PILOT STUDY FOR POSTAR INCLUSIVITY PROJECT
FINAL REPORT: FEBRUARY 2003

Paul Barber and Adrian Dickenson
School of Psychology
Birkbeck College
(University of London)

SUMMARY

Three tasks were used to explore the possibility of devising a data collection method that would:

(a) produce convincing measures of poster visibility;
(b) produce data that agree with the results from an established visibility method, such as the hit-rate measure from an eye-tracking method;
(c) be portable across poster environments.

The results of all three satisfied these criteria, as well as others relating to cost and availability. Two of them were search tasks in which the respondent was asked to make a response indicating that a poster in a scene had been found. Both entailed a check on the accuracy of the response after it had been made, by the respondent pointing to the location of the poster. The third was a perceptual judgement task in which the respondent made numerical assessments of the prominence of posters within scenes.

One of the search tasks and the prominence task provided data that showed high levels of reliability, and showed remarkably high levels of agreement with eye-tracking data, as well as being as sensitive to key variables (such as poster location and panel size) as the eye-tracking data. The second search task showed rather weaker relationships, which was attributed to the higher cognitive demands on the respondent imposed by the post-response component of the task. The data from a search task yield a hit rate-like measure, but a simple regression equation may be obtained to map the data from both search and prominence tasks on to the hit rates obtained from the more basic eye-tracking method.

This means that the prospect of a visibility currency conversion process is in view that meets the aim of inclusivity with respect to poster environments. The practical implications of the results are that both search and perceptual judgement tasks are viable for the purpose of estimating poster visibility, but that it may be optimal to target tasks to single objectives lest otherwise the task for the respondent becomes unwieldy and the utility of the results is compromised. Further research needs to extend the application of the new method(s) to new untested environments, to optimise the methods, to clarify any minor anomalies with the existing methods, and possibly to build on the study in developing a method to produce data from multiple poster scenes, and mobile/dynamic display poster technologies.

INTRODUCTION

The main aim of this study was to assess the feasibility of using a directed visual search task as a means of assessing poster visibility for the Inclusivity Project. The rationale for the method adopted was explained in a previous report that described a set of possible alternative approaches (Appendix 1). Six possible methods were outlined in that report and the pilot study is an implementation of a minor variation of "Approach 6".

Many variants of visual search tasks have been devised, but in the basic form of the task, as employed here the observer is asked to report whether a specified target is present in an extended array; the speed and accuracy with which targets are correctly reported are taken as the measures of search performance. As a post-search check, the observer may, as here, be asked to identify the actual location of the target. The search tasks adopted for the pilot research, and as proposed as Approach 6 were essentially as follows:

The task for the observer in the pilot investigation was to detect whether or not each of a series of photographic scenes contained a poster, and to respond accordingly as fast and accurately as possible. An additional requirement was to point to the location of the poster on the screen but this will be explained in detail in the Method section. A proportion (25%) of the scenes contained one (sometimes more than one) poster, and the scenes were selected to ensure a reasonable balance with respect to poster size and location. A fixed amount of time (a few seconds) was allowed for the task before the scene was removed from view, and the observer's search time was noted. Given an interval as long as used in the study, nearly 100% of the observers can find the poster in a given scene, provided there is no ambiguity about what is a poster and the target is not seriously obscured. Naturally fewer observers can detect the poster in that scene if the time is limited, and for example, within the first quarter-second of the interval the proportion who succeed in finding the target is considerably less than 100%. This percentage will clearly increase as the length of this "deadline" interval increases. A cut-off can be imposed at any convenient point between 0 and 5 seconds, and its unadjusted visibility score can be derived as the detection percentage, or "hit rate" for that cut-off. A poster panel that is very prominent for whatever reason (e.g., low surrounding clutter) gets a high score on this measure; indeed the method orders panels according to their visibility, and observers very likely need to fixate the panel to respond. This means that the "eyes on panel" requirement for a visibility index is met.

The task requires deliberate searching for a poster in contrast to what is required in an eye-tracking study, where the observer is given an unrelated task to have in mind while viewing the scene and posters are not mentioned as of interest or concern. The objective of the pilot research was to ascertain whether or not a directed visual search task could provide data that show a satisfactory degree of agreement with the results of an undirected eye-tracking study. The results of such a study exist in the form of the Postar pedestrian visibility study, so it was decided to use the same scenes as used for that study, with a new pool of observers but having to carry out a directed search task. It is possible that the results will not agree, but the relative visual salience of objects in any scene will not be affected by the type of task, and so it may be expected that observers would spot prominent posters quickly when asked to locate them, just as they would tend to fixate them when more casually inspecting the same scene while their eye movements were being monitored. Some modest level of agreement might be predicted on this account. The next stage may be to refine the method, possibly to increase the agreement by adjusting the task (imposing less direct demands, as in Approach 5 in Appendix 1). An alternative would be to find a statistical means of optimising the relationship between the results from the two methods, such as using a multiple regression technique, supplementing the search data with other measures of the scenes.

