An ACT-R Model of Individual Differences in Changes in Adaptivity due to Mental Fatigue

Linda Jongman (linda@tcw3.ppsw.rug.nl)
Experimental and Work Psychology, University of Groningen
Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands

Niels Taatgen (niels@tcw3.ppsw.rug.nl)
Cognitive Science and Engineering, University of Groningen
Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands

Abstract

In this paper we show that adaptivity is reduced when people become fatigued. Fatigued people adapt worse to changing probability distributions as compared to non-fatigued individuals. In an ACT-R model of the task we show that this decreased adaptivity is due to a decrease in the use of one specific strategy. We argue that the use of this strategy is decreased, because it places high demands on working memory. In previous research we also found indications that mental fatigue is related to changes in working memory functioning. We argue that modeling individual differences in performance will provide better insight in the processes involved in mental fatigue.

Introduction

In this paper, mental fatigue is defined as the subjective feeling of being fatigued, combined which negative changes in performance, apart from the influences of time of day, or investment of physical effort. Many research projects concerning mental fatigue have failed to show decreases in performance as a result from fatigue. It appears that people are able to maintain adequate performance for a substantial amount of time. A growing number of investigations reveal indications that it is the way in which this performance is attained that changes when people become fatigued, as was already suggested by Bartlett (1943) and Broadbent (1979). In the 1970's, Shingledecker and Holding (1974) showed that after 24-32 hours of continuous work on a mentally loading task battery, people changed the order in which they tested possibly defective components on a fault-diagnosis task. The task consisted of finding the defective resistor in three banks of resistors containing one, two and three resistors respectively. All resistors had an equal probability of being defective, so the probabilities for the three banks of containing the defective resistor were respectively 17, 33 and 50 percent. The difficulty of the calculations that had to be made for finding the defective transistor were easiest for the bank with one transistor and most difficult for the bank containing three transistors. It appeared that participants, in the beginning of the experiment, chose to start testing the bank with three transistors which was most likely to contain the defective component. At the end of the experiment, however, they started more often with the bank with only one transistor which was the easiest one to test.

In a more recent article, Schunn and Reder (1998) show for a number of different tasks that people adapt their strategies to changed success rates of these strategies. They also showed that this, what they call extrinsic adaptivity, is a source of differences across individuals and that working-memory capacity and reasoning ability are good predictors of this adaptivity ability.

We developed a task, which is a combination of these two approaches, to investigate whether adaptivity is influenced by changes in mental circumstances, in this case by mental fatigue.

The Coffee Task

In stead of diagnosing transistors as was done in the Shingledecker and Holding experiment, participants have to weigh packets of coffee. On each trial, participants are shown three balances containing a tray with one, two and three packets of coffee respectively. The weights of the six packets and the three trays differs for each trial. The task is to find the one packet that has the same weight as the tray it is on. Participants cannot weigh individual packets, but are only allowed to weigh the whole tray. To find the weight of a specific packet, the balance has to be weighed, the packet must be taken of the balance and the balance has to be weighed again and the difference in weight has to be calculated. Packets cannot be put back on the balance. The task was designed in this way to ensure that calculations for the balance with three packets is hardest, like in the Shingledecker and Holding experiment. Figure 1 shows the interface of the task.

To investigate adaptivity, the probability of success for the three balances is manipulated. At the beginning of the experiment, the probability that the goal packet is on a certain balance is 10% (for the balance with one packet), 20% (for the balance with two packets) and 70% (for the balance with three packets.) So, the probability per packet is highest at the balance with three packets. However, after every five trials, the probabilities are changed according to which balance the participant chooses to weigh first. The balance that is started with most often is reduced in probability. Participants are told that the probability changes in this direction, but not precisely when the probabilities are changed and how big this change in probability is. They are pointed out that it is wise to start with the balance with the highest probability per packet and they are instructed to complete as many trials as possible, making as few mistakes as possible.

