CRITIQUE OF BRUNEL DRIVER DISTRACTION STUDY

By Dr Paul Barber

September 2008

commissioned by

Outdoor Advertising Association
Summit House
27 Sale Place
London, W2 1YR

FOREWORD

In recent years outdoor advertising has played an ever increasing role within the UK advertising industry. The Outdoor Advertising Association (OAA) exists to ensure that all of its members, some 95% of the total industry in revenue terms, are responsible and are treated in a fair and equitable way.

Outdoor advertising is controlled in the interests of amenity and public safety under the provisions of the 2007 Advertisement Regulations. Outdoor advertising is a visual medium and it has often been suggested that advertising has the potential to distract drivers and thereby create a road safety hazard. The OAA is concerned to ensure that statements regarding advertising and driver distraction are supported by robust research to ascertain the validity of such opinions. There have over the years been a number of research projects aimed at examining or establishing whether there is a causal link between the occurrence of road traffic accidents and the location of advertising displays. None of the research projects reveal any empirical evidence to support the often voiced view that roadside advertising is, by its very nature, a distraction that causes accident. Additionally, to date, any research that has been carried out around the world looking into this issue has been, at the very least, inconclusive.

In 2007 researchers at Brunel University, sponsored by The Rees Jeffreys Road Fund, undertook a simulator based research study to examine driver reaction to roadside advertising. The published document is entitled “Driven to Distraction: Determining The Effects Of Roadside Advertising On Driver Attention”. The study results suggested, based on a small sample group, that in some circumstances roadside advertising can present a distraction to drivers.

More recently the Highways Agency commissioned WSP to conduct a more thorough report, which was published in July 2008 entitled “The Impact of Roadside Advertising on Driver Distraction.” A representative from the Highways Agency summarised the report findings thus;

“The aims of the study (Highways Agency) were to review existing policy, guidance and previous research and explore with stakeholders and members of the public whether and how much roadside advertising acts as a distraction when driving on the strategic road network. The outcome of the research has been generally inconclusive. Although the impact of roadside advertising on road safety has been researched at length, the broad range of views investigated/considered showed no overall consensus. It is widely perceived that adverts were a distraction to some drivers in certain circumstances, and as such, could potentially pose a road safety risk. However, the level or significance of this risk, has not to date been quantified. The lack of data or scientific evidence means that the case against roadside advertising generally relies upon anecdotal evidence and a perception that it can be distracting.”

The HA report looked at earlier research papers and caste doubt over the veracity of the Brunel Study, stating the following:

“Whilst the study (Brunel) has provided some new evidence on the impact of roadside advertising on driver distraction, it is considered to be inconclusive on some elements. Principally, as the authors did not complete a long term study but rather a short term study where each trial only lasted between five to six minutes, each participant would have sat only a 15 minute trial in total. Therefore the results may only be representative of a relatively short driving journey that would not closely reflect the more long distance nature of inter-urban journeys for example.”

The OAA is aware that despite the shortcomings of the Brunel Study and the conclusions of the Highways Agency on the report's weaknesses, it is occasionally held up as authoritative and conclusive evidence proving an irrefutable link between roadside advertising and road traffic accidents. For this reason, and to provide our organisation and Members with a critical assessment of the Brunel Report, the OAA commissioned a detailed critique from Dr Paul Barber, Emeritus Reader in Psychology at Birkbeck College, London.

I commend Dr Barber's paper to all practitioners and those involved in the outdoor advertising sector. The unequivocal message from Dr Barber is that the Brunel Study contains many flaws and for that reason cannot be relied upon or treated as a definitive piece of research that proves roadside advertising prejudices road safety.

Alan James
Chief Executive
Outdoor Advertising Association

April 2009

Dr Paul Barber was employed at Birkbeck in the University of London where for over 30 years he taught courses on perception, psychological statistics and experimental design, computing and ergonomics. He was head of Birkbeck's Department of Psychology from 1988 to 1992.

He is now Emeritus Reader in Psychology at Birkbeck. Dr Barber was research supervisor to 16 PhD students, his own doctoral research being on visual input processes. He is an Associate Fellow of the British Psychology Society and is a Chartered Psychologist.

Dr Barber was author/co-author of a number of psychology textbooks and many research papers in refereed scientific journals; he was Psychology Editor for the journal *Ergonomics* for over ten years.

He has been a consultant for Postar since 1995, closely identified with its programme of research on poster panel visibility.

