settings and the flows through the valves. There are two sets of such relationships, one for the case where flow is being constrained downstream and another for when the flow is being constrained upstream. Each can be illustrated by example. In the downstream case, if the initial valve, VA, is set to 10 and $VA1$ and VA 2 are set to 1 and 4, respectively, then the flows will be as follows: $FA = 5$, $FA1 = 1$, and $FA2 = 4$. In the upstream case, if the initial valve, VB, is set to 9 and $VB1$ and $VB2$ are set to 4 and 8, respectively, then the flows will be as follows: $FB = 9$, $FB1 = 3$, and $FB2 = 6$. Thus, in the downstream case, the relationship between the valve settings ($VA1$, $VA2$) and the flows ($FA1$, $FA2$) is a simple one, namely 1:1. In the upstream case, on the other hand, this relationship is more complex, being determined by the ratio of the relative settings of the valves. As we will discuss below, this difference has important psychological implications. For now, the primary conclusion to keep in mind is that the problem of what values to set the valves at is underspecified. There are an infinite number of valve settings to satisfy a given output demand ($D1$, $D2$) pair.

Psychological constraints. The system constraints are not sufficient to provide a unique solution to the FWS configuration subtask. As a result, additional criteria must be introduced to deal with the remaining degrees of freedom. It is at this point that psychological constraints come into play. For each of the three issues identified earlier, we can propose normative psychological criteria that can be used by operators to resolve the remaining degrees of freedom in a way that minimizes the information processing demands imposed on them.

The first issue of how many valves to use is a relatively straightforward one (Pawlak & Vicente, 1994). Although there is no unique solution to the problem based on system constraints alone, psychological constraints can be used to narrow down the possibilities in a meaningful fashion. Each valve that is brought into play adds information processing demands. For example, operators would have to decide what value to set the valve at. Also, they would have to remember to monitor its state. Thus, it seems that an adaptive operator would minimize the

number of valves for each of the demand pair modes identified earlier. However, there is a potential disadvantage associated with using the minimum 3 valves in the first mode. The flows to the two reservoirs become coupled because a single FWS is being used to feed both of them. This makes it more likely that the upstream valve mode will be experienced. As a result, a change in valve setting (e.g., *VA*1 or *VA*) to affect the flow going to one reservoir (e.g., reservoir 1) could unintentionally affect the other reservoir (e.g., reservoir 2). Using 4 valves obviates this problem because each FWS would be driving a single reservoir. There would be no interactions to worry about. Although it is difficult to be sure, it seems that the disadvantages of using only 3 valves outweigh the advantages. Thus, for Mode 1, we would predict that adaptive operators would probably use 4 valves. For Mode 2, they should use the minimum 4 valves, and for Mode 3, they should use the minimum 5 valves.

The next issue is which valves should be used, given that their number has been established. The systems analysis (see above) showed us that there are basically two options, so once again there is no unique solution. However, a cursory examination of the degrees of freedom allows us to narrow down the solution based on psychological criteria. In this case, we can rely on stimulus-response compatibility. For one of the options, the upper (or lower) stream would feed the upper (or lower) reservoir. For example, *VA* & *VA*1 would be used to feed reservoir 1. For the other option, the upper (or lower) valve would feed the lower (or upper) reservoir. For example, *VA* & *VA*1 would be used to feed reservoir 2 instead. Clearly, the first option is far better since it makes control of the system much more natural because of the directness of the mapping (i.e., to control the top reservoir, use the top stream). In contrast, the second option – while equally viable from a systems point of view – is problematic from a psychological viewpoint. It would result in an inverse mapping (e.g., to control the bottom reservoir, use the bottom stream). Thus, the choice for the adaptive operator seems clear.

The final issue is what quantitative values should the valves be set at. In this case, the systems analysis showed us that there are a number of algebraic constraints that must be taken into account in making this decision. Nevertheless, there is still a bounded but infinite number of possibilities. How can we narrow down this number? Again, we can introduce additional psychological criteria to deal with the remaining degrees of freedom. The relevant criterion in this case is to minimize the computational effort involved. Recall that the FWS can operate in two qualitatively different ways, one in which flow is constrained upstream and the other in

which flow is constrained downstream. The upstream case is more complex, involving flow split ratio calculations, whereas the downstream case is considerably simpler because of the identity relationship between valve setting and flow rate. Note that these two cases are equivalent in terms of the range of demand pairs they can satisfy so, from a systems perspective, there is no difference between them. But since the downstream configuration imposes considerably fewer computational demands (see below), it should be always be adopted by the adaptive operator.

Summary. This normative analysis of the FWS configuration task has spanned both system and psychological constraints. We began by identifying the constraints that the system imposes on successful performance, since these constraints must be respected if there is any hope all of performing the task successfully – the laws of physics can be ignored but not escaped. We described a number of systems constraints that must be taken into account, but we also learned that there are many ways of accomplishing the task, an infinite number in fact. From a systems perspective, there is no basis for choosing between these different solutions – they are all equivalent. But since DURESS II is being controlled by human operators, we considered how the remaining degrees of freedom can be closed by introducing psychological criteria. Our assumption was that an adaptive operator would select solutions that, not only get the job done, but get it done as economically and easily as possible. Table 5 summarizes the criteria we identified for each of the issues associated with the FWS configuration subtask. These criteria help us identify, out of all the feasible ways of performing the task, the way(s) that an adaptive operator is most likely to adopt.

Degrees of Freedom	Psychological Criteria
How many valves?	minimize monitoring demands and interactions
Which valves?	stimulus-response compatibility
What settings?	minimize computations

Table 5: A summary of the psychological criteria used for the normative analysis of the FWS configuration subtask.

If we integrate all of the considerations summarized in Table 5 and add in the fact that D1 never exceeded 10 in our experiment, we are able to use up all of the degrees of freedom in performing the task. In other words, we can identify a normative algorithm that an adaptive operator might use to accomplish the FWS configuration subtask. To show that this is the case, we again borrow from the work of Payne et al. (1993) by defining a set of elementary

information processes (EIPs) that can be used as a basis for the specification of an algorithm. Table 6 shows the EIPs that we developed for this particular task. Table 7 shows how these EIPs were used to compose a normative algorithm for the FWS configuration subtask, based on the criteria we discussed above. This algorithm provides one possible to answer to the question we posed at the outset – what function does (or should) an operator use to go from a set of output demand pairs to a set of valve settings?

EIP	Description
READ	Read a value into short-term memory.
ADD	Add two values.
DIVIDE	Divide one value by another.
SUBTRACT	Subtract one value from another.
COMPARE	Compare two values (greater than or less than).
ASSIGN	Finding appropriate component and assigning a setting to it.

Table 6: EIPs for performing the FWS configuration subtask in DURESS II. The READ, ADD, and COMPARE EIPs were borrowed from Payne et al. (1993).

```
ASSIGN      VA = 10
ASSIGN      VB = 10
READ        D1
READ        D2
ASSIGN      VA1 = D1
COMPARE     D2 to 10
    if D2 ≤ 10
        ASSIGN      VB2 = D2
    else (D2 > 10)
        ASSIGN      VB2 = 10
        SUBTRACT    D2 – 10
        ASSIGN      VA2 = (D2 – 10)
```

Table 7: Normative algorithm for the FWS configuration subtask, based on the adaptive criteria discussed above.

