Malleable Attentional Resources Theory: A new explanation for the effects of mental

underload on performance

MARK S. YOUNG[1] and NEVILLE A. STANTON[2]

[1]Railway Safety, London, England

[2]Department of Design, Brunel University, England

This paper proposes a new theory to account for the effects of underload on performance.

Malleable attentional resources theory posits that attentional capacity can change size in

response to changes in task demands. As such, the performance decrements associated

with mental underload can be explained by a lack of appropriate attentional resources.

These proposals were explored in a driving simulator experiment. Vehicle automation

was manipulated at four levels, and mental workload was assessed with a secondary task.

Eye movements were also recorded to determine whether attentional capacity varied with

mental workload. The results showed a clear decrease in mental workload associated

with some levels of automation. Most striking, though, were the results derived from the

eye movement recordings, which demonstrated that attentional capacity varies directly

with level of mental workload. These data fully supported the predictions of the new

malleable attentional resources theory. The implications of this theory are discussed with

regard to capacity models of attention, as well as to the design of future vehicle systems.

Keywords: attention, automation, driver behaviour, mental workload, multiple resources

AUTOMOTIVE AUTOMATION

With the new millennium upon us, major motor manufacturers are offering new vehicle automation devices. Adaptive Cruise Control (ACC) has already been released, offering total longitudinal control of the vehicle. Soon, we will see lateral control devices such as Active Steering (AS) taking to the roads. Although automation is usually implemented with the intention of reducing workload and thereby improving performance, current evidence suggests that excessively low mental demands are actually detrimental to performance. Rather than being a consequence of out-of-the-loop behaviour (Ephrath and Young, 1981; Kessel and Wickens, 1982), some believe that mental underload is in itself responsible for performance decrements (e.g., Brookhuis, 1993; Hancock and Caird, 1993). Indeed, underload is possibly of greater concern than overload, as it is more difficult to detect (Hancock and Parasuraman, 1992; Hancock and Verwey, 1997).

Although there is widespread concern about mental underload, relatively few researchers seem to be actively involved in exploring the issue in the context of vehicle automation. There are, however, a handful of experiments which have found degraded performance when using vehicle automation, mostly in recovery from automation failure. The explanations for such results vary around expectations about the automation (Nilsson, 1995), mobilisation of effort (Desmond, Hancock and Monette, 1998), complacency (de Waard, van der Hulst, Hoedemaeker and Brookhuis, 1999), and mental workload (Stanton, Young and McCaulder, 1997).

Despite these few studies, the literature has so far failed to produce a coherent and parsimonious explanation for why mental underload should be detrimental to

performance. This paper uses the currently pertinent subject of vehicle automation as an

applied focus for a significant theoretical proposition.

Malleable Attentional Resources Theory (MART)

       The breadth of the underload problem in the literature suggests it is not limited to

situations involving automation. An example is recent research at the University of

Michigan's Transportation Research Institute. Participants were sometimes slower to

complete in-car tasks when stationary, than when trying to perform them while driving

(O. Tsimhoni, personal communication, September 14 1999). It seemed that participants

were not as efficient without the additional driving workload even though they were

trying to complete the task as quickly as they could. Thus there appears to be an

underlying cognitive problem associated with underload, which has not yet been

satisfactorily identified. So, in an attempt to fill this gap in the literature, the current

paper offers a new theory based on attentional resource theory.

       Attentional resource theories make a common basic assumption about

performance: if demands exceed resource capacity, performance degrades. In the

original capacity model of attention, Kahneman (1973) suggested that attentional

capacity was positively associated with physiological arousal. Further work since has

found that resource size may change with long-term fluctuations in mood or age (Hasher

and Zacks, 1979; Humphreys and Revelle, 1984). However, most applied research on

attention has implicitly assumed that the size of resource pools is fixed (cf. Wickens,

1984; 1992). It is posited here, though, that this limit may change in the relatively short

term, depending on task circumstances. This introduces the concept of <u>malleable</u>

<u>attentional resource pools</u>.

Evidence is accumulating that simply reducing demand is not necessarily a key to

improving performance. It has been proposed (e.g., Young and Stanton, 1997a; 1997b;

1999a; 1999b; 2000) that resources may actually shrink to accommodate any demand

reduction, in a converse of the 'work expands to fill the time available' tenet. This could

explain the apparent degradation of attention and performance observed in low demand

tasks. If the maximum capacity of an operator has been limited as a consequence of the

task, it is not surprising that they cannot cope when a critical situation arises. Malleable

attentional resources theory (MART) therefore potentially explains why mental

underload can lead to performance degradation.

Imagine someone driving a car with ACC and AS engaged. This is a situation

which considerably reduces MWL (Young and Stanton, 1997a). Assuming an attentional

demand model of MWL (cf. Liao and Moray, 1993; Young and Stanton, 2001), this

translates to low demand on resources. Now, MART posits that the size of the relevant

resource pool will temporarily diminish, as it is not required. This could result in poorer

performance on any subsidiary tasks, or problems if the driver is suddenly faced with

increased demand (e.g., if the automation fails).

The idea that the level of task demands can influence cognitive processing has

been hinted at in previous research. Buck, Payne and Barany (1994) quoted the 'par

hypothesis' to explain some of their results. This states that, as demands fluctuate,

operators increase or decrease the amount of effort invested in a task to maintain

performance at a set level. This level represents an operator's personal par for that task.