In anticipation of the possible eventual need to adjust the method, it was decided to add a more reflective method, recruited from the Postar maximum visibility distance study. This was a technique from classical psychophysics known as magnitude scaling. The observer is free to inspect the scene as long as needed, with the requirement to assess the ease with which a poster in the scene can be spotted. Speed and accuracy of response are not in issue. The observer is simply asked to "scale" the series of images according to the prominence of posters contained in them. The literature on such techniques is supportive of the view of the human brain as a highly valid and reliable measuring tool in just this sense. The question here is whether measurements of perceptual prominence would agree with those from the search or eye-tracking studies.

The rationale for expecting an agreement between data from the different tasks is that the eyes tend to land on regions that contain information that is of use or is essential to perception (Yarbus, 1967; Buswell, 1935). These early eye-tracking studies have been confirmed by research showing that the rate of fixating an item is related to its judged informativeness (Mackworth and Morandi, 1967; Antes 1974). A series of further studies, by Loftus and Mackworth (1978) and more recently by Henderson, Weeks and Hollingworth (1999), suggests that this may extend to semantic informativeness independently of visual informativeness. In short, there is strong evidence that what one observer looks at is also what another observer judges to be informative. Hence, it makes sense to assess whether this generalizes to viewers of scenes containing posters, and whether observers can reliably asses the prominence of posters in scenes. To the extent that they can, a metric of some generality may become devised that is generally useful to the industry, aside from the present investigation.

The two central sections of the report describe the methods and the results of the investigation, and they are structured according to the three tasks employed. The original specification of the research was to use a directed, speeded, visual search task; two additional tasks were devised, one a variant of the search task, the other a modified version of a task used in the visibility distance study (Barber and Dickenson, 1997).

METHOD

Two types of task were used, a visual search or reaction time task and a perceptual judgement task. The targets for the search task were posters in photographic scenes presented on a computer monitor. The task was to locate the target (if any) as quickly as possible, and then (as a check) to indicate its location on screen. There were two variants of this checking process. The perceptual judgement task used the same scenes, the task being to rate the prominence of the posters. In all cases the task was completed twice, and most respondents were asked to complete both tasks.

The images used were those from the POSTAR Pedestrian Visibility Study (see Method). As a result there were 42 “target” scenes (i.e. containing a poster), and 14 “decoy” scenes (i.e., without a poster); however, see the Results for a minor qualification of this. Hence the ratio of target to decoy scenes was 0.75. The viewing interval was set at 5 seconds (compared to 6 seconds in the eye-tracking study). To make the parallel more complete, the possibility of increasing the interval to 6 seconds was considered, but the data from the pilot strongly suggest that this would be unproductive. For the first 13 observers, there was a total of 1456 responses, of which only 52 were in excess of 5 seconds; so there would be little return in offering more time than this. The final results will be found to vindicate the decision to retain the 5-second viewing interval.

Materials

Table 1 summarizes the composition of the 40 photographic scenes containing poster panels that had originally been selected from a larger collection commissioned by POSTAR for the Pedestrian Visibility Study. There were in addition 16 decoy scenes with no poster.

Figure 1: Classification of scenes containing posters by panel size and poster location

Number of target scenes	Left of centre	Line of sight	Right of centre
6 sheet	8	8	8
48 sheet	8	-	8

Subjects

The subjects (respondents) were recruited from the Birkbeck College subject panel and with posters positioned around the vicinity of the college. Subjects were paid £7 for completing the experiment. All reported normal or corrected-to-normal vision.

Search Task 1: 40 subjects (15 male, 25 female), mean age 24.8).
Prominence Judgement Task: 30 subjects (9 male, 21 female), mean age 27.1.
Search Task 2: 20 subjects (6 male, 14 female), mean age 30.0.

Of the 40 who did Search Task 1, 10 were then tested on the Prominence Judgement task. The remaining 20 who did the Prominence Judgement task, were then tested on Search Task 2. There were thus 60 subjects in total.