Results

The obvious question is: Did any of the subjects in our experiment use the algorithm described in Table 7? In this subsection, we address this question by comparing subjects' actions with the predictions made by the normative analysis. We calculated an error score to measure the difference between the valve settings used by subjects and the valve settings predicted by our analysis. An error score of zero would indicate perfect agreement with the normative analysis. We then plotted this error score vs. trial number for each subject. The results are shown in Figure 42. We can see that the normative analysis predicts the behaviour of TL quite well. Although he started off with a different strategy, the behaviour of IS is also well accounted for during the latter third of the experiment. As for AV, there is good agreement during the first third of the experiment and moderate agreement during the remainder. The data for the remaining subjects (AS, ML, and WL) clearly show that they were not following the normative strategy.

Interestingly, the degree of fit to the adaptive norm seems to best for the most proficient subjects. TL was clearly the best in the P group, AV usually the best in the P+F group, and IS a close second. This observation suggests that there may be a connection between subjects' performance and the extent to which their behaviour was adapted to the system constraints <u>and</u> their own information processing limitations. In the next subsection, we will pursue this conjecture in more detail.

Descriptive Analysis

Using the three issues identified in Table 5, we conducted more detailed descriptive analyses to inductively determine to what extent subjects were following the adaptation criteria we identified in our normative analysis. This, in turn, should allow us to identify what strategies subjects were following. We began with the issue of how many valves to use. Figures 43, 44, and 45 plot the number of valves subjects used per trial for trials with demand pairs of Mode 1, Mode 2, and Mode 3 (see above), respectively. Recall that the minimum number of valves for Mode 1 is 3. Figure 43 shows that no subject ever used this strategy. Nevertheless, the data in Figure 43 reveal several interesting facts. First, subjects were remarkably consistent in terms of the number of valves that they used. There was no learning whatsoever. The number of valves that subjects started off using was the number that they kept on using for the duration of the

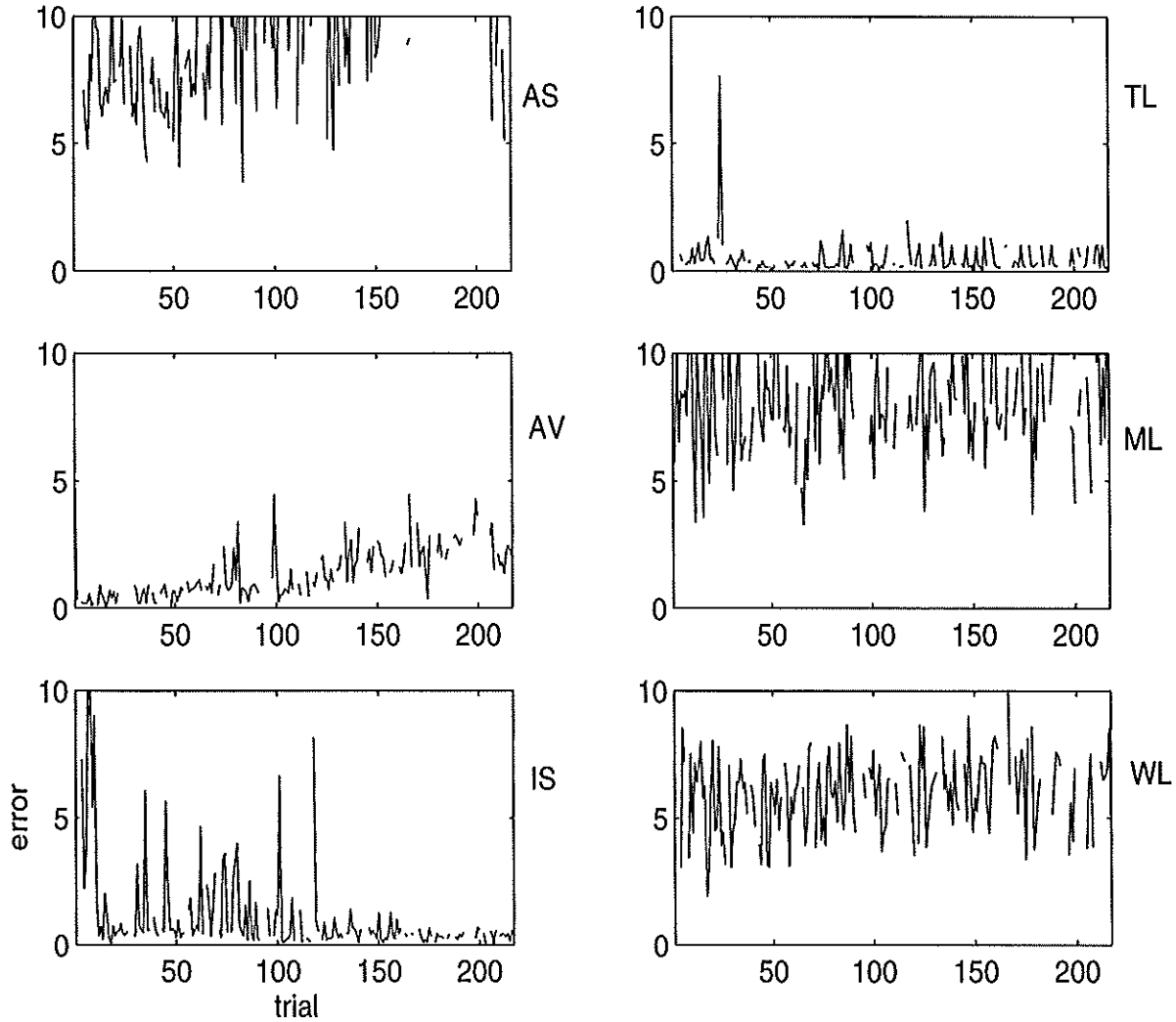

Figure 42: Error between practical valve settings and those predicted by Table 7

experiment. Second, three subjects (AV, IS, and TL) consistently used only 4 valves. Given our normative analysis, it seems that these subjects adopted the criteria of minimizing monitoring demands and interactions in determining how many valves to use. The other three subjects (AS, ML, and WL) almost always used all 6 valves to configure the FWS. They did not seem to adapt their strategy to the unique possibilities associated with Mode 1 demands.

Figure 44 shows the data for the trials with Mode 2 demand pairs. Recall that the minimum number of valves for this Mode is 4. As with the previous graph, we see that all subjects were very consistent in the number of valves that they decided to use. AV, IS, and TL generally used only 4 valves to perform the task, just as in Mode 1. In contrast, AS, ML, and WL used all 6 valves for Mode 2 trials, even though they could have accomplished the task with only 4. Thus, these subjects did not exploit the economies made possible in Mode 2 demands.

Figure 45 shows the results for the trials with Mode 3 demand pairs. The minimum number of valves for this Mode is 5. Once again, subjects were very consistent in their actions. AV, IS, and TL always used only 5 valves to perform the task. On the other hand, AS, ML, and WL almost always used 6 valves, one more than was necessary to complete the task. Collectively, these results show that AV, IS, and TL adapted their actions according to the demands imposed by the demand pairs, thereby minimizing their information processing requirements whenever the trials allowed it. In contrast, AS, ML, and WL showed no such sensitivity. Their actions were consistent, independent of the demand pairs. As a result, they unintentionally experienced greater information processing demands than the other subjects.

A second set of analyses was conducted to address the issue of which valves were used. This analysis was only pertinent to AV, IS, and TL because the other subjects always used all 6 valves. The results were very straightforward. When these subjects used only 4 valves, *VA*2 and *VB*1 were always the valves that were not used. When they used 5 valves, *VB*1 was always the one that was not used. These findings are in perfect agreement with the stimulus-response compatibility criterion identified earlier (see Table 5). Subjects always chose the direct mapping, thereby avoiding the non-intuitive reverse mapping.