There is some support for this notion. Liao and Moray (1993) found that participants invest more effort with higher time pressure, which may increase attentional capacity. Conversely, Desmond and Hoyes (1996) concluded that a decrease in performance at low levels of demand might be due to a failure to mobilise effort appropriately to match the task. More recently, Wickens, Xu, Helleberg and Marsh (2001) suggested effort conservation as the explanation as to why pilots in a simulated aviation task actually devoted more visual attention to an instrument panel when the task was easier. Finally, the concept of fluctuating resource pools has also been aired by Karashima and Saito (2001), in their proposed Flucutation of Working Memory (FLOWM) hypothesis. Whilst these authors present a very detailed explanation of how capacity fluctuates, they fail to offer any reasoning as to why (see also Young and Stanton, 2002, for a discussion of working memory versus attentional resource models).

The malleable resources hypothesis reflects these attitudes, but is a little more parsimonious with respect to current knowledge. Taking inspiration from Kahneman's (1973) original theory, it does not appeal to extraneous concepts such as effort or motivation, instead relying purely upon task demands to govern the size of attentional resources. In its 'strongest' form, then, MART states that: the size of attentional resource pools varies positively with MWL up to a finite limit, such that excessive reductions in MWL shrink capacity (which in turn can cause performance decrements), and that this is independent of variations in arousal or effort. A simpler form would merely state that attentional capacity and MWL are directly related.

The present experiment was intended as a first step to determining whether the malleable resources hypothesis has enough credence to warrant deeper investigation.

Vehicle automation was investigated in a simulated environment, and care was taken to control for alternative explanations as much as possible. By systematically varying the level of automation, factors such as boredom, fatigue, and motivation should be accounted for. Similarly, since all conditions imposed similar levels of physical demand, it is unlikely that physiological arousal will have an affect on the results.

METHOD

<u>Design</u>

This experiment was designed to investigate the potential differences in attention and workload amongst drivers under different conditions of automation. It is expected that MWL will decrease as more levels of automation are introduced, and that such reductions may be associated with decrements in performance. Previous research (e.g., Nilsson, 1995; Stanton et al., 1997; Ward, Fairclough & Humphreys, 1995) has largely concentrated on the effects of ACC. The Southampton Driving Simulator (SDS) was used in the present study to extend this paradigm by also considering lateral control (i.e., AS).

A within-subjects design was used. Level of automation constituted the independent variable, with four levels: manual (the participant controls speed, headway, and steering), ACC (longitudinal control is automated), AS (lateral control is automated), and ACC+AS (both longitudinal and lateral control are automated). The latter condition essentially constitutes a fully automated vehicle. Order of presentation of these conditions was randomised to counterbalance practice effects.

A number of dependent measures were recorded to assess performance, MWL, and attention. The SDS software records a wealth of primary task data, but for the purposes of the present experiment the variables considered to be most relevant to driving performance were speed, lateral position, and distance headway. Various derivatives of these measures have been successfully used in previous studies of driving performance (e.g., Bloomfield and Carroll, 1996; Fairclough, 1997; Verwey and Veltman, 1996; Wierwille and Gutmann, 1978). Furthermore, given the nature of the automated systems under examination here, measures of longitudinal and lateral control are most appropriate.

MWL was measured using a self-paced secondary task. This consisted of a rotated figures task (as used by Baber, 1991, and proved by Stanton et al., 1997, in the SDS), presented in the lower left corner of the screen. Each stimulus was a pair of stick figures (one upright; the other rotated through 0°, 90°, 180° or 270°) holding one or two flags. The flags were simple geometrical shapes, either squares or diamonds. The task was to make a judgement as to whether the figures were the same or different, based on the flags they were holding. Responses were made via buttons attached to the steering column stalks, and brief visual feedback was provided before presentation of the next stimulus. The subsidiary task technique was used in an attempt to measure spare attentional capacity. Participants were thus instructed to attend to the secondary task only when they had time to do so. The secondary task was visual-spatial, requiring a manual response, and, as such, was intended to occupy the same attentional resource pools as driving. This ensured that the task was indeed a measure of spare mental capacity (based on multiple resources theory), and not some alternative cognitive resource.

Attentional gaze was measured by means of an on-board miniature camera directed at the participant's face. Total time spent looking at the secondary task was coded by video analysis. The experimenter timed how long the participants' eyes were directed at the bottom left of the screen. These data were used to investigate the malleable resources hypothesis. This assumes that visual gaze reflects allocation of attention, as stated by Underwood and Everatt (1996). Static resource pools would predict that the proportion of time spent looking at the secondary task should correlate directly and positively with the number of correct responses. Any other form of association would imply that the size of attentional resource pools can change.

The design of this experiment serves--to a certain extent--to rule out competing explanations for the results. By monitoring performance on both primary and secondary tasks, it will be possible to determine whether boredom or motivation have influenced the results. If so, there should be a general decline in performance across both tasks. Malleable attentional resources theory, however, would predict a specific shrinkage in capacity (as inferred from the secondary task) while primary task performance is maintained at a constant level. The counterbalanced presentation of conditions also helps reduce the influence of fatigue. Finally, given the low levels of physical interaction with the simulator in all conditions, it was assumed that physiological arousal would be roughly constant throughout.

The Southampton Driving Simulator (SDS)

The SDS is a medium-fidelity, fixed-base driving simulator. The simulator consists of the front half of a Ford Orion. An Acorn Archimedes computer runs the

simulation and generates the display image. A medium-resolution colour monitor

displays a view of the road and a simulated instrument panel on a forward projection

screen, providing a visual area of approximately 40º horizontal by 20º vertical.

The SDS software records data at a rate of 2Hz. The following data are logged:

speed, lateral position on the road, distance from the vehicle in front, distance from

oncoming vehicle, steering wheel and pedal positions, and collisions. The simulator was

set up to run with automatic transmission at all times.