Apparatus

All stimulus presentation and response recording was carried out using an IBM Pentium II PC running the E-Prime experiment generator software under the Windows 2000 operating system. Responses were made using the number pad of the PC keyboard and a “MicroTouch” touch sensitive screen. The keyboard was placed with the number pad approximately in line with screen centre although subjects were allowed to move this slightly if they felt uncomfortable. Picture stimuli were displayed on a 21 inch monitor (visible diagonal 49.5cm) with a screen resolution of 1024 x 768 pixels. Each image was full 1024 pixels wide (40cm) and 670 pixels high (26cm). Response latencies (reaction times) were recorded to millisecond accuracy by E-Prime with image onset synchronised with screen refresh. Refresh rate was 60Hz.

Procedure

The details of the three tasks investigated will be given in the order in which the findings are prioritised, and are reported in the Results section. The key difference between the search tasks was that for Search Task 1, the scene was erased immediately after the target-located response and the poster located was indicated on the blank monitor and therefore from immediate memory. For Search Task 2 the target-located response did not blank the screen, and the target and any other posters noticed were indicated with the scene in view.

1. Search Time Task 1

Individual trials consisted of the following components (All text was 18 point Courier font in white on a black background):

A. Fixation. A single "+" character displayed for 1000ms at screen centre.

B. Image Display. Images were displayed for a maximum of 5000ms or until a "yes" or "no" response was made. Responses were made using the "1" and "2" keys of the keyboard number-pad.

C. Touch display. The screen was cleared and the message "Touch Screen Now" displayed for 5000ms or until the screen was touched by the subject. A line of text reading "[No Poster]" was displayed at the bottom centre of the screen, to be pressed if the subjects had responded "no" **. X and Y coordinates (in pixels) of the response were recorded.

** This feature was introduced after the 9th subject when it became apparent that touching the screen after each trial allowed the subjects to develop a rhythm in their responses that appeared to aid responding on all trials.

D. Blank Screen. The screen remained blank for 2000ms before the fixation point was displayed and the process repeated.

Subjects completed 8 practice trials in the company of the experimenter who advised if the correct procedure was being followed and initiated another set of practice trials if deemed necessary. Two experimental blocks consisting of one display each of the 56 images in random order were completed by each subject.

2. Prominence Judgement Task

Subjects were given instructions 2 (below) and then shown a series of example images (not subsequently used in the main part of the study) further to explain their task.

A. Image display. Each image was displayed for 6000ms before the screen was cleared.

B. Response display. A standard Windows "Answer Box" was displayed with the heading of "Enter Prominence (0-100):" displayed about a text box into which the subject entered their numeric rating of prominence for the image just seen. After the number was entered the screen was cleared and remained blank for 2000ms before the next trial began.

Subjects completed 2 blocks of 56 trials (one trial for each image in an individually randomized order).

3. Search Time Task 2

Each trial began with the same fixation and image display screens as for Search Task 1, although in this case the screen was not cleared after a Yes/No response was made (B).

A. Fixation. A single "+" character displayed for 1000ms at screen centre.

B. Image Display. Images were displayed for a maximum of 5000ms. Responses "yes" and "no" were made using the "1" and "2" keys of the keyboard number-pad.

C. Poster-touching. With the image still visible subjects were requested to touch each of the posters on the screen starting with the one that they had seen first. On making their "Yes/No" response, a white box (measuring 7cm high and 3cm wide) was presented on top of the image at the bottom right-hand corner of the screen with the word "End" in 18 point text at its mid-point. Subjects were requested to touch the screen on the box when they had either touched all posters on the screen or immediately if they had responded "No" to a poster being present.

A maximum of four poster-touch responses were accepted before the screen was cleared and the message "That's 4 touches." was displayed for 2 seconds.

D. Next Trial Warning. The message "Get Ready…" was then displayed for 2 seconds.

E. Blank Screen. The screen remained blank for a further 2 seconds before the next trial.

As with the first search task each subject was given 8 practice trials under the supervision of the experimenter with the option of completing more if they had any problems. They then completed two blocks of trials.

RESULTS AND DISCUSSION

1. Search time task 1

The search time for each scene was noted, together with the response (yes vs no, i.e., the scene contains a poster vs. the scene does not contain a poster).

For each scene, and for each of the two presentation blocks, a distribution of search times was formed. These were based on the individual search times for each of the 40 observers.

For a given scene, cut-offs were imposed on each distribution at 0.75, 1.0, 1.25, 1.5 and 2.0 seconds. These cut-offs were decided by inspection, the goal being to avoid ceiling or floor effects, and to set a value that appeared to ensure good separation (discrimination) between the distributions for the 56 scenes.