CRITIQUE OF BRUNEL DRIVER DISTRACTION STUDY - "DRIVEN TO DISTRACTION: DETERMINING THE EFFECTS OF ROADSIDE ADVERTISING ON DRIVER ATTENTION"

The Brunel Study is the latest of a series of reports on driver distraction. Most of the previous work has been on internal distraction (mobile phones, in-vehicle navigation devices, etc.) but a number of investigations have also addressed external factors, including roadside advertising. The report's sub-title states the objective of the study clearly, namely "to determine the effects of roadside advertising on driver attention". The authors contend that their study shows that "roadside advertising has a clear detrimental effect on lateral control, increases mental workload and eye fixations". The evidence for these effects will be reviewed in this critique of the study, along with other crucial aspects of the research. It will be argued that the evidence is weak in most key respects, equivocal in others, and based on a flawed methodology.

CORE ISSUES

There are a number of core issues in evaluating this report:

- Does the simulated environment satisfactorily resemble a driver's real visual world and the objects it contains?
- Are the measures used to assess performance satisfactory?
- Do the results from the simulator in general map satisfactorily on to the behaviours of drivers on the road?
- With regard to the specific aspects addressed by the researchers, are the results believable?

In addition to these general questions that are accessible to the generalist reader, there are several technical problems in the report regarding the design and conduct of the research and the analysis of the results, which are not subjected to close scrutiny here. While the review does not avoid conclusions of its own, it is intended to aid neutral readers to draw their own judgements about the research.

1. DRIVING SIMULATOR AS METHOD OF CHOICE

It is fundamental to this study that it used a driver simulator to represent the task as well as the natural circumstances of driving, and to collect data on the performance of driving. The study could succeed only insofar as it achieved these goals convincingly. The authors do not present a rationale for their choice of methodology, and its pros and cons will not be reviewed fully here (for a balanced appraisal see Hole, 2007). Notwithstanding, it is important to see the significance of this for what the study delivers. It goes without saying that most is likely to be delivered by the most technically advanced simulators; the system used for this study, albeit claimed by the authors to be of "high-fidelity", does not approach the standards that are achievable[1].

The driving environment

[1] E.g., the Toyota driving simulator – http://en.wikipedia.org/wiki/Driving_simulator; see also http://www.youtube.com/watch?v=CZ5uqMLcbGo&feature=related.

The shortfall between the naturalistic world of driving and that portrayed by the Brunel study is suggested by the illustrations in the Brunel report (Figures 1-3); these have the air of a set of cartoon frames, samples from the world of the video game. This is not to say nothing can be learned by the use of such imagery, however, what is signalled in the present case is the gap between pretend and reality. The risk is that an invitation is offered to participants in the research to treat the exercise as a game, with footling consequences one way or another. This prospect is underlined by the fact that drivers can "crash" in the course of a simulated drive, seemingly with impunity. It is not a convincing framework for the research.

Tasks and driving requirements

Drivers were required to keep to a target speed for each road type. While 70 mph is perfectly appropriate for the UK Motorway condition, a limit of 40 mph would seem unusual for the Urban condition. The fact that the participants managed to drive at 38 mph in the latter condition suggests that the environment was not representative of the average UK urban situation, where mostly the limit is 30 mph and average driving speeds rarely exceed this. While London is the most extreme case in the UK (and Europe), it is appropriate to mention that the average speed was recently reported as 11.6 mph, because this is where the participants were recruited.

It is well-known that advertising billboards are not allowed on the UK motorway system, and there are no moves to change this arrangement. Any that are there are illegal and do not have the backing of legitimate advertisers. To this extent the inclusion of the Motorway condition in the research is redundant.

Three sample screenshots of the routes are provided in the report. The cartoon-like appearance of the visual environment is typical of mid-range driving simulators. There were enhancements to help the resemblance to the real-world (such as an audio background and a device to suggest - but not create - vibration) but the task and presentation seems very much like a video game.

Research scale

There were 48 driver participants in the study, not many compared with the numbers used in survey style research but this would be typical of and normally sufficient for research of this sort. Obviously the study could be scaled up and the WSP Report for the Highways Agency (2008) notes that the sample was small.