A third set of analyses was conducted to tackle the question of what valve settings were used. As indicated by Figure 42, the data for AV, IS, and TL were well-captured by the normative strategy. We tried to inductively identify algorithms that would capture the behaviour of the other three subjects by replaying their trials and by relying on their verbal reports and

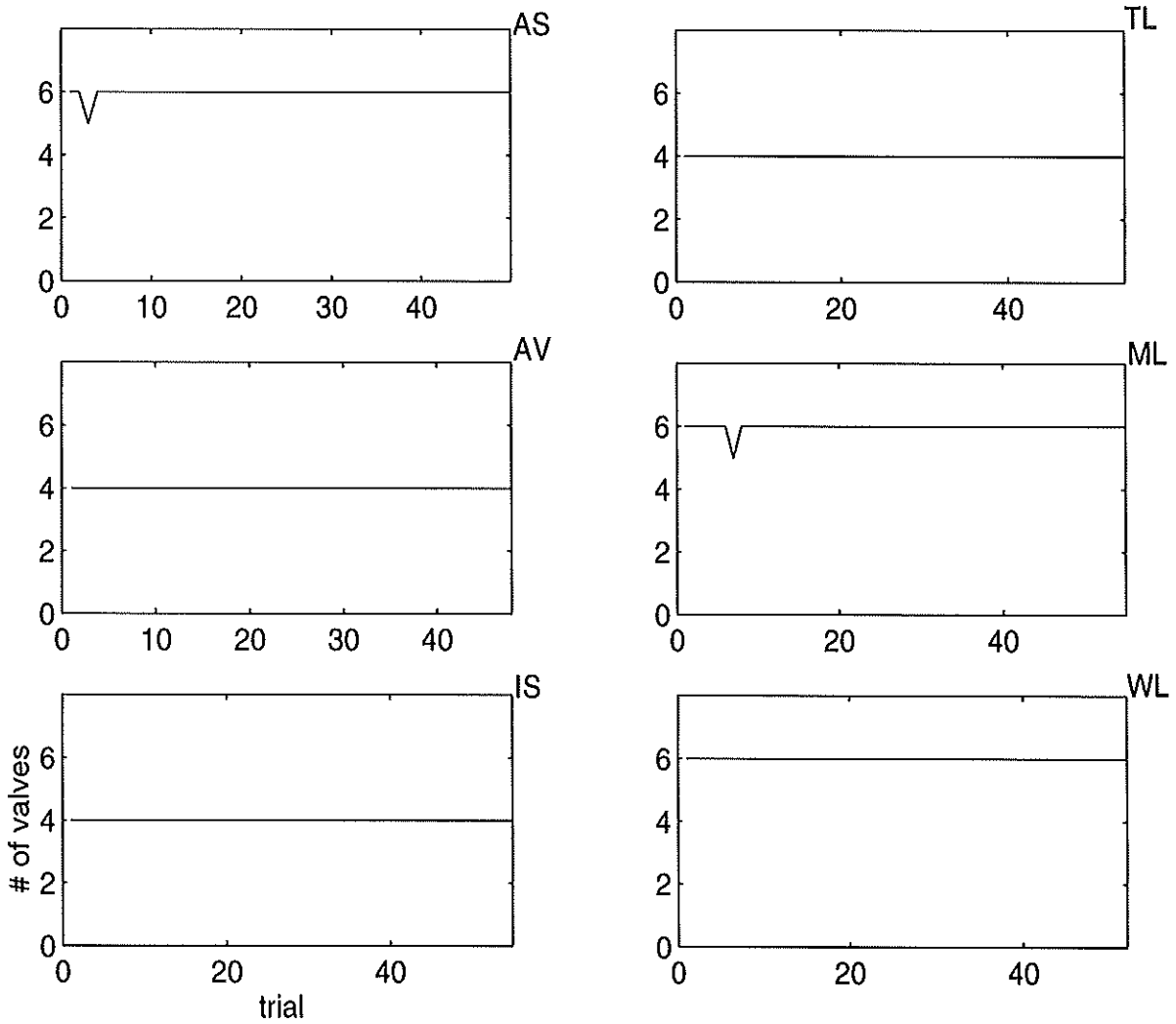

Figure 43: Number of valves used at trials with demand pairs of mode 1

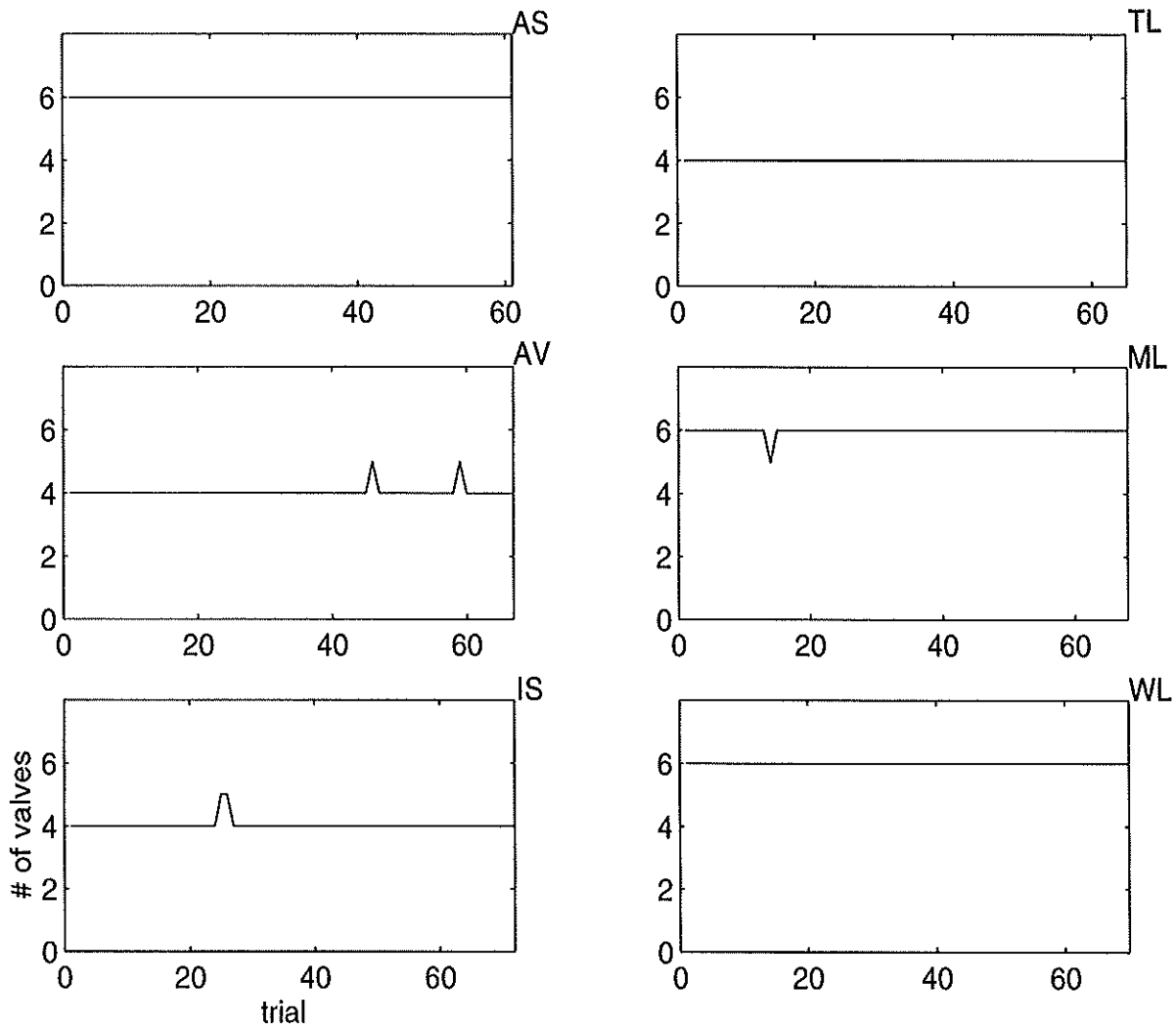

Figure 44: Number of valves used at trials with demand pairs of mode 2

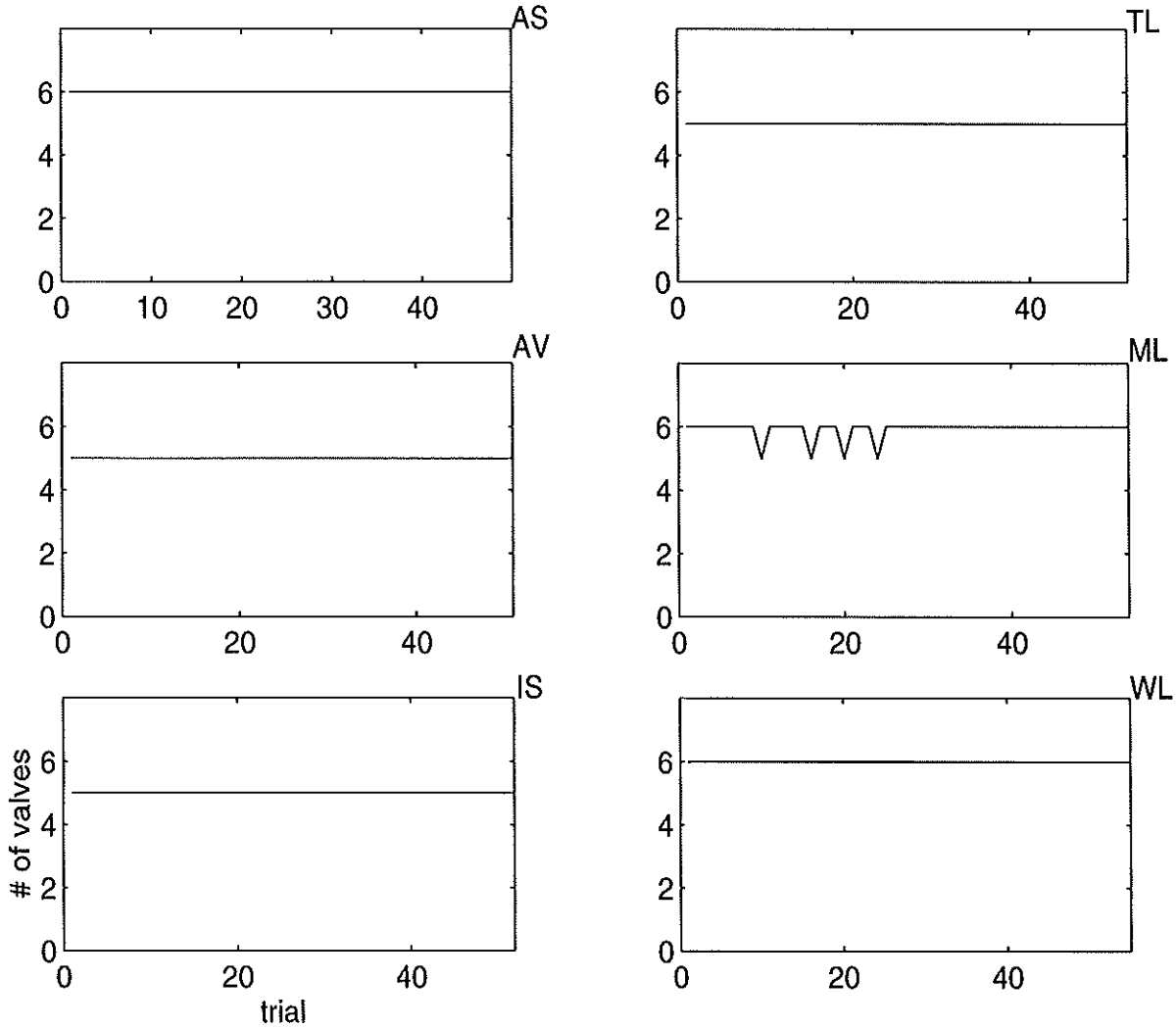

Figure 45: Number of valves used at trials with demand pairs of mode 3

control recipes (Christoffersen et al., 1994). The algorithm that was identified for ML is listed in Table 8, again using the EIPs in Table 6.

```
READ        D1
READ        D2
ADD         D1 + D2
DIVIDE      (D1 + D2) / 2
ASSIGN      VA = (D1 + D2) / 2
ASSIGN      VB = (D1 + D2) / 2
COMPARE     D2 to 10
    if D2 ≤ 10
        ASSIGN      VA1 = D1
        ASSIGN      VB1 = D1
        ASSIGN      VA2 = D2
        ASSIGN      VB2 = D2
    else
        DIVIDE      D1 / 2
        ASSIGN      VA1 = (D1 / 2)
        ASSIGN      VB1 = (D1 / 2)
        DIVIDE      D2 / 2
        ASSIGN      VA2 = (D2 / 2)
        ASSIGN      VB2 = (D2 / 2)
```

Table 8: The algorithm inductively developed to account for ML's strategy for performing the FWS configuration subtask.

To see how well this algorithm accounted for the subjects' behaviour, we again calculated an error score and plotted it against trial number. The results for all 6 subjects are given in Figure 46. As we expected, the algorithm in Table 8 provided an excellent fit to the data of ML, beginning at around trial #100. The fit with the data for most of the other subjects was quite poor. The sole exception was WL, whose behavior on certain trials was moderately well captured by the algorithm in Table 8. As we will see, this occasional fit occurs because WL's strategy is similar to ML's under certain conditions.

The algorithm we inductively identified for WL is listed in Table 9. Figure 47 shows how well this algorithm fit the subjects' data. The agreement with WL's behaviour is very good and