The principal measure of performance on the search task for a particular scene was defined as the proportion of observers who responded accurately (i.e., they correctly classified the scene as containing a poster, or they correctly identified it as a scene not containing a poster) within a given deadline or cut-off interval. For a given scene, this proportion will be referred to as the search hit rate. Since there were five cut-offs, there were five of these hit rates per scene. These measures were obtained separately for the two presentation blocks. The hit rates for the 5 second maximum viewing interval were also calculated.

For a given target scene there was also a hit rate based on the eye-tracking performance of the observers who participated in the Pedestrian Visibility Study.

Errors (when the wrong or no response was given) are of interest but there appear to be no trends that seriously qualify the results reported below, and a detailed analysis of the errors has not been undertaken. It seems likely that they would serve as a source of noise in the data, possibly marginally lowering the strength of any relationships reported, and making conclusions marginally more conservative than they might be. Notwithstanding, further analysis of the detail will be needed to identify the source of difficulties for the observers.

1.1. Target hit rates

Correlation coefficients were calculated between the six search hit rates and eye-tracking hit rate (PEDHRATE), as well as a number of other measures from the Pedestrian Visibility Study (Table 1). Correlations are also reported for the median search times for Blocks 1 and 2.

Table 1: Selected correlation coefficients for search task measures and the hit rate from the Pedestrian Visibility Study

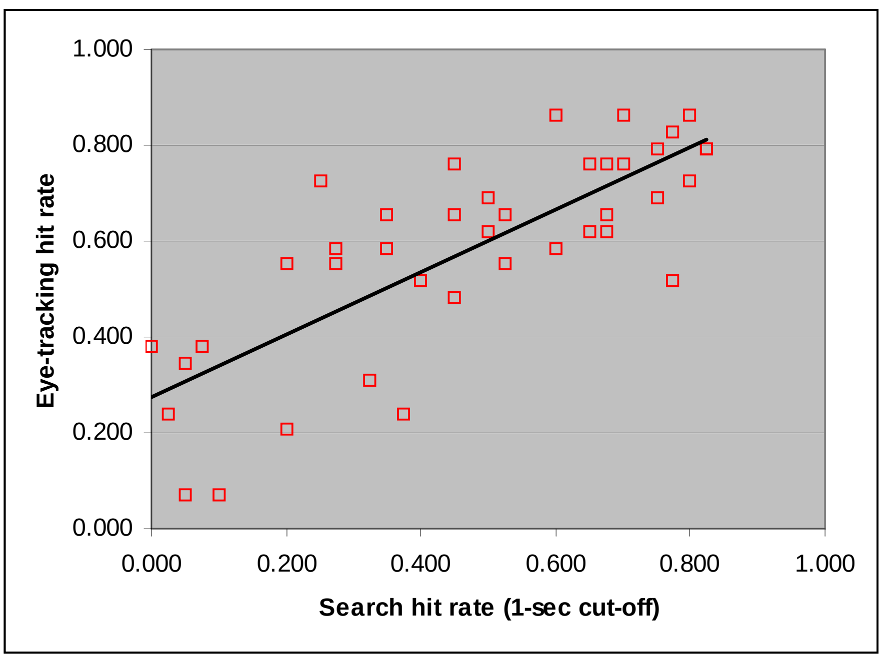
	AVTSTART	*NHITS*	*AVDURPS*	***PEDHRATE***
BLOCK1 N<0.75	-.630	.668	.733	**.716**
BLOCK1 N<1.00	-.698	.693	.713	**.780**
BLOCK1 N<1.25	-.713	.616	.649	**.741**
BLOCK1 N<1.50	-.744	.615	.638	**.720**
BLOCK1 N<2.00	-.739	.499	.538	**.631**
BLOCK1 N<5.00	-.683	.413	.443*	**.592**
BLOCK2 N<0.75	-.664	.651	.580	**.767**
BLOCK2 N<1.00	-.703	.598	.578	**.786**
BLOCK2 N<1.25	-.724	.568	.578	**.760**
BLOCK2 N<1.50	-.764	.551	.545	**.683**
BLOCK2 N<2.00	-.731	.489*	.509*	**.630**
BLOCK2 N<5.00	-.620	.499*	.485*	**.664**

MEDIAN RT BLOCK1	.740	-.610	-.641	**-.748**
MEDIAN RT BLOCK2	.752	-.605	-.554	**-.766**

* Correlation is significant at the 0.01 level (2-tailed).
All other correlations are significant at the 0.001 level (2-tailed)