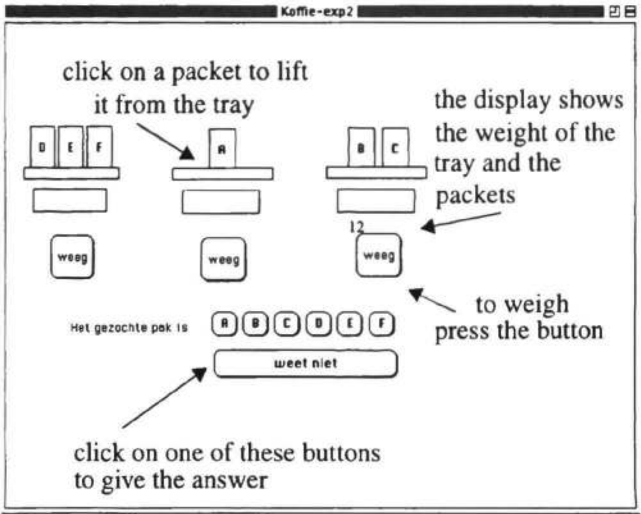

Figure 1: The interface of the coffee task

The Experiment

32 undergraduate students participated in the experiment distributed over two conditions. In both conditions participants performed the coffee task for 25 minutes at the beginning of the experiment (the PRE-test) and at the end of the experiment (the POST-test). Before both tests, participants had to rate how fatigued they felt on a 150-point word-anchored scale. In the time between the PRE-and POST-test, participants of the experimental condition had to continuously solve complex scheduling problems under time-pressure for two hours (for a description of the task see Taatgen, 1997). Participants in the control condition could watch video tapes or read books for two hours. All participants were trained on the task for 3 times 25 minutes on the day preceding the experiment.

Results

Analysis of the reported feelings of fatigue revealed a main effect of session ($F(1,30) = 23.937$, $p<.001$) and an interaction of session and condition ($F(1,30) = 4.343$, $p<.05$), indicating that the fatigue-manipulation had the intended effect. As a whole group the participants are more fatigued on the POST-test, and this effect is stronger for the participants from the experimental condition. Figure 2 shows the fatigue ratings.

As for the strategy measures, contrary to the findings of Shingledecker and Holding (1974), no difference in global preference for one of the balances could be found on the POST-test as compared to the PRE-test.

More interesting is how well participants adapt their choices according to the changing probability distribution. The neutral probabilities for the three balances are 17, 33 and 50 percent respectively (as used in the Shingledecker and Holding experiment). If a participants always chooses the balance with the highest probability per packet, the distribution will remain close to 17,33 and 50 percent. Large deviations from this neutral probability distribution indicate that the participant often chooses a balance that was not optimal. This deviation for participant i at trial (j) can be calculated according to formula (1).

$$\text{Deviation}_i(j) = 50 - P3 \tag{1}$$

In this formula, *P3* represents the probability (as a percentage) that the goal-packet is on the balance containing three packets. The deviation is zero when $P3 = 50$, as is the case in the neutral distribution. The deviation is plotted positive if the participant chose the optimal balance and negative if the participant chose a non-optimal balance. A deviation close to zero means that the participant adapts to the changing probabilities, whereas a deviation far from zero means he is not. Figure 3 shows an example of a deviation plot, where the participant starts out with a large deviation, but attains a performance close to zero deviation in the second half of the test.

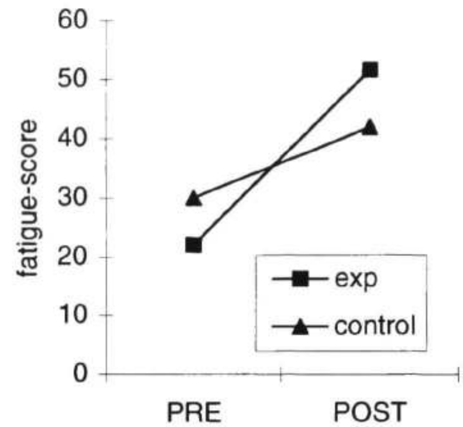

Figure 2: Subjective measurements of fatigue for the PRE- and POST-test.