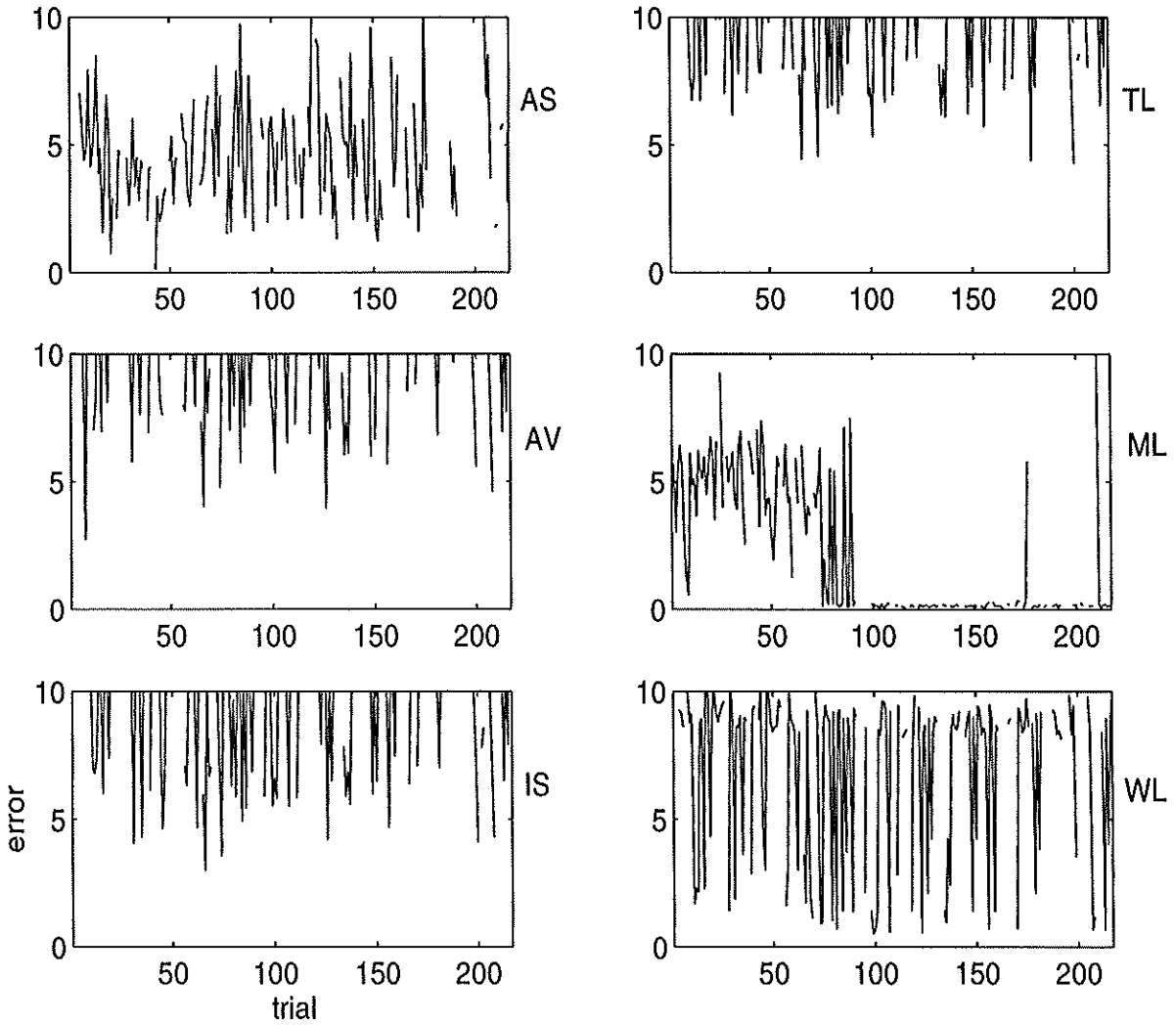

Figure 46: Error between practical valve settings and those predicted by Table 8

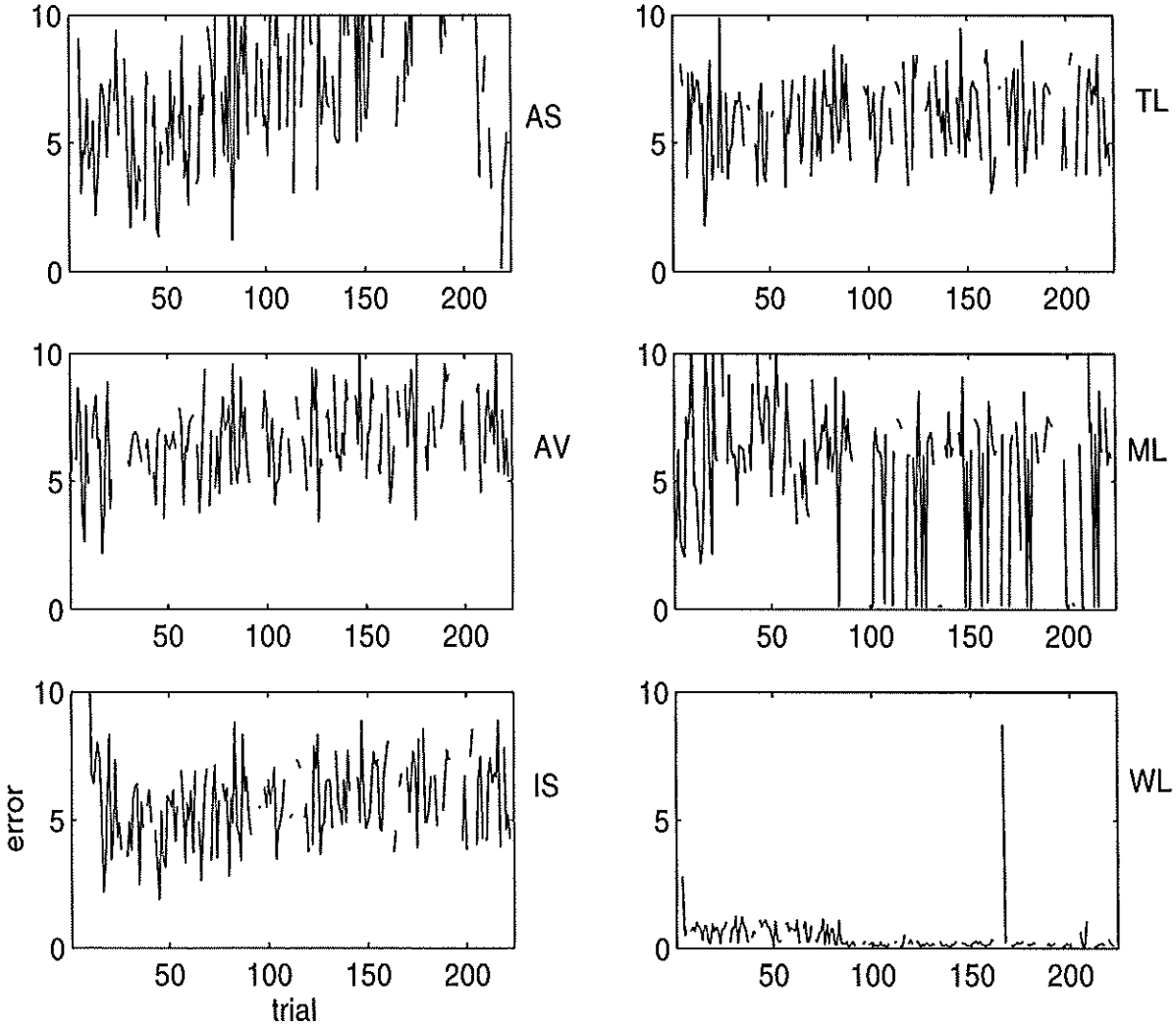

Figure 47: Error between practical valve settings and those predicted by Table 9

quite consistent. ML's behaviour is well accounted for on some trials for the reason already mentioned. The fit to the behaviour of all of the other subjects is clearly inadequate.

```
ASSIGN      VA = 10
ASSIGN      VB = 10
READ        D1
READ        D2
COMPARE     D1 to 1
COMPARE     D2 to 1
    if D1 = 1
        ASSIGN      VA1 = 1
        DIVIDE      D2 / 2
        ASSIGN      VA2 = D2 / 2
        ASSIGN      VB2 = D2 / 2
    if D2 = 1
        ASSIGN      VA2 = 1
        DIVIDE      D1 / 2
        ASSIGN      VA1 = D1 / 2
        ASSIGN      VB1 = D1 / 2
    else (D1 & D2 > 1)
        DIVIDE      D1 / 2
        ASSIGN      VA1 = D1 / 2
        ASSIGN      VB1 = D1 / 2
        DIVIDE      D2 / 2
        ASSIGN      VA2 = D2 / 2
        ASSIGN      VB2 = D2 / 2
```

Table 9: The algorithm inductively developed to account for WL's strategy for performing the FWS configuration subtask.

We also tried to inductively determine what strategy AS was doing, but we were unable to do so. Figures 42, 46, and 47 clearly reveal that AS was not following any of the previously identified algorithms. Although it is possible that there is some unknown consistency in his behaviour, as far as we were able to tell, his actions were driven by a trial and error strategy across trials.