The scale on which the research was done is also manifest in the amount of testing which was imposed on each participant. The simplest way of viewing this is to consider the total distance driven in the simulator and the total time spent on the task. The total distance was 2 x (3.0 + 5.7 + 2.8) = 23 miles and the average total time per participant was just under 30 minutes (in addition to what is termed a short practice run). Clearly the total exposure to the task overall and simulator environment is limited, and exposure to any one of the six test conditions was very brief indeed – a cursory look at the research literature will show that few published studies in experimental psychology or ergonomics would be so restricted. It is noteworthy that the 2008 Highways Agency report recommends an increase in the scale of future research (notably testing participants over a period of some days).

Demand characteristics

The concept of demand characteristics is used in psychology and similar disciplines to flag the possibility that research participants may unwittingly – and without any intent on the part of the researcher - come to behave in response to subtle cues as to

the researcher's expectations as to the outcome of the study. This may be in addition to or in opposition to any experimental variables that have been manipulated and hence the effect may be to supplement, mask or destroy the effect of the experimental variables. Researchers should naturally take steps to prevent those taking part in their research from acting in this way, which may mean a degree of concealment of purpose. In the Brunel study, the investigators no doubt did not at the outset declare their interest in billboards as a potential source of driver distraction; had they done so the participants would very likely have been on the look-out for billboards. It would also have been wise for the investigators not to mention in advance that recall of sign/poster information would be required, otherwise a similar alerting effect would have been probable. The problem is that once recall was tested, the drivers would have known that this might well happen again, as it did through the experiment. It is entirely possible that the participants came to use this information to structure their attentional resources in the course of the study. Indeed this could explain many if not all of the marginal effects reported.

2. MEASURES

Three sorts of variables were measured: there were five measures of driving performance, four eye-movement measures, one measure of recall, and one overall measure of task workload. The results are discussed below after the following assessment of the measures.

The variables used to quantify driving performance included two that depend on "time-to-contact" (TTC). This variable, in some contexts referred to as "time-to-collision", generally entails the presence of a lead vehicle or a specified obstacle, with which the test vehicle could come into contact (for example, if the lead vehicle stopped without warning). It is not clear how TTC was involved since no lead vehicle is mentioned that would presumably have been required for TTC to be measured, and none is apparent in the sample screenshots. In any case details should have been provided to explain TTC as implemented. It is of interest that an earlier study from the same laboratory used TTC with a lead vehicle, and with full details of the circumstances and how TTC was derived. What explains the omission is unclear, however, it is important to an understanding of the overall operational context of the study – the data may well, if the foregoing is correct, have been collected by the inclusion of critical events such as a vehicle or pedestrian arriving in view in the road ahead.

The researchers reported two measures, average and minimum TTC; these would undoubtedly be positively correlated, the average value would suffice, and it is not clear why minimum TTC was included. In assessing the contribution of these two variables, one should be mindful that a strong degree of redundancy would probably apply.

In addition there were two driving performance variables scoring time spent out of lane and number of lane excursions; these are also likely to be positively correlated, so their joint contribution should be assessed accordingly.

These measures seem fit for purpose, however, it is important to stress the overlap between the two TTC measures and between the two lane performance measures. In each case the findings may reinforce one another but they hardly amount to independent sources of evidence.

The number of "crashes" was the fifth driving performance measure that was noted for each condition. Although it is conventional enough to note episodes of this type during simulated driving, the possibility of such an event must always raise serious questions about the validity of the task. It is obvious that the simulated drive can be designed so that crashes occur with an incidence appropriate to the real-world to which the study is intended to apply. The authors are careful to avoid equating the idea of a simulator "crash" with an accident in the real-world, but it is inevitable that a link will be made and seems to be intended[2].

The problem is surely that the consequences of a simulator crash are completely different to that of a real crash, in terms of emotional impact, potential cost, disruption and maybe injury or life. Yet in a simulator the session may well continue, perhaps at the footling cost to the participant of momentary loss of face. The applicability of any findings from a simulator study must be constrained by considerations such as these. This will be discussed further in the section on Results.

It should also be noted that the number of crashes was recorded over the entire route. It is not clear why crash incidence – and the other driving performance measures – were not recorded contingent on passing the billboard(s) for comparison with data from the equivalent part of the no-billboard drive. This would have revealed potential effects in a suitably targeted fashion. Aggregating over the whole route allows overall extraneous factors – such as participants' desire to be cooperative - to influence the outcome.