A more detailed description of the SDS, along with the results of a validation

study, can be found in Stanton, Young, Walker, Turner and Randle (2001).

Participants

There were 30 participants (17 male) in this experiment, with a mean age of 25.3

(SD = 6.53). All participants held a full UK driving licence, for 6.9 years on average (SD

= 5.92), and their mean annual mileage was 5650 (SD = 3985).

Participants were recruited via a poster campaign in and around the University of

Southampton, and an attempt was made to balance age and gender as far as possible. The

experiment met the ethical criteria of both the Department of Psychology at the

University of Southampton and the British Psychological Society.

Procedure

After entering the SDS, participants were first given a minimum 15 minute

practice run (this could be extended, if either participant or experimenter felt that it was

needed), to acclimatise to the conditions of driving a simulated vehicle. Following this,

experimental instructions were given, including advice on how to operate the automated devices and how to respond to the secondary task. To check that the participants had understood the nature of the secondary task, three example stimuli were presented prior to the experimental trials beginning. Once participants were sufficiently familiar with the operation of the simulator, the experimental trials would begin. In accordance with the design, there were four conditions, each lasting 10 minutes.

In all of the experimental conditions, participants were faced with a single-carriageway road which was a mixture of curved and straight sections. The track was quite simple, with no hills or wind gusts to disturb longitudinal or lateral control. Participants were instructed to first catch up and then follow a leading vehicle, which was travelling at a constant 70 mph (cf. Stanton et al., 1997), for the duration of the trial (10 minutes). There were no other vehicles in the participants' lane (so no overtaking was necessary), although oncoming traffic was encountered infrequently, encouraging participants to remain in their own lane. Participants were required to maintain a constant distance from the lead vehicle, although the choice of that distance was left to the individual. There were a number of advantages to this approach. Firstly, it meant that participants did not have to disengage the automatic devices (for instance, in order to overtake), thus avoiding contamination of conditions. Secondly, following a car motivated participants to drive at a relatively constant speed, thereby controlling objective demand across conditions. Otherwise, participants may have compensated for increased workload by reducing speed, which again would contaminate results. Finally, a constant speed implied that participants all drove approximately equal distances, again controlling for workload and attention differences which may otherwise have been

incurred. While driving, participants were expected to attend to the secondary task <u>only</u>

<u>when they felt able to do so</u> (this instruction was emphasised to participants in order to

minimise secondary task interference).

At the end of each 10 minute trial, participants were informed of the conditions

for the following trial. When all four trials were completed, participants were thanked for

their time and debriefed about the purpose of the experiment. The whole procedure

lasted approximately 75 minutes.

<u>Data reduction</u>

There are a number of driving measures which may be used as primary task

performance variables. These can be divided into measures of longitudinal and lateral

control.

Longitudinal control measures involve speed and headway. However, simple

measures of location (i.e., mean, median) do not necessarily provide evaluative

information about how well participants are performing. Given the instructions to

participants (maintain constant speed and headway), it would be logical to adopt a

measure of consistency (or rather, inconsistency) for these variables. Fortunately,

Bloomfield and Carroll (1996) described such a measure, in their derivation of <u>instability</u>.

"A linear equation that is the line of best fit for a series of points on the track of a vehicle

can be used to describe the position of the vehicle relative to the center of the lane"

(Bloomfield and Carroll, 1996; p. 336). A similar line can be calculated for vehicle

speed. The sampling rate of the SDS allows such equations to be calculated for the 1200

data points on each of the speed and headway variables. The standard error around this

line represents the driver's ability to maintain stability in the measure. This is a better

measure of driving performance than standard deviation, as it reflects the drivers'

consistency in their own performance, rather than deviation from an absolute measure (J.

R. Bloomfield, personal communication, December 15 1999).

　　　　For lateral control, it was considered that instability measures would not be an

appropriate reflection of driving performance on a road which involves both curved and

straight sections. Popular measures of lateral control (such as instability, RMS error, or

time-to-line-crossing) assume that 'good' driving performance is characterised by the

vehicle remaining consistently in the centre of the lane. These measures would be

confounded if participants used modern driving techniques to negotiate the curves on the

road. Instead, then, simple measures of lane excursions were used to evaluate lateral

control, with the assumption then being that good driving performance is rewarded with

fewer lane excursions. Total number of lane excursions, and time spent out of lane, were

the dependent variables for lateral control. All of the driving performance measures were

filtered for outliers and extreme values, and these data points were removed prior to

analysis (i.e., any values outside two standard deviations from the mean).

　　　　For an assessment of the malleable resources hypothesis, a more complex analysis

was required. Video analysis was performed on a subset of participants ($\underline{N} = 20$). For

the remaining participants, the video data were not clear enough to make a reliable

recording (e.g., due to the participant wearing glasses). The number of correct responses

on the secondary task was divided by the total duration of glances directed at that task (a

measure which can vary independently from secondary task responses), and this

composite score is referred to here as an <u>attention ratio</u> (see Figure 1). A null hypothesis

of no difference between the automation conditions on this score would be consistent

with a model of fixed attentional resource pools. If resources fluctuate with MWL, then

the pattern of attention ratio scores should reflect the pattern of MWL results.

$$AR = \frac{ST_{cr}}{ST_t}$$

Where AR = Attention Ratio
ST = Secondary Task
cr = correct responses
t = time

Figure 1. Derivation of Attention Ratio, used to infer attentional resource capacity.