The differences in the ways in which the subjects approach the issue of valve settings can be summarized by tabulating the number of EIPs associated with the algorithms listed in Tables 7 to 9. The results are listed in Table 10. Because the issue of settings is nested under the issues of how many valves to use and which valves to use, this summary actually represents the overall differences in strategies across subjects. We can see that the strategy used by AV, IS, and TL is more adaptive in the sense that it minimizes the number of EIPs required to do the task. The other subjects are doing more work than is required to perform the task.

EIP	AV, IS, & TL	ML	WL
ADD	0	1	0
ASSIGN	4 or 5	6	5 or 6
COMPARE	1	1	2
DIVIDE	0	1 or 3	1 or 2
READ	2	2	2
SUBTRACT	0 or 1	0	0
TOTAL	7 or 9	11 or 13	10 or 12

Table 10: A comparison of the EIP count for the FWS configuration subtask strategies for all subjects, except for AS.

Although there are probably differences in the amount of effort associated with different EIPs (e.g., READ is probably easier than DIVIDE), we have not tried to weight the different types of EIPs to account for these differences (cf. Payne et al., 1993). If weights are added, it may very well be that the differences in information processing demands between the strategies in Table 10 are even greater than the totals would indicate.

Discussion

For goal-oriented behaviour to be successful, it must be sensitive to the constraints imposed by the environment (in this case, DURESS II). Operators cannot change the physical principles governing the behaviour of the system. They have to work within them. In short, they have to adapt to the structure of the system if they are going to be proficient performers. We saw empirical evidence of this in the previous section.

Having said that, there are many situations in complex work domains where the constraints in the environment underspecify how a particular task is to be done. There can be many different

ways of performing the very same task. From a systems perspective, these alternative paths are functionally equivalent because they all get the job done. Thus, operators are left with choices about how to perform the task, given all of the degrees of freedom that are still left over. The important point, however, is that operators generally only pick a very small number of the solution paths to perform the task, out of all of the ones that are available. Even though there may be a bounded but infinite number of ways of doing the task, operators will not exhibit the vast majority of these behaviours. This is a very important insight because it suggests that there may be some systematic criteria that operators are using to select away the remaining degrees of freedom. Following Payne et al. (1993), we hypothesized that adaptive behavior involves performing the task successfully but in a way that minimizes the information processing demands. In other words, operators adopt psychological criteria to govern strategy choice, thereby dealing with the available degrees of freedom in a way that makes the task more manageable to perform.

The analyses we have presented in this section clearly support this view. Almost all of the subjects exhibited very consistent strategies for dealing with the FWS configuration subtask. A normative analysis revealed that it is possible to introduce psychological criteria to remove essentially all of the available degrees of freedom, thereby leading to the identification of a minimal effort solution to the task. By comparing this normative algorithm to the data exhibited by subjects throughout the experiment, we found that three of the subjects showed signs of following this algorithm. AV, IS, and TL all adapted their behaviour to get the task done in a way that was consistent with the psychological criteria we had identified. In short, these subjects were adapted, not only to the system constraints, but to their own information processing limitations as well. They tried to make the task as simple as possible.

We also conducted a descriptive analysis to see if we could identify the strategies being used by the remaining three subjects. In one case, AS, we were not successful. There does not seem to be any consistent pattern in the way in which AS deals with the FWS configuration subtask across trials. In the other two cases, ML and WL, we had more success. Algorithms were developed that provided very good fits to the behavioural data. An examination of the information processing demands imposed by these algorithms shows, however, a poor degree of adaptation to minimal effort criteria. These subjects were performing the task in a way that was more difficult than it needed to be. They were not sensitive to the differences between trial

modes, and they used strategies that required a larger number of EIPs than was necessary. Thus, while they were able to perform the task successfully, they did not do so in a cognitively economic fashion.

If we compare these results to the findings from the performance analyses presented earlier, an obvious pattern comes to the fore. The three subjects who exhibited signs of adaptation to information processing limitations were generally the most proficient performers. TL was always the best in the P group, AV was usually the best in the P+F group, and IS was often a close second. The three subjects who made the FWS configuration subtask more difficult than it needed to be, on the other hand, were the least proficient performers. WL and ML alternated being the worst in the P group, and AS was clearly the worst in the P+F group. Thus, there seems to be a correlation between performance and adaptation to information processing limitations.

Although we do not have any direct evidence to causally link these two variables, it seems very likely that such a link exists. All subjects have information processing limitations. Furthermore, all subjects were given the same task to perform on the same system. The constraints imposed by these factors cannot be escaped. The only place where there was discretion was in how to deal with the task demands. Some subjects dealt with those demands in a way that showed that they were sensitive to making the task as easy as possible. This involved systematically adapting their behaviour as a function of the context. Other subjects deal with the task demands in a way that made the task more difficult than it needed to be. Their behavior was not as systematically adapted as a function of the context. Thus, it is plausible that the subjects who performed the task in a more economic fashion would outperform those who performed the task in a more challenging fashion.

8. CONCLUSIONS

The purpose of this research was to develop novel measures that could be used to investigate long-term operator adaptation. We explored many different alternatives, many more than are described in this report. Each measure was tested against the data from our six-month longitudinal experiment to see if it shed any light on subjects' adaptation. Although many measures turned out not to be particularly informative, several were successful. Thus, we were able to learn a great deal, not just about the methodological issue of measurement, but also about

the empirical issue of what our subjects were doing, and thus, about the theoretical implications of these findings. In this concluding section, we will describe the contributions, implications, and limitations of this research, as well as some promising ideas for future research.

Contributions

This research makes a number of significant contributions to several different areas. From a methodological perspective, the most important contribution is the synthesis of the AH and the dynamical systems approach to create a principled approach to adaptation measurement. As far as we know, this the first time that the AH has been used in this way. Our analyses indicate that this approach is very powerful since it revealed insights of great theoretical significance that simply could not be directly observed with any of the many other measures we investigated. Because the AH provides a systematic approach to the measurement adaptation, it has the promise to be a useful tool in a wide variety of domains.

Still in a methodological vein, the productive synthesis of the dynamical systems approach and the AH also shows that dynamics is a generalizable tool. Most applications of dynamics in psychology have been in the areas of perception and action (see the section on Coordinative Structures and Intentional Dynamics in the Bibliography in volume 1). While there are some discussions of how it can be applied to cognitive phenomena (Port & van Gelder, 1995), empirical evidence showing the value of the dynamical systems approach to problem solving is hard to come by. Thus, the analyses presented here provide further support for the utility and generalizability of the dynamical systems approach to cognition.

From a theoretical perspective, this research made an important contribution by providing empirical evidence of control based on coordinative structures. Although there is a great deal of such evidence in the motor control literature, as far as we know this is the first time that coordinative structures have been identified in the cognitive engineering literature. This result is surprising if we consider the many differences between motor control and process control (e.g., the magnitude of the lags involved, the psychological resources being used, the characteristics of the object of control). One would think, for instance, that walking would have very little to do with controlling DURESS II. Although there certainly are important differences, our data show that both situations pose a <u>coordination problem</u> and that there may be a generic solution to this problem in the form of a coordinative structure. This insight holds great promise for developing a generic theory of coordination that spans both biological and human-machine coordination.

Such a theory would support the old, but relatively unexplored, view that thinking can be productively viewed as a skill (Bartlett, 1958).

Still in a theoretical vein, we have extended the work of Payne et al. (1993) on adaptive strategy selection to a more complex setting. Most, if not all, of the experiments conducted by Payne et al. have been with static, one-shot decision tasks. Our research shows that their work can be generalized to predict adaptive steady state behaviour in a dynamic, closed loop system. Furthermore, we have extended Payne et al.'s concepts to the study of expertise. Our findings show that subjects who adapt their behaviour to their information processing limitations are more proficient controllers than subjects who are less sensitive to the discretionary demands imposed by a task.