The key entries in Table 1 are the six correlations shown in bold in the rightmost column. The highest correlations (0.780 and 0.786) with the pedestrian eye-tracking hit rate are for the search hit rates based on the 1-second cut-off for the first and second block, and the lowest correlation (0.592) is for the 5-second cut-off for the first block. It is of interest that the six correlation coefficients peak at the same point (the 1-second cut-off), and decrease between 1 and 2 seconds within blocks as the cut-off is delayed. Figure 1 is a scatterplot for the strongest of these relationships in Block 1 (when of course the respondent is least familiar with the material). The following account will therefore focus primarily on results for this cut-off. This will be reinforced by the subsequent finding that the parallel between the hit-rates estimated by eye-tracking and search methods is closest for this cut-off.

Figure 1: Scatterplot for search hit rates from Session 1 of the search task (1-second cut-off) and eye-tracking hit rate from the Pedestrian Visibility Study.



Linear regression equations (N=40 images) for the eye-tracking hit-rate (EMHR) as a function of the search time hit-rate at the 1-second cut-off (SHR) were:

Block 1: EMHR = 0.651 x SHR + 0.275

Block 2: EMHR = 0.635 x SHR + 0.213

(the respective standard errors of estimate for the slopes were 0.085 and 0.081 and for the intercepts 0.045 and 0.051). Notwithstanding the level of agreement, it is clear that there are a number of deviant observations, and closer analysis is needed to identify the source of these discrepancies.

Any change in the search results between the two blocks is not crucial at this juncture though it will bear on how the method might be applied, and the extent to which the two sets of results agree also serves as a useful index of their reliability (repeatability). In the event the respective hit-rates for Block 1 were 0.499 and 0.604, while the median search times were 1.032 and 0.925 seconds. These differences were highly significant (further details below) indicating faster location of the targets when the images were seen for a second time.

An indication of the impressive reliability of measurement for the key search measures is provided by correlations between Blocks 1 and 2 for the 40 poster images. For the search time hit-rate at the 1-second cut-off, the correlation was 0.932, and for the median search time it was 0.941.

Panel size

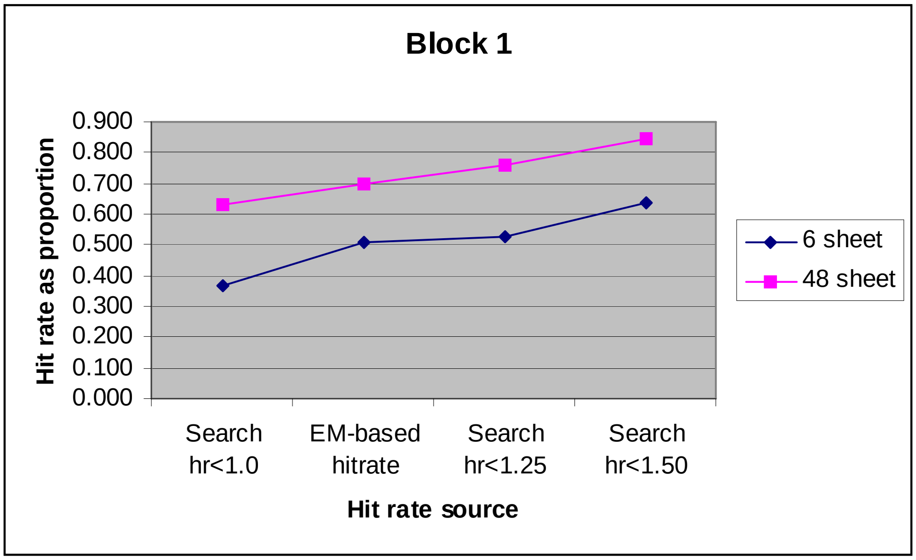
Table 2 shows results for the 6 sheet and 48 sheet scenes. The search task hit rates are shown for Blocks 1 and 2 for both panel sizes, together with the corresponding eye-tracking hit rates. Median search times are also shown for the two panel sizes for each block. Hit rates were higher for the 48 sheets in all cases, and the hit rate derived from the search task increased significantly between the two presentation blocks. It is clear that median search times were much lower for the 48 sheets, and decreased between first and second presentations. A split-plot analysis of variance, with Panel size as a between-items factor and Block as a within-items factor of the data for the 1-second cut-off, showed both the panel size and block effects to be highly significant (Panel size: $F[1,38]=10.84$, $p<0.001$, $MS_{error}=0.104$; Block: $F[1,38]=45.48$, $p<0.001$, $MS_{error}=0.0046$). The interaction of Block and Panel Size was not significant. This was mirrored by similar results for the median search time data, for which both panel size and block were significant (Panel size: $F[1,38]=8.928$, $p<0.001$, $MS_{error}=143498$; Block: $F[1,38]=41.02$, $p<0.001$, $MS_{error}=5344$).