The eye movement variables included the usual measures of glance duration (the time spent fixating in a target area) and number of fixations (in the target area). These were scored for each of the three 20 degree sectors into which the screen was divided. It is evident that duration and fixation frequency will tend to be negatively correlated. In the event over 90% of the fixations were in the central third of the screen, about 6% in the left third and 2.5% in the right third (duration data were not reported, but they would no doubt tend to be in similar proportions). It is not explained how these measures reflect driving or safety, and no criteria relating to fixation frequency are presented to decide when excessive risk has occurred. It is obvious that gazing away from the road may be risky, though failing to deal with hazards in the visual periphery can equally expose the driver to danger. It would seem reasonable, for example, to specify safe limits in terms of gaze duration and fixation frequency. One could, for example, contrast protracted gazing in order to read a notice with quick glancing to appraise a situation, which would require a duration criterion; a similar criterion relating to fixation frequency could also be adopted. In the absence of criteria such as these the evidence from these variables is equivocal.

The measure of task workload was the well-documented and much-used NASA-TLX *subjective* workload scale; this is commonly summated over a number of rating-scale items: mental demand, physical demand, temporal demand, performance, effort and frustration level). The aggregate scale score is often reported and analysed, along with sub-scale results since clearly they may pinpoint effects of interest. No details were provided of the potentially diagnostic sub-scale results, but the more probing analysis that might have been expected[3] was not undertaken. In the absence of any indications to the contrary, the effect on workload must have been restricted to the aggregate workload level, which leaves the explanation of the findings quite obscure.

[2] The first sentence of the Executive Summary states that "Distraction in driving is a frequently cited factor in crashes."

[3] Other reports from this laboratory have dealt intensively with workload and driving.

The measure of recall poses further difficulties. Participants were asked "at the end of the run to recall the last road sign they had passed and, in the case of billboard conditions, the last advert they saw". Presumably this means in the latter condition recall of just the last advert, but it could mean both traffic sign and advert, in which case a participant's recall strategy could affect his/her deployment of attention.

What is more important, however, is how recall was scored. To judge from the sample screenshots, the adverts carried the product name (i.e., Pepsi and 7 Up), so this verbal property could be used to define correct recall. How road sign recall was measured is a different matter, verbal recall of symbolic or directional information being much harder to score. It is not clear whether there were just four adverts that were permuted somehow, or how many road signs were used, or how many different signs were used. Indeed it is not clear how many road signs were encountered on each run although the Design section of the report indicates that the two versions were identical apart from the inclusion of billboards. For a proper comparison one would need an assurance that recall measurement was on a common basis; this seems unlikely on the basis of information in the report.

In any event, a chance level for recall would either exist in advance or it would develop during the experiment. Either way the interpretation of recall levels between conditions would be compromised or simply impossible. It therefore comes as no surprise that the recall results are patchy and almost incomprehensibly confused. To suggest that they hint at a trading relationship between billboard and road-sign recall is a step too far. The reader would not require much in the way of scepticism to decide to ignore these data.

3. RESULTS

The main statistical findings of the Brunel study are summarized in the table below. The data were mostly subjected to statistical treatment by the analysis of variance. The research design combined Road Type (Urban/Motorway/Rural) and Advert Condition (Billboard vs. no Billboard). The effect of each of these factors was assessed, as was their interaction effect (that is, their effect in combination. The interaction effect is an indication of whether the Billboard vs. no Billboard difference varies according to Road Type. It could for example be significantly greater for one road type than the other two, in which case the analysis would reveal an interaction between AC and RT. The Table summarizes the statistical conclusions that were drawn. The critical significance level was 0.05 (signifying a 1 in 20 chance level for the result); this is far from decisive and possibly a more stringent criterion should have been adopted, such as the 0.01 level (signifying a 1 in 100 chance). The authors also flag higher levels of significance (this is quite common practice) but no effects involving the billboard factor reach these higher significance levels. Indeed, when assessing the billboard factor, the authors have opted for an even more generous than customary significance level (that is $\alpha = 0.1$, signifying a 1 in 10 chance); this is another common practice but one that potentiates hazardous lines of reasoning. When this was done in the report, the following table shows the decision as NONSIG/SIG to underline their highly equivocal status. The point here is that while the authors implicitly invite the reader to decide that the finding is significant, a more cautious reading would be that it is not significant.

One central aim of the research is to influence policy and the onus is on the authors to obtain results that are clear-cut and not just statistically marginal. Had they applied the more stringent 0.01 significance level a quite different set of conclusions

would have been reached. All of the effects involving the Advert Condition factor would have been declared not statistically significant. What would remain would be the set of findings concerning Road Type.