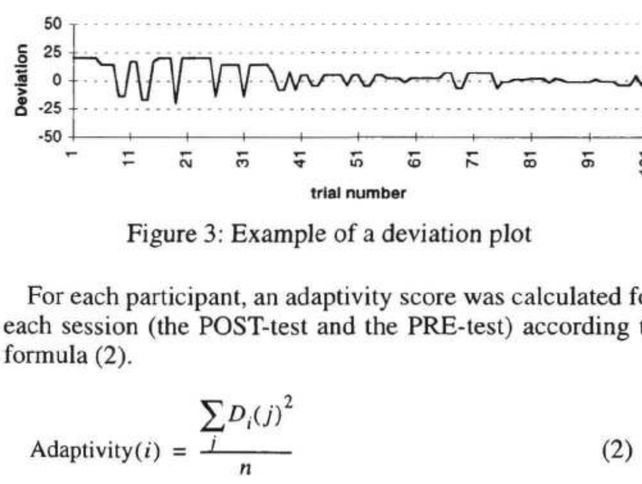

Figure 3: Example of a deviation plot

For each participant, an adaptivity score was calculated for each session (the POST-test and the PRE-test) according to formula (2).

$$\text{Adaptivity}(i) = \frac{\sum_j D_i(j)^2}{n} \tag{2}$$

$D_i(j)$ is the deviation score for participant i on trial j. This adaptivity is the mean squared deviation score for a whole session. N is the total number of trials the participant completed. We chose to take the squared deviation in order to get rid of the sign and to stress large deviations. Figure 4 shows how this adaptivity changes from the PRE-test to the POST-test.

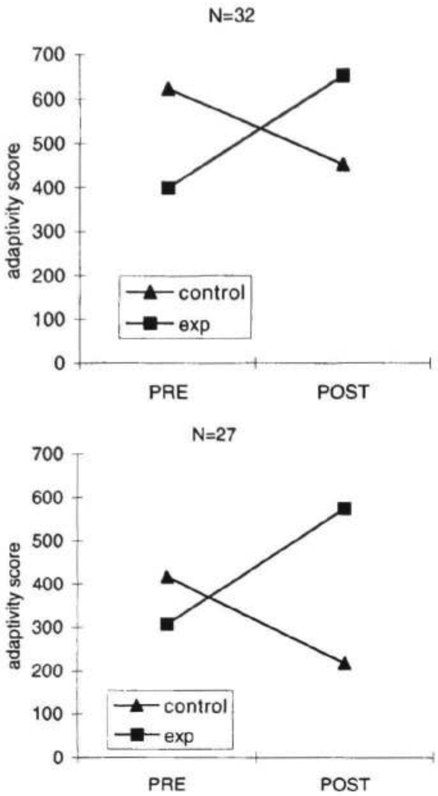

Figure 4: Adaptivity scores for the PRE- and POST-test for all participants (upper figure), and only participants who adapt to the changing probabilities (lower figure)

While no main effect of session (PRE-test, POST-test) could be found, there was a significant interaction of session by condition ($F(1,30) = 9.912$, $p<.01$). As can be seen in the upper half of the figure, participants from the control condition adapt better on the POST-test as compared on the PRE-test, while participants from the experimental condition do adapt worse on the POST-test. The difference between the two conditions on the PRE-test was non-significant and could be attributed to five participants who did not adapt at all to the changing probabilities. The lower part of figure 4 shows the adaptivity for the two conditions when these five participants are removed from the set.

These results indicate that strategy adaptivity is reduced when people become fatigued. Moreover, the decrease in performance correlates with the change in reported feelings of fatigue ($r=.50$, $p<.01$).