Number of correct responses on secondary task were divided by total duration of glances

directed at that task

RESULTS

Primary task data

Each of the four evaluative performance variables (number of lane excursions,

time spent out of lane, speed instability, headway instability) was evaluated via ANOVA

with 'condition' as a within-subjects factor. Simple contrasts, with the Manual condition

as the reference category, were used to determine the nature of any significant effects.

Since we were interested in the effects of automation on individual components of the

driving task, this was preferred over a 2x2 ANOVA collapsing across lateral or

longitudinal control. By using simple contrasts, the Manual condition serves as a

baseline by which to compare any performance changes due to automation. Descriptive

statistics for evaluative performance variables for which a significant result was observed

are in Table 1 (nonsignificant data are not presented for reasons of clarity).

TABLE 1

Descriptive statistics for evaluative performance variables. Means are shown in regular type, standard deviations are shown in italics.

Cells: Mean				
<u>SD</u>	<u>Manual</u>	<u>ACC</u>	<u>AS</u>	<u>ACC+AS</u>
Number of lane	42.0	39.7	1.04	1.61
excursions	<u>16.5</u>	<u>14.2</u>	<u>1.00</u>	<u>1.37</u>
Time spent out	111.1	101.7	2.68	4.69
of lane	<u>48.8</u>	<u>52.1</u>	<u>3.45</u>	<u>4.73</u>
Speed	10.2	9.17	8.92	7.13
instability	<u>3.38</u>	<u>3.31</u>	<u>1.73</u>	<u>0.741</u>

Turning first to the measures of lateral driving performance, number of lane excursions produced a significant main effect of automation ($\underline{F}(3,72) = 163.2$, $\underline{p} < 0.001$). There was no difference between the Manual and ACC conditions ($\underline{F}(1,24) = 0.004$, $\underline{p} = 0.950$), however number of lane excursions significantly reduced in the AS ($\underline{F}(1,24) = 192.2$, $\underline{p} < 0.001$) and ACC+AS conditions ($\underline{F}(1,24) = 194.9$, $\underline{p} < 0.001$). The existence of lane excursions in the AS conditions is most likely due to noise in the data which escaped the filtering process (e.g., variation in lateral position before the system had been engaged).

Time spent out of lane displayed a similar pattern, with a significant main effect of automation ($\underline{F}(3,66) = 69.1$, $\underline{p} < 0.001$). Again, the difference between Manual and

ACC conditions was nonsignificant ($F(1,22) = 1.73$, $p = 0.202$), whilst time spent out of

lane significantly reduced in both the AS ($F(1,22) = 96.0$, $p < 0.001$) and ACC+AS

conditions ($F(1,22) = 88.2$, $p < 0.001$). As with number of lane excursions, the fact that

there are some nonzero values in the AS conditions is most likely due to variations in

performance before the system had been engaged.

The analysis of speed instability also demonstrated a main effect of automation

($F(3,63) = 7.07$, $p < 0.001$). Compared to Manual driving, speed instability only

decreased significantly in the ACC+AS condition ($F(1,21) = 16.3$, $p < 0.005$), although a

marginal decrease was observed when using AS ($F(1,21) = 3.88$, $p < 0.1$). The test

between Manual and ACC conditions was nonsignificant ($F(1,21) = 2.16$, $p = 0.156$).

Finally, the ANOVA for headway stability did not produce a significant main effect

($F(3,63) = 1.78$, $p = 0.161$).

Secondary task data

The dependent variable for secondary task data was number of correct responses.

This was determined to be the most appropriate measure of spare capacity, as it reflected

the level of concentration participants were investing in the secondary task. Total

number of responses may have been inflated if participants were randomly responding

without thinking about their answers. A visual inspection of the data, however, suggested

that error rate was quite constant (around 5%), and a statistical examination revealed no

difference between conditions ($F(3,57) = 1.33$, $p = 0.273$). Therefore, in the end it would

have made little difference to the analyses if the dependent variable was the number of

correct responses or total number of responses. Proportion correct would not have revealed any effects as a dependent variable, due to the consistency in error rates.

Number of correct responses was therefore entered into a repeated measures ANOVA, with level of automation as the within-subjects variable. This time, repeated contrasts were used to determine the nature of any differences, with the conditions ordered from fully manual to fully automated control (ordering of partial automation conditions was based on rank). Repeated contrasts were chosen for these MWL data because we were now concerned with any stepwise decreases in demand with increasing levels of automation. Remember throughout these analyses that an increase in the number of correct responses implies more spare attentional capacity, or lower MWL.