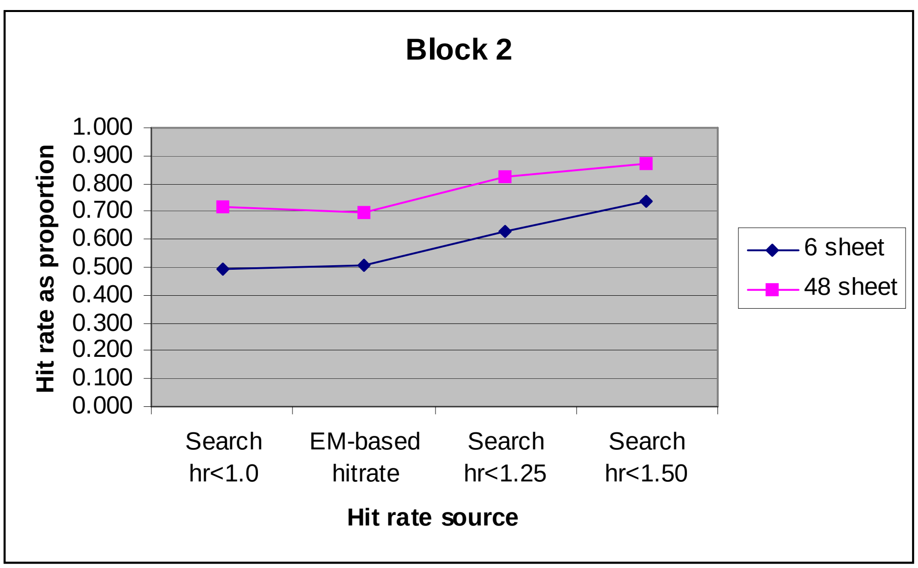
Table 2: Performance on the search task as a function of panel size: Hit rates and median search times (ms), with the eye-tracking hit rates for reference.

		6 SHEETS	48 SHEETS
BLOCK 1	N<1000	0.369	0.630
	N<1250	0.525	0.761
	N<1500	0.635	0.845
	Median RT	1175.8	887.4
BLOCK2	N<1000	0.492	0.716
	N<1250	0.627	0.825
	N<1500	0.740	0.873
	Median RT	1038.9	810.6
EYE-TRACKING HIT RATE		0.507	0.696

It is of particular interest that the two types of hit rate are not drastically misaligned. Figure 2 shows the eye-tracking hit rates (which also of course show the panel size difference) in company with the various distribution characteristics formed by the search hit rates for the first presentation block, with the eye-tracking data appropriately interposed between the data for the three cut-offs.

Figure 2: Hit rate as estimated from the search task and the eye-tracking data as a function of panel size





Panel location

Another comparison within the new data and relative to the old is provided by Table 3, which summarizes the hit rates and median search times for 6 sheets at different horizontal positions in the scenes (left of centre, straight ahead, or right of centre). It appears that hit rates are rather higher for panels that are directly in the initial line of sight, and search is faster in this case. Furthermore hit rates for left of centre panels tend consistently to be higher than for right of centre panels. The outcome for the hit rate data is underlined by the results for median search time (which of course produces an inverse version of the hit rate profiles). The interest in the present context is not so much in these details as in the level of agreement between the new and old methods, at least in terms of the various profiles of performance. Analyses of variance revealed significant effects of panel location (for hit rates at the 1-second cut=off F[2,21]=4.149, p<0.03, MS_{error}=0.103; for median search times F[2,21]=4.839, p<0.019, MS_{error}=150782)

Table 3: Performance on the search task for 6 sheets as a function of panel position: Hit rates and median search times (ms), with the eye-tracking hit rates for reference.

6 SHEETS		Left of centre	Line of sight	Right of centre
BLOCK 1	N<1000	0.378	0.506	0.222
	N<1250	0.547	0.672	0.356
	N<1500	0.672	0.747	0.488
	Median RT	1163.6	984.4	1379.4
BLOCK2	N<1000	0.500	0.672	0.303
	N<1250	0.672	0.775	0.434
	N<1500	0.747	0.844	0.628
	Median RT	1013.0	823.6	1280.1
EYE-TRACKING HIT RATE		0.556	0.621	0.345

Although it is somewhat premature to conduct close analyses of the results, the next Table is included for completeness, and as again suggestive of a possible trend. No scenes were used with 48 sheet panels positioned directly in the line of sight, but there were equal numbers to left and right of centre. The new and old data are summarized in Table 4 for these large format panels. The effect of panel location was not significant, but the left of centre panels again seem at a small advantage, consistent with the result for the eye-tracking data (a significant advantage to the left of centre panels).