Strategies

An interesting question is how to explain the difference in adaptivity between the two conditions. One approach is to look at which strategies are possible to do the task. We hypothesize that there are two possible strategies: choose the same balance as on the previous trial (P), and choose the balance which contained the answer on the previous trial (A). The latter of the two is an adaptive strategy. Because these strategies overlap in their predicted responses, responses can be categorized into the following four categories:

(1) $P \wedge A$ (PA)
(2) $P \wedge \neg A$ (PnA)
(3) $\neg P \wedge A$ (nPA)
(4) $\neg P \wedge \neg A$ (nPnA)

Based on the distribution of the responses into these four categories, we estimated the use of the two strategies (A) and (P). Responses in category (3) strongly indicate the use of the adaptive A-strategy. Participants choose the balance that contained the answer on the previous trial. Responses in category (2) indicate the use of the non-adaptive P-strategy. Category (4) is some kind of rest category in which a different balance is chosen, but not the one that contained the answer on the previous trial. Responses in this category can indicate the use of a different strategy as the two mentioned before.

As was described in the introduction, we wanted to see whether adaptivity is influenced by mental fatigue. The results of the experiment indicated that the experimental group had a decreased adaptivity score. An interesting question is whether this reduction in their adaptivity scores could be explained by a reduction in the use of the adaptive A-strategy. If so, this should be visible by a decrease in responses in the nPA-category. We must note that not all participants were fatigued to the same degree by the experimental manipulation. Therefore, we have split the participants in a high-fatigue group and a low-fatigue group, based on the median increase in fatigue scores for the experimental group. Although we did not find a main effect of session in the number of responses in the nPA category, there was a significant interaction of session and the two fatigue groups ($F(1,30) = 5.548$, $p=.025$). So, only the high-fatigue group showed a decrease in responses in the nPA category.

Furthermore, the four different categories correlate strongly with the adaptivity scores of the participants as calculated according to formula (2):

	PRE adaptivity	POST adaptivity
PA	-.53**	-.72***
PnA	.90***	.86***
nPA	-.67***	-.61***
nPnA	-.42*	-.44*

* p<.05, ** p<.01, *** p<.001

As this table shows, the A-strategy, indicated by nPA responses, has a strong negative correlation with the adaptivity score, which implies using the A-strategy has a positive effect on performance. The P-strategy on the other hand has a very negative effect on performance, as can be concluded from the positive correlation between PnA and performance.

The ACT-R Model

In order to explore the question whether the proposed strategies fully characterize the behavior of participants on this task, we developed an ACT-R model to simulate the behavior of individual participants.

ACT-R (Anderson & Lebiere, 1998) is a hybrid cognitive architecture based on a production system. It has been used to explain a wide range of cognitive phenomena by producing models that make precise predictions about choices, latencies and errors. The main mechanism we will use is ACT-R's conflict resolution that will be used to choose between strategies.

The basis for this choice between strategies consists of the following three rules:

(1) A rule that proposes to start with the same balance as the previous trial, corresponding to the P-strategy
(2) A rule that proposes to use the answer to the previous trial as the basis for the choice, corresponding to the A-strategy
(3) A rule that picks a random balance to start with, which differs from the balance chosen first in the previous trial. We will call this the rest (R) strategy.

This last rule is used to represent the nPnA cases, for which it is not clear what strategy the participant pursues.

In order to choose between rules, ACT-R (Anderson & Lebiere, 1998) uses a conflict-resolution mechanism based on the expected gain of a rule. The expected gain of a rule is calculated by taking the following factors into account: an estimate of the probability that the rule achieves the current goal, an estimate of the costs that are involved in achieving this goal, and that value of the goal itself. Basically, the rule with the highest expected gain is selected. However, since noise is added to the expected gain, the best rule not always fires, it only has the highest probability of firing, governed by the following equation:

$$\text{Probability of choosing rule } i = \frac{e^{E_i/t}}{\sum_j e^{E_j/t}} \qquad (3)$$

In this equation, E_i represents the expected gain of rule i, and the t parameter determines the level of noise.

From the experiment we have, for each participant, the proportion of times they chose a response in the categories PA, PnA, nPA and nPnA for both the PRE- and the POST-test. These values can be used to estimate the probability the participant uses the P-strategy, the A-strategy, or the R-strategy. Consequently, these estimates can be used to calculate suitable expected gains for the three rules that choose the strategies.