The number of correct responses on the secondary task demonstrated a significant main effect of level of automation ($\underline{F}(3,78) = 98.1$, $\underline{p} < 0.001$). Repeated contrasts revealed that there was no difference between Manual and ACC conditions ($\underline{F}(1,26) = 1.16$, $\underline{p} = 0.291$), but there were stepwise increases in responses from ACC to AS ($\underline{F}(1,26) = 49.8$, $\underline{p} < 0.001$), and from AS to ACC+AS conditions ($\underline{F}(1,26) = 51.4$, $\underline{p} < 0.001$). Mean numbers of responses in each condition were 108.5 (Manual; $\underline{SE} = 9.36$), 111.3 (ACC; $\underline{SE} = 9.25$), 181.8 (AS; $\underline{SE} = 8.43$), and 214.3 (ACC+AS; $\underline{SE} = 5.97$). These data are represented graphically in Figure 2.
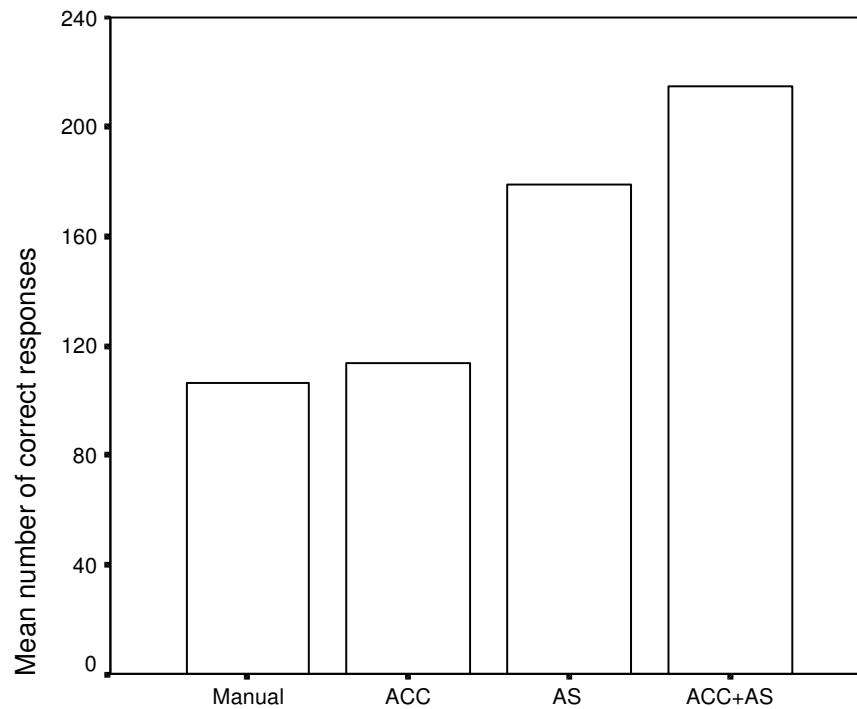
Figure 2. Secondary task scores in each condition. Higher score implies more spare attentional capacity, and thus lower MWL

Attention ratio data--the test for MART

The attention ratio is derived from the secondary task score and the amount of visual attention directed at the secondary task. The latter is gathered from the video data, which could not be analysed for all participants. Therefore, the attention ratio analysis is performed on a subset of participants. Before presenting that analysis, it is necessary to repeat the ANOVA for the secondary task data within this subset, to check that the secondary task patterns are consistent with the main group.

Unsurprisingly, there was still a main effect of automation on secondary task performance ($\underline{F}(3,57) = 64.9$, $\underline{p} < 0.001$). As before, there was no difference between the Manual and ACC conditions ($\underline{F}(1,19) = 0.245$, $\underline{p} = 0.627$), although responses did increase in the AS condition ($\underline{F}(1,19) = 26.7$, $\underline{p} < 0.001$), and further in the ACC+AS condition($\underline{F}(1,19) = 45.7$, $\underline{p} < 0.001$). Mean numbers of correct responses in each condition were 103.4 (Manual; $\underline{SE} = 11.9$), 107.2 (ACC; $\underline{SE} = 11.4$), 170.3 (AS; $\underline{SE} = 10.7$), and 207.7 (ACC+AS; $\underline{SE} = 6.81$). In sum, the pattern of secondary task responses within this subset directly reflected that of the main sample. Therefore, the analysis of attention ratio may proceed with confidence.

The attention ratio data were entered into a repeated measures ANOVA. Here, as with the MWL data from the secondary task, repeated contrasts were used to determine if adjacent levels of automation were significantly different from one another.

A significant main effect of automation was found ($\underline{F}(3,57) = 34.7$, $\underline{p} < 0.001$). Repeated contrasts revealed no difference between Manual and ACC conditions ($\underline{F}(1,19) = 1.02$, $\underline{p} = 0.325$), however a significant reduction in attention ratio was observed from ACC to AS ($\underline{F}(1,19) = 14.2$, $\underline{p} < 0.005$) and from AS to ACC+AS ($\underline{F}(1,19) = 37.9$, $\underline{p} < 0.001$). These effects are illustrated in Figure 3.
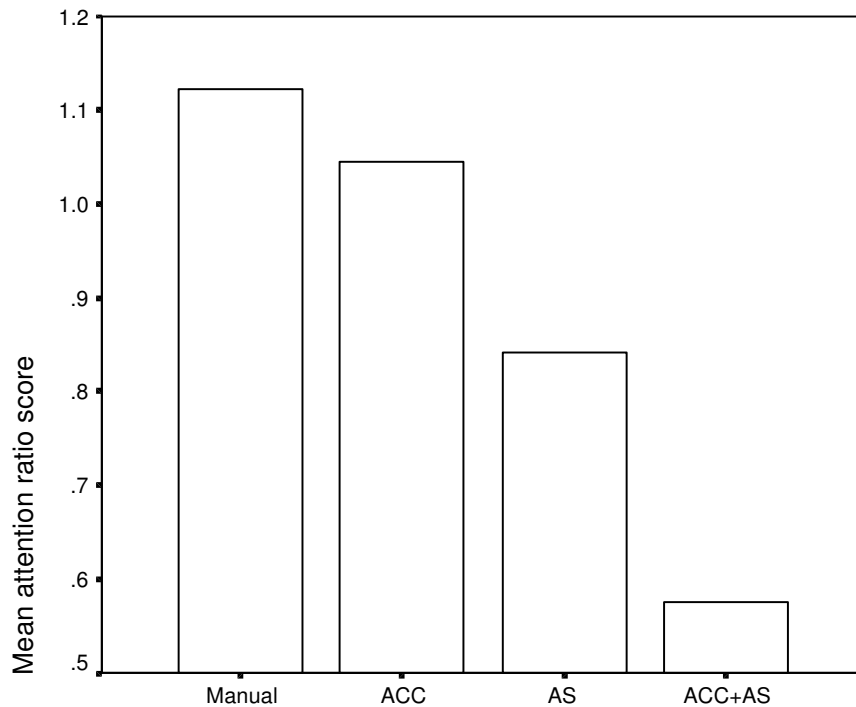
Figure 3. <u>Attention ratio score in each condition. Lower score implies smaller attentional resource capacity</u>

These are striking data, directly in line with the secondary task data themselves.