Finally, the findings presented throughout this document show that expertise can, and perhaps should, be viewed as adaptation to the goal-relevant constraints in the environment. Although this is far from a new idea (e.g., Bartlett, 1958; Bernstein, 1967; Gibson, 1991), it has been relatively unappreciated in the expertise literature in cognitive science. The focus has traditionally been more on understanding human information processing limitations than on the constraints on goal-oriented behaviour imposed by the environment. However, the view of expertise as adaptation is gaining more support in recent years (e.g., Ericsson & Lehmann, 1996). However, this newer literature frequently speaks of adaptation without providing a rigorous and comprehensive description of the environment. But since the environment is the referent for adaptation, the claim that human behaviour is adaptive is very difficult to substantiate without such a comprehensive analysis. The work we have presented here shows that the environment needs to be studied in at least as much detail as the organism if we are to study adaptation.

Implications

This investigation of long-term operator adaptation also has implications for the design and evaluation of human-machine systems in NPPs. Because of the limitations of this research (see below), these implications should be considered as tentative hypotheses to be tested more rigorously rather than definitive conclusions.

It is possible for operators to achieve a very high level of performance under normal operating conditions, while using qualitatively different strategies. This is not a new finding (e.g., Edwards & Lees, 1974), but it is worthwhile repeating. Just because there are no obvious deficiencies under routine conditions, it does not follow that there is no cause for concern.

It is also possible to exhibit a very high level of adaptation to system structure and to information processing limitations under normal operating conditions, while again using qualitatively different strategies. Thus, extreme adaptation to goal-relevant constraints can take different forms, some of which are probably more robust or flexible than others.

Specific types of measures to show that operators who demonstrate the same high level of performance and adaptation under normal conditions may be using different strategies to control the system. The synthesis between the AH and the dynamical systems approach seems to provide a systematic, and potentially generalizable, means of addressing this issue.

The dynamical systems approach to performance and adaptation measurement is very promising and complements the insights obtained using other measures (e.g., task completion time, number of errors, verbal protocols, knowledge elicitation tests).

It may be easier and more efficient for operators using a traditional interface based on physical information to improve their performance and adaptation under normal conditions by following a rote set of procedures than trying to understand the basis for their actions. Because of the impoverished information in the interface, it may require too much cognitive effort to understand how, and why, the system reacts in the way it does.

It seems that the highest level of performance and adaptation with an EID interface is achieved by operators who adopt an execution-driven control strategy based on a coordinative structure. In contrast, the highest level of performance and adaptation with a traditional interface is achieved by operators who adopt a program-driven control strategy based on a detailed set of procedures. Thus, these two types of interfaces seem to induce qualitatively different types of expert control strategies.

Limitations

Although this research has generated a number of important contributions and implications, there is no doubt that it is limited in several significant ways. First, there is the issue of low sample size. Because of the duration of the experiment, it was not possible to include more than 6 subjects. Thus, it is difficult to make any strong claims about the generalizability of the results we have observed. Second, we have also not reported any statistical tests in this document. All of our inferences are based on visual inspection of graphs and informal comparisons of quantitative data. We do not know whether the patterns we have detected are statistically reliable. Third, we have deliberately limited the scope of the investigation to normal startup

trials. We do not know to what extent the measures we have developed, or the findings we have uncovered, generalize to other tasks, such as tuning and fault management.

Future Research

Given these limitations, there are a number of promising issues that should be addressed in future research. First, a literature review of methods for statistical analysis of data needs to be conducted. Traditional approaches such as ANOVA have been subject to strong criticisms over the years, and several alternative approaches to data analysis have been proposed. We need to determine which type of data analysis is best suited to goals of our research and the characteristics of our data. Second, we should evaluate whether the measures of adaptation and performance we have developed here can be applied to study tuning and fault trials. This is an issue of high priority because research in motor control has shown that evidence to distinguish control via motor programs from control via coordinative structures is best obtained by perturbing the system and studying subjects' reactions. Tuning tasks involve a perturbation of the goal state, whereas fault trials involve a perturbation of system structure. Thus, these tasks should be particularly sensitive discriminators of program-driven and execution-driven control strategies. Finally, we should also explore further the theoretical similarities between motor control and control of complex, dynamic processes. We could continue to use previous experimental and theoretical work in motor control as a model for future research on long-term operator adaptation. Conversely, our research may offer novel contributions to the study of motor control (e.g., the application of the AH as a systematic technique for the measurement of adaptation). Furthermore, the empirical results we obtain with DURESS II may show that theories in motor control generalize much more broadly than human movement scientists ever suspected. A general systems theory of coordination may be attainable.

REFERENCES

Bartlett, F. (1958). Thinking: An experimental and social study. London: Allen & Unwin.

Bisantz, A. M., & Vicente, K. J. (1994). Making the abstraction hierarchy concrete. International Journal of Human-Computer Studies, 40, 83-117.

Christoffersen, K., Hunter, C. N., & Vicente, K. J. (1994). Research on factors influencing human cognitive behaviour (I) (CEL 94-05). Toronto: University of Toronto, Cognitive Engineering Laboratory.

Edwards, E., Lees, F. P. (1974). The human operator in process control. London: Taylor & Francis.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. Annual Review of Psychology, 47, 273-305.

Gibson, E. J. (1991). An odyssey in learning and perception. Cambridge, MA: MIT Press.

Howie, D. E. (1996). Shaping expertise through ecological interface design: Strategies, metacognition, and individual differences (CEL 96-01). Toronto: University of Toronto, Cognitive Engineering Laboratory.

Howie, D. E., & Vicente, K. J. (in press). Measures of operator performance in complex, dynamic microworlds: Advancing the state of the art. Ergonomics.

Itoh, J., Sakuma, A., & Monta, K. (1995). An ecological interface for supervisory control of BWR nuclear power plants. Control Engineering Practice, 3, 231-239.

Keele, S. W., & Summers, J. (1976). The structure of motor programs. In G. E. Stelmach (Ed.), Motor control: Issues and trends (pp. 109-142). New York: Academic Press.

Moray, N. (1980). The use of information transmission as nonparametric correlation in the analysis of complex behavior: A preliminary report (ONR Final Control Report). Cambridge, MA: Man-Machine Systems Laboratory, MIT.

Newell, K. M., & Hancock, P. A. (1984). Forgotten moments: A note on skewness and kurtosis as influential factors in inferences extrapolated from response distributions. Journal of Motor Behaviour, 16, 320-335.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception, 5, 437-459.

Pask, G., & Scott, B. C. (1972). Learning styles and individual competence. International Journal of Man-Machine Studies, 4, 217-253.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The adaptive decision maker. Cambridge: Cambridge University Press.

Port, R. F., & van Gelder, T. (1995). Mind as motion: Explorations in the dynamics of cognition. Cambridge, MA: MIT Press.

Rasmussen, J. (1986). Information processing and human-machine interaction: An approach to cognitive engineering. Amsterdam: North-Holland.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). Cognitive systems engineering. New York: Wiley.

Sanderson, P. M., Verhage, A. G., & Fuld, R. B. (1989). State-space and verbal protocol methods for studying the human operator in process control. Ergonomics, 32, 1343-1372.

Shaw, R. E., Kadar, E., Sim, M. & Repperger, D. W. (1992). The intentional spring: A strategy for modeling systems that learn to perform intentional acts. Journal of Motor Behavior, 24, 3-28.

Tuller, B., Turvey, M. T., & Fitch, H. L. (1982). The Bernstein perspective: II. The concept of muscle linkage or coordinative structure. In J. A. S. Kelso (Ed.), Human motor behavior: An introduction (pp. 253-270). Hillsdale, NJ: Erlbaum.