To see how well the model can estimate the adaptivity score for each participant in each test, the model was run for 50 times with the three expected-gain parameters estimated for each participant and each test. The result is shown in figure 5. Each point in the graph corresponds to one test (PRE or POST) of a single participant). The correlation between the data and the model predictions is 0.77, which is not particularly high, although encouraging. The problem is, that there is a lot of randomness involved in the model. Even if the average score for one of the models is 200, values for individual runs may range from 100 to 500. So we decided to see how far apart the experimental score and each model prediction was in terms of the standard deviation of the model (based on the 50 runs for each score). The result was, that 66% of the experimental scores was within one S.D. of the model prediction, and 97% within two S.D.'s, exactly what one would expect in a normal distribution.

To get a better idea of how the model's performance can be compared to what goes within a test, we did a second run of 20 simulations for each participant and each test. In stead of averaging these simulations, we picked the simulation which adaptivity score was closest to the adaptivity score in the experiment. This "best of 20" strategy, nor surprisingly, boosts the correlation between the data and the model to 0.99. Figure 6 shows the match between the model and the data.

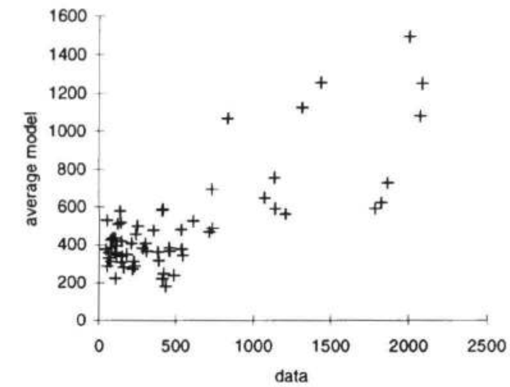

Figure 5: Predictions of the model compared to the data

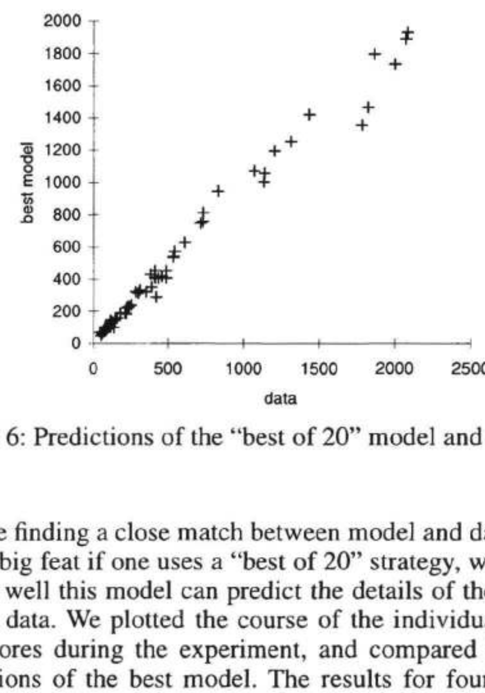

Figure 6: Predictions of the "best of 20" model and the data

Since finding a close match between model and data is not such a big feat if one uses a "best of 20" strategy, we looked at how well this model can predict the details of the experimental data. We plotted the course of the individual deviation scores during the experiment, and compared it to the predictions of the best model. The results for four participants are depicted in figure 7. The left-hand column shows data from the experiment: a PRE- and POST-test for each participant. The right-hand column shows the predictions of the model for each of the individual runs. The four participants shown are all from the high fatigue-group and all performed worse on the POST-test than on the PRE-test, as measured by the adaptivity score. A total of six participants satisfied both of these criteria, so two-third of the "interesting" group is shown in figure 7.