It appears that, when MWL decreases, the allocation of attention to the secondary task

becomes less efficient. This could either represent shrinkage of attentional resources (as

predicted by MART), or simply a change in strategy by the participants--perhaps

reflecting a speed-accuracy trade-off. However, since the analysis of secondary task

error rates found that the percentage of correct responses remained stable in all four

conditions, it is unlikely that such a strategy change was occurring. Therefore, malleable

attentional resources theory seems to be the most likely explanation for these data.

DISCUSSION

Implications: Mental workload and performance

The present results indicate that automation does indeed have a significant effect on driver MWL, although the specifics of this effect depend on the level of automation. Nilsson (1995) found no significant differences in subjective MWL between manual and ACC-supported driving. Yet in the study by Stanton et al. (1997), ACC reduced workload as evidenced by the secondary task results. The present results favour Nilsson's (1995) conclusions, as there was no effect of ACC on the secondary task.

It is apparent that in this situation, AS has a far greater influence on workload than ACC. Again, this finding contradicts Desmond et al. (1998), who found that even a combination of lateral and longitudinal automation did not affect subjective MWL. However, the present findings are in accordance with the reduced subjective demands of an automated highway system, reported by de Waard et al. (1999). Furthermore, an analysis of vehicle dynamics demonstrates that steering is a second-order tracking task (i.e., control of acceleration; cf. Wickens, Gordon and Liu, 1998), whereas headway control is first-order (i.e., velocity control). Since tracking difficulty increases with control order, it makes sense that AS should relieve MWL to a greater extent than ACC.

In general, this experiment suggests that there are no adverse consequences of reductions in MWL for performance. Whilst automated lateral control reduced MWL as measured by the secondary task, this did not have any effect on speed and headway maintenance (the remaining manual subtask in the AS condition). In comparison with

previous studies, this does little to clear the water, as there have been contradictory

results regarding the effects of automation on MWL and performance. Both Desmond et

al. (1998) and Nilsson (1995) found degraded performance with automation in the

absence of reductions in MWL, whilst de Waard et al. (1999) and Stanton et al. (1997)

found that automation affected MWL as well as performance. However, each of these

experiments employed critical driving scenarios (e.g., automation failure), and these may

have influenced the subjective ratings used by Desmond et al. (1998) and Nilsson (1995).

In the present study, normal driving performance was assessed, so the secondary task

may have been more sensitive to MWL differences between the automation conditions.

The fact that reduced MWL did not affect performance may have been due to the fact that

normal driving (rather than responses to critical situations) was under scrutiny, and

attentional resources were shrinking to <u>match</u> task demands, as MART predicts.

Implications: Malleable attentional resources theory

        The attention ratio score is possibly the single most important result to emerge

from this experiment, and it provides the first piece of evidence in favour of Malleable

Attentional Resources Theory. The fact that participants' responses on the secondary

task do not vary consistently with the amount of attention they are directing to the task

suggests that the size of the resource pool can change. On the basis of the malleable

attentional resources hypothesis, it was expected that the attention ratio score would

decrease in line with the MWL data from the secondary task. This prediction was

directly upheld by the observed data, providing strong evidence for an association

between task demands and attentional resource capacity.

22

These are very encouraging results for MART. Further support is provided by the primary task performance data, as these reductions in demand are not accompanied by changes in driving performance. It could be argued that these results are due to different attention allocation strategies, or a qualitative change in the driving task (from active operator to passive monitor), allowing more time to be devoted to the secondary task in the light of a perceived reduction in driving demands. If participants' allocation policies were inappropriate to the relative task demands though, either a decrement in driving performance or an improvement in secondary task performance should be observed. This was not the case--driving performance remained constant regardless of attention ratio score, and no improvement in secondary task error rate was observed. Therefore, all attention devoted to the secondary task really did represent <u>spare</u> capacity. Furthermore, the fact that driving performance did not improve with reductions in task demands implies that <u>all</u> spare capacity was allocated to the secondary task. It is reasonably safe to assume, then, that the sum of primary and secondary task demands reflected the total attentional capacity of the driver. Given this assumption, and the fact that increases in secondary task scores were not proportional to increases in visual attention, it is logical to conclude that attentional capacity had shrunk.

An alternative perspective might appeal to the classic vigilance decrement (e.g., Mackworth, 1948; Singleton, 1989) as an explanation for the results obtained here. However, the present experimental design does not qualify it as a vigilance task. Observations elsewhere (Singleton, 1989; Warm, Dember and Hancock, 1996) typically find that a vigilance decrement sets in after 20-30 minutes. Given the 10 minute trials in the current study, it is unlikely that vigilance would have caused a problem.

Furthermore, Parasuraman (1987) argued that continuous, dynamic tasks do not lend themselves to vigilance problems, and it is easily arguable that the task of driving fits these criteria. Therefore, malleable attentional resources theory seems to be a more likely explanation for these data.