Turvey, M. T. (1988). Simplicity from complexity: Archetypal action regimes and smart perceptual instruments as execution-driven phenomena. In J. A. S. Kelso, A. J. Mandel, & M. F. Shlesinger (Eds.), Dynamic patterns in complex systems (pp. 327-347). Singapore: World Scientific.

Turvey, M. T. (1990). Coordination. American Psychologist, 45, 938-953.

Turvey, M. T., Fitch, H. L., & Tuller, B. (1982). The Bernstein perspective: I. The problems of degrees of freedom and context-conditioned variability. In J. A. S. Kelso (Ed.), Human motor behavior: An introduction (pp. 239-252). Hillsdale, NJ: Erlbaum.

Valdes-Perez, R. E., & Conant, R. C. (1983). Information loss due to data quantization in reconstructability analysis. International Journal of General Systems, 9, 235-247.

Vicente, K. J. (1990). Ecological interface design as an analytical evaluation tool. In Proceedings of the American Nuclear Society Topical Meeting on Advances in Human Factors Research: Nuclear and Beyond (pp. 259-265). LaGrange, IL: ANS.

Vicente, K. J., & Pawlak, W. S. (1994). Cognitive work analysis for the DURESS II system (CEL 94-03). Toronto: University of Toronto, Cognitive Engineering Laboratory.

Vicente, K. J., & Rasmussen, J. (1990). The ecology of human-machine systems II: Mediating "direct perception" in complex work domains. Ecological Psychology, 2, 207-249.

Yu, X., Khan, F. S., Lau, E., Vicente, K. J., & Carter, M. W. (1997). ADAPT user's manual: A data analysis tool for human performance evaluation in dynamic systems (CEL 97-03). Toronto: University of Toronto, Cognitive Engineering Laboratory.

Cognitive Engineering Laboratory

CEL TECHNICAL REPORT SERIES

CEL 93-01 "Egg-sucking, Mousetraps, and the Tower of Babel: Making Human Factors Guidance More Accessible to Designers"
• Kim J. Vicente, Catherine M. Burns, & William S. Pawlak

CEL 93-02 "Effects of Expertise on Reasoning Trajectories in an Abstraction Hierarchy: Fault Diagnosis in a Process Control System"
• Klaus Christoffersen, Alex Pereklita, & Kim J. Vicente

CEL 94-01 "Cognitive 'Dipsticks': Knowledge Elicitation Techniques for Cognitive Engineering Research"
• Klaus Christoffersen, Christopher N. Hunter, & Kim J. Vicente

CEL 94-02 "Muddling Through Wicked Problems: Exploring the Role of Human Factors Information in Design"
• Catherine M. Burns

CEL 94-03 "Cognitive Work Analysis for the DURESS II System"
• Kim J. Vicente & William S. Pawlak

CEL 94-04 "Inducing Effective Control Strategies Through Ecological Interface Design"
• William S. Pawlak

CEL 94-05 "Research on Factors Influencing Human Cognitive Behaviour (I)"
• Klaus Christoffersen, Christopher N. Hunter, & Kim J. Vicente

CEL 94-06 "Ecological Interfaces for Complex Industrial Plants"
• Nick Dinadis & Kim J. Vicente

CEL 94-07 "Evaluation of a Display Design Space: Transparent Layered User Interfaces"
• Beverly L. Harrison, Hiroshi Ishii, Kim J. Vicente, & Bill Buxton

CEL 94-08 "Designing and Evaluating Semi-Transparent 'Silk' User Interface Objects: Supporting Focused and Divided Attention"
• Beverly L. Harrison, Shumin Zhai, Kim J. Vicente, & Bill Buxton

CEL 95-01 "An Ecological Theory of Expertise Effects in Memory Recall"
• Kim J. Vicente & JoAnne H. Wang

CEL-SP "Strategic Plan"
• Cognitive Engineering Laboratory

CEL-LP "Cognitive Engineering Laboratory Profile"
• Cognitive Engineering Laboratory

CEL 95-04 "A Field Study of Operator Cognitive Monitoring at Pickering Nuclear Generating Station-B"
• Kim J. Vicente & Catherine M. Burns

CEL 95-05 "An Empirical Investigation of the Effects of Training and Interface Design on the Control of Complex Systems"
• Christopher N. Hunter

CEL 95-06 "Applying Human Factors to the Design of Medical Equipment: Patient-Controlled Analgesia"
• Laura Lin, Racquel Isla, Karine Doniz, Heather Harkness, Kim J. Vicente, & D. John Doyle

CEL 95-07 "An Experimental Evaluation of Transparent Menu Usage"
• Beverly L. Harrison & Kim J. Vicente

CEL 95-08 "Research on Factors Influencing Human Cognitive Behaviour (II)"
• Christopher N. Hunter, Michael E. Janzen, & Kim J. Vicente

CEL 95-09 "To the Beat of a Different Drummer: The Role of Individual Differences in Ecological Interface Design"
• Dianne Howie

CEL 95-10 "Emergent Features and Temporal Information: Shall the Twain Ever Meet?"
• JoAnne H. Wang

CEL 95-11 "Physical and Functional Displays in Process Supervision and Control"
• Catherine M. Burns & Kim J. Vicente

CEL 96-01 "Shaping Expertise Through Ecological Interface Design: Strategies, Metacognition, and Individual Differences"
• Dianne E. Howie

CEL 96-02 "Skill, Participation, and Competence: Implications of Ecological Interface Design for Working Life"
• Peter Benda, Giuseppe Cioffi, & Kim J. Vicente

CEL 96-03 "Practical Problem Solving in a Design Microworld: An Exploratory Study'
• Klaus Christoffersen

CEL 96-04 "Review of Alarm Systems for Nuclear Power Plants"
• Kim J. Vicente

CEL 96-05 "DURESS II User's Manual: A Thermal-hydraulic Process Simulator for Research and Teaching"
• Lisa C. Orchanian, Thomas P. Smahel, Dianne E. Howie, & Kim J. Vicente

CEL 96-06 "Research on Factors Influencing Human Cognitive Behaviour (III)"
• Dianne E. Howie, Michael E. Janzen, & Kim J. Vicente

CEL 96-07 "Application of Ecological Interface Design to Aviation"
• Nick Dinadis & Kim J. Vicente

CEL 96-08 "Distributed Cognition Demands a Second Metaphor for Cognitive Science"
• Kim J. Vicente

CEL 96-09 "An Experimental Evaluation of Functional Displays in Process Supervision and Control"
• Catherine M. Burns and Kim J. Vicente

CEL 96-10 "The Design and Evaluation of Transparent User Interfaces: From Theory to Practice"
• Beverly L. Harrison

CEL 97-01 "Cognitive Functioning of Control Room Operators: Final Phase"
• Kim J. Vicente, Randall J. Mumaw, & Emilie M. Roth

CEL 97-02 "Applying Human Factors Engineering to Medical Device Design: An Empirical Evaluation of Two Patient-Controlled Analgesia Machine Interfaces"
• Laura Lin

CEL 97-03 "ADAPT User's Manual: A Data Analysis Tool for Human Performance Evaluation in Dynamic Systems"
• Xinyao Yu, Farzad S. Khan, Elfreda Lau, Kim J. Vicente, & Michael W. Carter

CEL 97-04 "Research on the Characteristics of Long-Term Adaptation"
• Xinyao Yu, Renée Chow, Greg A. Jamieson, Rasha Khayat, Elfreda Lau, Gerard L. Torenvliet, Kim J. Vicente, & Michael W. Carter