As one can see in figure 7, the plots of the deviation scores show huge individual differences. The model, however, captures these differences quite nicely, especially given the fact

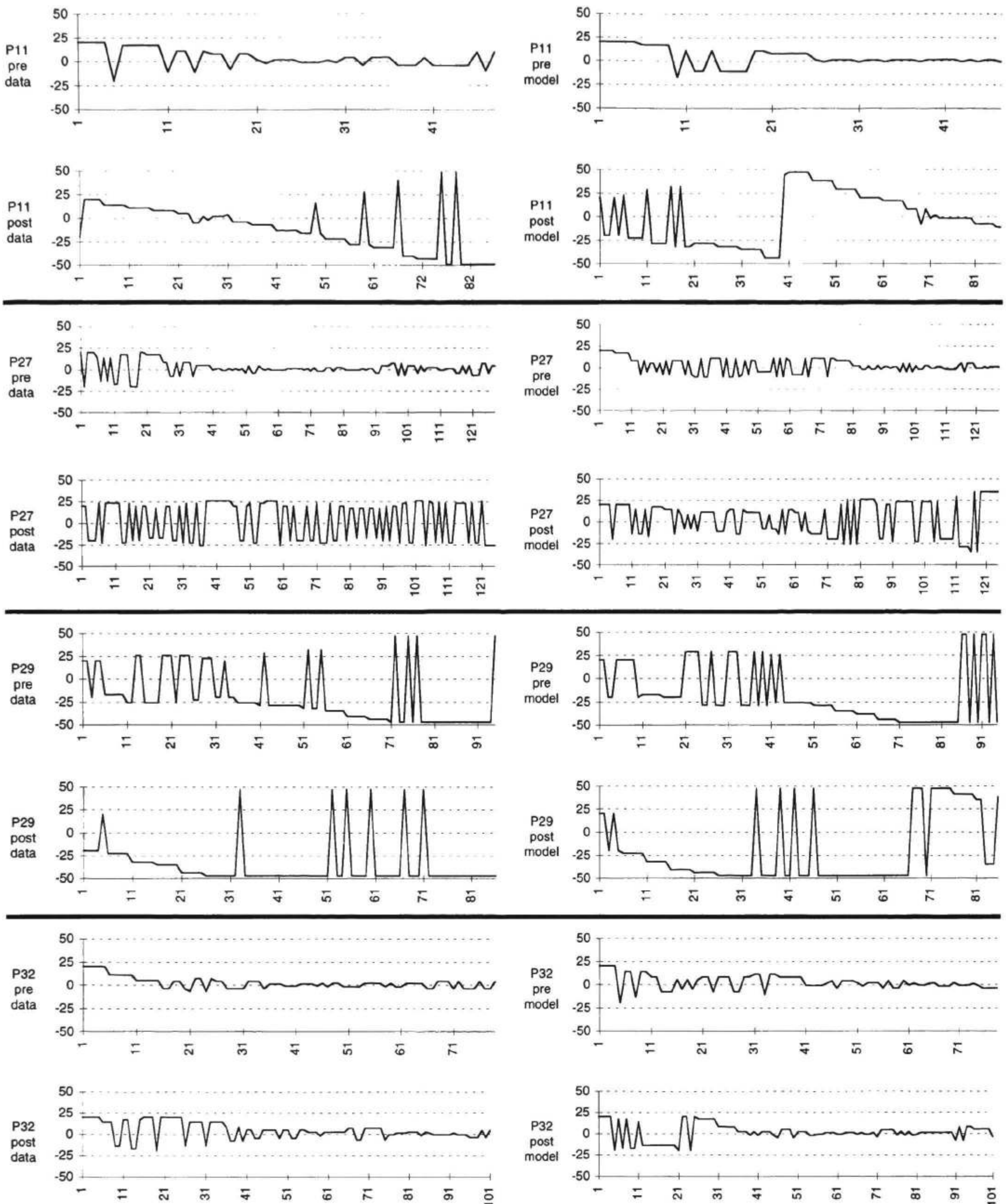

Figure 7: Deviation scores during the course of the experiment for the experiment and the model's predictions. The four participants shown are all from the experimental condition, all reported fatigue at the post-test, and all had a worse performance on the post-test compared to the pre-test.

that no information about the course of the deviations has been put into the model. Basically, each model is based on four parameters from the experimental data: the PC, nPC, PnC, and the adaptivity score.