Other alternatives relate to the nature of the task, and centre around issues of motivation and arousal. One might suggest that participants were simply bored or less motivated to maintain performance on the secondary task in the underload conditions. If this were the case, it would be expected that a lack of motivation would have a general effect on performance. Since performance on the primary (driving) task was not affected, the balance of evidence favours MART. Similarly, although physiological arousal was not measured in the present study, all of the experimental conditions posed fairly equal levels of physical demand. There was no reason to believe a priori that physiological arousal would vary, and the counterbalanced conditions should have mitigated any confounding effects of motivation or arousal. However, it is acknowledged that mental demands might only have influenced attentional capacity via an effect on arousal. Future research could strengthen the argument for MART by including measures of motivation and arousal.

From a theoretical point of view, the possible influence of arousal actually gives MART more support from the established literature. Kahneman's (1973) original model of attention explicitly identified physiological arousal as a factor in attentional capacity. If future investigations into MART cannot dissociate mental demands from arousal in determining resource size, then the contribution of this research will be limited to resurrecting ideas that have been glossed over in the applied literature since. In these

authors' opinion, such a contribution is no less valuable than validating the new theory
itself. Indeed, there are practical advantages in the link with arousal, since it is more
readily detected than mental underload, and may help to design monitoring systems.

In terms of the design implications, then, one might controversially conclude that
drivers should be given extra tasks in their automated cars in order to maintain capacity at
optimal levels. Of course, this denies all best practice in human factors, which prescribes
a holistic approach to designing systems around the driver. Most ergonomists now agree
that systems should be designed to support the driver, rather than replace them (Alm,
Svidén and Waern, 1997; Grote, Weik, Wafler and Zolch, 1995; Stanton and Young,
1998). The technology behind ACC could feasibly be used to present distance and/or
relative speed information to the driver, rather than actively controlling the vehicle itself.
This solution would also satisfy the driver's need for feedback, a classic problem with
automated systems (Norman, 1990).

CONCLUSIONS

The analyses presented here provide a detailed insight into the relationship
between automation, MWL, and performance, and strong support for the proposed theory
of malleable attentional resources was found. The data on attentional gaze provided a
very convincing indication that the size of attention resource pools can shrink directly in
line with reductions in MWL. If these results can be replicated, and the pattern of
attentional capacity with MWL adequately mapped, then MART will really begin to have
some predictive power. However, it remains to be seen whether the 'strong' version of

MART, which favours the role of task demands over arousal or effort in determining

attentional capacity, will be supported.

It has been demonstrated thus far that the introduction of automation into the

automobile can significantly reduce driver MWL, and that this has the potential of

affecting the attentional capacity of drivers. Previous research suggests that in

automation failure situations, many drivers cannot reclaim control of the vehicle, and a

collision is the inevitable result (Nilsson, 1995; Stanton et al., 1997). Given the theories

being developed in this paper, the reason for such performance decrements would be a

reduced ability to devote appropriate levels of attention to the situation. This potentially

offers a definitive explanation of mental underload, as well as the opportunity to make

practical predictions for performance with automation. Further research in the SDS has

made a structured investigation of responses to automation failure, and this will be

reported in a later paper. By linking together these strands of information, it is hoped that

a body of evidence can be built to refine and uphold Malleable Attentional Resources

Theory as proposed in this paper.

REFERENCES

Alm, H., Svidén, O., & Waern, Y. (1997). Cognitive ITS: On cognitive integration of ITS

   functions around the driver's task. In Y. I. Noy (Ed.), Ergonomics and safety of

   intelligent driver interfaces. (pp. 231-237). Mahwah, NJ: Lawrence Erlbaum

   Associates.

Baber, C. (1991). Speech technology in control room systems: a human factors

   perspective. Chichester: Ellis Horwood.

Bloomfield, J. R., & Carroll, S. A. (1996). New measures of driving performance. In S.

   A. Robertson (Ed.), Contemporary ergonomics 1996. (pp. 335-340). London:

   Taylor & Francis.

Brookhuis, K. A. (1993). The use of physiological measures to validate driver

   monitoring. In A. M. Parkes & S. Franzen (Eds.), Driving future vehicles. (pp.

   365-376). London: Taylor & Francis.

Buck, J. R., Payne, D. R., & Barany, J. W. (1994). Human performance in actuating

   switches during tracking. International Journal of Aviation Psychology, 4(2), 119-

   139.

de Waard, D., van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver

   behavior in an emergency situation in the Automated Highway System.

   Transportation Human Factors, 1, 67-82.

Desmond, P. A., Hancock, P. A., & Monette, J. L. (1998). Fatigue and automation-

   induced impairments in simulated driving performance. Transportation Research

   Record, 1628, 8-14.

Desmond, P. A., & Hoyes, T. W. (1996). Workload variation, intrinsic risk and utility in

a simulated air traffic control task: evidence for compensatory effects. Safety

Science, 22(1-3), 87-101.

Ephrath, A. R. & Young, L. R. (1981). Monitoring vs. man-in-the-loop detection of

aircraft control failures. In J. Rasmussen & W. B. Rouse (Eds.), Human detection

and diagnosis of system failures (pp. 143-154). New York: Plenum Press.

Fairclough, S. (1997). Monitoring driver fatigue via driving performance. In Y. I. Noy

(Ed.), Ergonomics and safety of intelligent driver interfaces (pp. 363-379).

Mahwah, NJ: Lawrence Erlbaum Associates.

Grote, G., Weik, S., Wafler, T., & Zolch, M. (1995). Criteria for the complementary

allocation of functions in automated work systems and their use in simultaneous

engineering projects. International Journal of Industrial Ergonomics, 16(4-6), 367-

382.

Hancock, P. A., & Caird, J. K. (1993). Experimental evaluation of a model of mental

workload. Human Factors, 35, 413-429.