Measure	Road Type (RT)		Advert Condition (AC)		Interaction RT x AC
Time out of lane	**SIG (p<0.001)**		**SIG (p=0.050)**		**NONSIG (p=0.117)**
Lane excursions	**SIG (p<0.001)**		**NONSIG/SIG (p<0.1)**		**NONSIG (p=0.195)**
Average Time-to-contact (TTC)	**SIG (p<0.001)**		**NONSIG (p=0.919)**		**NONSIG (p=0.959)**
Minimum TTC	**SIG (p<0.001)**		**NONSIG (p=0.325)**		**NONSIG (p=0.261)**
N crashes	**EYE-BALL TEST**		**EYE-BALL TEST**		**EYE-BALL TEST**
Glance duration left	**NONSIG**		**NONSIG**		**NONSIG**
Glance duration centre	**NONSIG**		**NONSIG**		**NONSIG**
Glance duration right	**NONSIG**		**NONSIG**		**NONSIG**
N fixations left	**SIG (p<0.001)**		**SIG (p<0.05)**		**NONSIG**
N fixations centre	**NONSIG**		**SIG (p<0.05)**		**NONSIG**
N fixations right	**SIG (p<0.05)**		**SIG (p<0.05)**		**NONSIG/SIG (p<0.1)**
Mental Workload	**NONSIG/SIG (p<0.1)**		**SIG (p<0.05)**		**NONSIG (p=0.480)**
Recall (ALL SIGNS)	**NONSIG?**		**NONSIG?**		**NONSIG?**
Recall (road-sign vs. advert)	**NON PARAMETRIC TESTS**		**NON PARAMETRIC TESTS**		**NON PARAMETRIC TESTS**

4. INTERPRETING THE RESULTS

The authors begin the Discussion by stating that “In terms of driving performance, the presence of billboards had a detrimental effect on lateral control, and also appeared to increase crash risk.” As is evident from the table, one measure of lateral control (time out of lane) was on the very margin of statistical significance and the other

(number of lane excursions) showed an even more marginal tendency in the direction of poorer control in the presence of billboards; both effects vanish if a suitably strict significance criterion is applied so the evidence does not seem strong enough for any conclusion about lateral control.

Crashes were too infrequent for statistical analysis, however, having noted this, the authors state that "there was an indication that more crashes occurred in the presence of billboards". In the Executive summary, this is stated unreservedly as "the presence of billboards … appeared to increase crash risk". The reality is that there was no statistical support for this, so this assertion is unwarranted and injudicious.

Crashes may have been too infrequent for analysis but the fact that they occurred at all must be discussed. Eight were recorded as the 48 drivers each traversed the requisite 23 miles, translating into 0.24 crashes per km overall. Comparison with official statistics is revealing; accident rates in 2003 on UK Motorways and A roads were 9 and 76 per 100 million vehicle km respectively. The corresponding crash rates in the Brunel study were 0.11 per km (Motorway) and 0.32 per km (Urban and Rural combined). The lower of these is more than 400 thousands times higher than the actual accident rates. The Urban/Rural figure suggests that crashes occurred – in the simulated driving conditions – once every 90 miles or so!
Such deviations from reality are not considered in the report. It is not trivial since it suggests the driving task was either much harder than real driving; or it was not treated as a serious exercise by the participants; or both. Even if the second possibility is discounted, reasonably enough, it is hard to see that the task could have the same significance as driving a real vehicle. In any event the above comparison suggests that the results may be seriously misleading, and possibly considerably overstate any deleterious effects of advertising.

5. OTHER COMMENTS AND CAVEATS

The Brunel report does not have the important status of having been refereed or peer-reviewed. This is not necessarily fatal to the authors' concerns with driver distraction but it could – and arguably does – mean that inadequacies in the research have gone unremarked and uncorrected. Indeed this review – which for clarity has ignored certain more technical points– suggests that a scientific journal would not find it acceptable for publication. The 2008 Highways Agency report treats the Brunel study as a possible model for further research, but does not articulate any technical criticisms beyond those of scale and the quality of the simulator used. These assessments are indeed valid, but as shown in the present report, the research is also disturbingly flawed in many other respects. The confidence with which its findings and conclusions are presented is unwarranted.

6. REFERENCES

HOLE, G. (2007) *The psychology of driving*. Psychology Press.

WSP Development and Transportation (2008) *The impact of roadside advertising on driver distraction: Final Report - Highways Agency.* WSP UK Limited, London.