Table 4: Performance on the search task for 48 sheets as a function of panel position: Hit rates and median search times (ms), with the eye-tracking hit rates for reference.

48 SHEETS		Left of centre	Right of centre
BLOCK 1	N<1000	0.681	0.578
	N<1250	0.813	0.709
	N<1500	0.888	0.803
	Median RT	852.6	922.3
BLOCK2	N<1000	0.756	0.675
	N<1250	0.850	0.800
	N<1500	0.891	0.856
	Median RT	761.5	859.8
EYE-TRACKING HIT RATE		0.785	0.608

Terminal accuracy

Accuracy (at the end of the 5-second viewing period) for poster scenes was 92.9% in Block 1, and 95.3% in Block 2. For 6 sheets, accuracy was 91.1% and 94%, and for 48 sheets 95.6% and 97.1% in the two respective presentations. Not surprisingly, it was easier to miss a 6 sheet than a 48 sheet. Error rates were clearly low, and as noted above, further analysis is needed to reveal their basis. There is one instance in the set of decoy scenes which illustrates this point.

1.2 Decoy scenes

The results for the decoy scenes are not of major interest of course but they do raise some important concerns. Table 5 shows the results for the three cut-offs for Blocks 1 and 2, and the median search times. It is evident from the latter that there was a considerable speeding of responses between the two presentations. This is reinforced by the clear increase in the relevant search hit rates. While this is consistent with a general speeding of performance with practice, and in particular to the increased familiarity of the scenes, there is one crucial feature of the results that merits attention. The median search times are both less than 2 seconds, which means that when there was no poster on view, the observers tended to respond much sooner than they might have done, given that 5 seconds was allowed before "time-out" was called. This is underlined by the distribution results, which showed that for Block 1 90% of correct responses had been made within 3 seconds, and for Block 2 the percentage had risen to 96.5. Very few respondents reserved their responses to the final two seconds of the interval allowed.

Table 5: Performance on the decoy scenes for the search task: Correct response rates and median search times (ms)

Decoy scenes	Block 1	Block 2
N<1000	0.116	0.236
N<1250	0.248	0.408
N<1500	0.372	0.563
Median RT (ms)	1628	1453

Finally, it transpired that among the decoys, one scene (Image B39) included what a sizeable majority of observers (33/40 and 34/40 in Blocks 1 and 2 respectively) judged to be a poster. This scene was therefore excluded from further analyses reported here.

Terminal accuracy

Excluding Image 39, the accuracy of decoy responses in Blocks 1 and 2 was 92.8% and 95.0% respectively. It seems that the instructions to observers were effective in specifying the requirements of the task, and the nature of the targets (i.e., poster panels) of the search task. This is reassuring to the extent that observers were not seriously misled into classifying other signage as posters; this was a concern in advance because several of the scenes contained other advertising poster-like material. It remains to see what the pointing task reveals on this score.

2. Prominence task

2.1 Target scenes

Correlation coefficients were calculated between prominence and selected measures from the pilot study as well as the criterion measure, the eye-tracking hit rate from the Pedestrian Visibility Study. An indication of the reliability of the prominence measure is provided by correlation between Blocks 1 and 2, which was 0.867 (N=40, p<0.001).

Table 6: Correlations between prominence and search measures (all are significant at p<0.001 level)

		Search time measures			
		Block 1		Block 2	
		Hitrate(<1.00)	Median RT	Hitrate(<1.00)	Median RT
Prominence	Block 1	0.747	-0.715	0.742	-0.750
	Block 2	0.807	-0.752	0.804	-0.788
	Mean	0.799	-0.756	0.795	-0.792

Table 6 presents correlations between prominence (for poster scenes) and the hitrate and median search times for Blocks 1 and 2 of Search Task 1. It should be emphasised that the prominence measure was estimated from 20 respondents and the search measures from a different set of 40 respondents. The high correlations reinforce the strong impression of a high degree of agreement between the varied set of measures in the present study.

The correlation between prominence as measured in Block 1 and Block 2 and the eye-tracking measure was 0.689 and 0.652 respectively, and 0.699 between prominence averaged over blocks and the eye-tracking hit rate.