Hancock, P. A., & Parasuraman, R. (1992). Human factors and safety in the design of

Intelligent Vehicle-Highway Systems (IVHS). Journal of Safety Research, 23(4),

181-198.

Hancock, P. A., & Verwey, W. B. (1997). Fatigue, workload and adaptive driver systems.

Accident Analysis and Prevention, 29(4), 495-506.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. Journal

of Experimental Psychology: General, 108(3), 356-388.

Humphreys, M. S., & Revelle, W. (1984). Personality, motivation, and performance: a

    theory of the relationship between individual differences and information

    processing. Psychological Review, 91(2), 153-184.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Karashima, M. & Saito, M. (2001). A Study on the Error Occurrence and Human

    Information Processing Time Influenced by the Fluctuation of Working Memory

    Resource Capacity. International Journal of Cognitive Ergonomics, 5(2), 91-109.

Kessel, C. J., & Wickens, C. D. (1982). The transfer of failure-detection skills between

    monitoring and controlling dynamic systems. Human Factors, 24(1), 49-60.

Liao, J., & Moray, N. (1993). A simulation study of human performance deterioration

    and mental workload. Le Travail Humain, 56(4), 321-344.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search.

    Quarterly Journal of Experimental Psychology, 1, 6-21.

Nilsson, L. (1995). Safety effects of adaptive cruise control in critical traffic situations.

    Proceedings of the second world congress on intelligent transport systems: Vol. 3

    (pp. 1254-1259).

Norman, D. A. (1990). The 'problem' with automation: inappropriate feedback and

    interaction, not 'over-automation'. Phil. Trans. R. Soc. London B, 327, 585-593.

Parasuraman, R. (1987). Human-computer monitoring. Human Factors, 29, 695-706.

Richardson, M., Barber, P., King, P., Hoare, E., & Cooper, D. (1997). Longitudinal driver

    support systems. Proceedings of Autotech '97 (pp. 87-97). London: IMechE.

Singleton, W. T. (1989). The mind at work: psychological ergonomics. Cambridge:

    Cambridge University Press.

Stanton, N. A. & Young, M. S. (1998). Vehicle automation and driving performance. Ergonomics, 41(7), 1014-1028.

Stanton, N. A., Young, M., & McCaulder, B. (1997). Drive-by-wire: The case of driver workload and reclaiming control with adaptive cruise control. Safety Science, 27(2/3), 149-159.

Stanton, N. A., Young, M. S., Walker, G. H., Turner, H., & Randle, S. (2001). Automating the driver's control tasks. International Journal of Cognitive Ergonomics, 5(3), 221-236.

Underwood, G., & Everatt, J. (1996). Automatic and controlled information processing: The role of attention in the processing of novelty. In O. Neumann & A. F. Sanders (Eds.), Handbook of perception and action. (pp. 185-227). London: Academic Press.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: a comparison of nine workload assessment techniques. Journal of Experimental Psychology: Applied, 2(3), 270-285.

Ward, N. J., Fairclough, S., & Humphreys, M. (1995). The effect of task automatisation in the automotive context: a field study of an autonomous intelligent cruise control system. International Conference on Experimental Analysis and Measurement of Situation Awareness. November 1, Daytona Beach, Florida.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications. (pp. 183-200). Mahwah, NJ: Lawrence Erlbaum Associates.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R.

Davies (Eds.), Varieties of attention (pp. 63-101). New York: Academic Press.

Wickens, C. D. (1992). Engineering psychology and human performance (2nd edition).

New York: Harper Collins.

Wickens, C. D., Gordon, S. E. & Liu, Y. (1998). An introduction to human factors

engineering. New York: Longman.

Wickens, C. D., Xu, X., Helleberg, J. & Marsh, R. (2001). Pilot visual workload and task

management in freeflight: A model of visual scanning. 11th International

Symposium on Aviation Psychology. Columbus, Ohio.

Wierwille, W. W., & Gutmann, J. C. (1978). Comparison of primary and secondary task

measures as a function of simulated vehicle dynamics and driving conditions.

Human Factors, 20, 233-244.

Young, M. S., & Stanton, N. A. (1997a). Automotive automation: Investigating the

impact on drivers' mental workload. International Journal of Cognitive

Ergonomics, 1(4), 325-336.

Young, M., & Stanton, N. (1997b). Taking the load off: Investigating the effects of

vehicle automation on mental workload. In S. A. Robertson (Ed.), Contemporary

Ergonomics 1997 (pp. 98-103). London: Taylor & Francis.

Young, M. S., & Stanton, N. A. (1999a). Look--no hands! Potential effects of future

vehicle automation on driver attention. Proceedings of Vision in Vehicles VIII (in

press).

Young, M. S., & Stanton, N. A. (1999b). Miles away: A new explanation for the effects

of automation on performance. In M. A. Hanson, E. J. Lovesey & S. A. Robertson

(Eds.), Contemporary Ergonomics 1999 (pp. 73-77). London: Taylor & Francis.

Young, M. S. & Stanton, N. A. (2001a). Mental workload: theory, measurement, and

application. In W. Karwowski (Ed.), International encyclopedia of ergonomics

and human factors: Volume 1 (pp. 507-509). London: Taylor & Francis.

Young, M. S., & Stanton, N. A. (2001b). Size matters. The role of attentional capacity in

explaining the effects of mental underload on performance. In D. Harris (Ed.),

Engineering Psychology and Cognitive Ergonomics Volume 5: Aerospace and

Transportation Systems (pp. 357-364). Aldershot: Ashgate.

Young, M. S. & Stanton, N. A. (2002). Attention and automation: New perspectives on

mental underload and performance. Theoretical Issues in Ergonomics Science,

3(2), 178-194.