Each of the four fatigued participants in the figure shows a slightly different pattern of fatigue. Participant 11 does quite well on the PRE-test, and keeps her deviation score quite close to zero. At the POST-test, however, she shows hardly any adaptivity at all anymore. Participant 27 also starts out very well on the PRE-test, but cannot maintain this performance in the POST-test, where he starts oscillating. Participant 29 already starts with poor deviation scores, but gets even worse in the POST-test, where she goes down to a deviation score of -50 with an occasional spike to +50. Participant 32, finally, exhibits a good performance on both the PRE- and the POST-test, but she takes slightly more time to arrive at a deviation of around zero in the post-test.

In all four cases the model shows the same pattern of the effect of fatigue as the data. This is an indication that although the deviation plots of the participants are all quite different, the essence is captured in the four parameters that are put into the model.

Discussion

In the introduction, we hypothesized that mental fatigue would influence adaptivity. The results from the experiment show that adaptivity is reduced when people become fatigued. Moreover, the increase in reported feelings of fatigue strongly correlates with a reduction of adaptivity. If we zoom in on adaptivity in more detail, we see that the reduction in adaptivity for fatigued people could be largely explained by a reduction in the use of the adaptive strategy. In the model, this adaptive strategy was defined as choosing the balance that contained the answer on the previous trial, which is a fairly simple implementation. However, it is possible, people do not realize that such a simple strategy will do the job. It is likely that people will base their decisions on which balance contained the answer on the last two or three trials. In that way, this adaptive strategy will place high demands on working memory. This is consistent with the findings of Schunn and Reder (1998) who report that working-memory capacity is a good predictor of adaptivity. A possible reason that this adaptive strategy is used less when people become fatigued, is that working memory functioning is impaired by mental fatigue. Jongman (1998) also found some indications that working memory functioning could play a role in mental fatigue.

As for the model, we hypothesized that there are two possible strategies to perform the task. Figure 7 shows that the model is able to capture many aspects of individual participants's performance on the task. So, the two strategies gave an adequate representation of people's performance. This was also confirmed by the strong correlations between the different response categories and the adaptivity scores. However, we also found a moderately significant negative correlation between the nPnA category and the adaptivity scores. This may indicate that participant uses a more elaborate adaptive strategy, related tot the A-strategy, but using more trials to base the decision on. Or it may indicate a totally different strategy, meaning people use at least a third strategy as well, which was not captured by our model, but which did have a positive influence on their adaptivity scores.

Overall, the model gave an encouraging fit of the data. Six people from the high-fatigue group showed huge changes in adaptivity from the PRE- to the POST-test. Although the pattern of change was different for the six persons, as shown for four persons in figure 7, the model fitted these different patterns quite nicely. So, the differences in these patterns can be adequately explained by changes in the frequencies these strategies are used.

Many research projects concerning mental fatigue show very specific changes in performance for different individuals. In this paper we showed that different patterns could be explained by changes in the use of a single strategy. We will argue therefore, that fatigue research and related fields will benefit from an approach that focuses on modeling individual differences, thus avoiding the risk of throwing the baby out with the bath water.

Acknowledgments

This research is part of the Netherlands Concerted Research Action "Fatigue at Work" funded by the Netherlands Organization for Scientific Research (NWO).

References

Anderson, J.R. & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Bartlett, F.R.S. (1943). Fatigue following highly skilled work. *Proceed. Royal Society,* 131, 247-257.

Broadbent, D.E. (1979). Is a fatigue test possible now? The Society Lecture 1979. *Ergonomics, 12,* 1277-1290.

Jongman, L. (1998) How to fatigue ACT-R? Proceedings of the Second European Conference on Cognitive Modelling. Nottingham: Nottingham University Press, 52-57.

Shingledecker, C.A. & Holding, D.H. (1974). Risk and effort measures of fatigue. *Journal of motor behavior, 6,* 17-25.

Schunn, C.D & Reder, L.M. (1998). Strategy adaptivity and individual differences. *The psychology of learning and motivation, 38,* 115-154.

Taatgen, N.A. (1997). A rational analysis of alternating search and reflection strategies in problem solving. *Proceedings of the 19th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.