Linear regression equations (N=40 images) for the eye-tracking hit-rate (EMHR) as a function of prominence (PROM) were:

Block 1: EMHR = 0.0106 x PROM – 0.199

Block 2: EMHR = 0.0131 x PROM - 0.377

(the respective standard errors of estimate for the slopes were 0.002 and 0.002 and for the intercepts 0.136 and 0.183).

For comparative purposes, other analyses follow a similar pattern to that for the search time task.

Panel size

Prominence values were higher overall for the 48 sheets than for 6 sheets (84.5 vs. 66.0), but this difference decreased marginally between Block 1 (86.0 vs. 65.3) and Block 2 (82.9 vs. 66.7). A split-plot analysis of variance, with Panel size as a between-items factor and Block as a within-items factor of the prominence data, showed the panel size effect to be highly significant ($F[1,38]=56.92$, $p<0.001$, $MS_{error}=114.9$). The interaction of Block and Panel Size bordered on significance ($F[1,38]=4.016$, $p<0.052$, $MS_{error}=22.97$).

Panel location

Analyses of variance of the prominence data for panel location (and block) were computed for the two panel sizes. These analyses of variance revealed a significant effect of panel location for 6 sheets ($F[2,21]=4.265$, $p<0.028$, $MS_{error}=128.4$) with posters in the line of sight (71.9) receiving higher prominence scores than those left of centre(66.0), which were in turn rated as more prominent than those right of centre (60.2). The 48 sheet panels on the left of centre (85.4) were rated as slightly more prominent than those on the right of centre (83.5), but this small difference was not statistically significant.

2.2 Decoy scenes

There is little practical interest in the data for these scenes, but they have a useful role in confirming the validity of the task. Image 39 was excluded from the analysis since many respondents identified an object as a poster (producing average prominence scores of 46.3 and 56.4 in Blocks 1 and 2 for that image).

After excluding image 39, mean prominence scores for the decoy scenes were 14.96 and 18.30 for Blocks1 and 2. These scores are well towards the bottom of the scale, and roughly speaking mean “Poster probably/definitely not present”.

Table 7 shows confidence limits for the mean prominence ratings for decoy scenes along with those for 6 and 48 sheet poster scenes. Clearly in both blocks 1 and 2 there was no overlap whatever between the three scene categories, and in particular the decoys scored well below the targets. The larger format posters were assigned mean ratings between the scale responses “Poster clear but not quite obvious” and “Poster obvious and without effort”, while the smaller format received ratings from just above “Poster eventually clear but not immediately” and somewhat above “Poster seen but only with real effort”. The relative grading of these values is what matters, and the actual labels are not to be taken too seriously.

Table 7: 95% confidence limits for prominence judgements for poster and decoy scenes for Blocks 1 and 2

	48 sheet	6 sheet	Decoy
Block 1	88.67-83.27	70.77-59.87	16.32-9.42
Block 2	86.02-80.20	69.71-56.84	18.83-6.55

3. Search time task 2

Full results for this task have not been provided, since although there is broad overall agreement, this is in the face of a concern about the method that will be discussed below. Since the concern arose after the completion of data collection and was prompted by analyses of the results, sufficient information has been provided to allow an evaluation of the method and the concern expressed below. The results are adequately represented by the next table (that parallels the treatment of the results for Search Task 1), and by an overview of other findings.

The hit-rates and median RT measures all correlated significantly with the pedestrian eye-tracking hit-rate but at lower levels. The correlations (0.780 and 0.786) for the search hit rates based on the 1-second cut-off for the first and second block were 0.335 and 0.469 respectively, and for median RTs (0.338 and –0.396). The correlation coefficients peaked at the same point as before (the 1-second cut-off) for Block 2, but rather later for Block 1 (0.353 at the 1.5 second cut-off). The interpolation of the eye-tracking hit rates amid the search hit rates, in the manner of Figure 2 would position the eye-tracking measure around the 1.25 second cut-off, rather later than the previous case. The cumulative distribution of search times begins at a lower level than before, but finally (as will be seen from the terminal accuracy scores) accelerates to a higher level than before. This and other evidence suggests that the two tasks impose demands that elicit different processing strategies, and this will need careful consideration in any future investigation and applications.
The reliability of measurement for this task was also lower than for Search Task 1; correlations between Blocks 1 and 2 for the 40 poster images for the search time hit-rate at the 1-second cut-off, the correlation was 0.785, and for the median search time it was 0.848.

Notwithstanding these weaker relationships (which were all statistically reliable), the profile of the results – in terms of the effects of panel size and location – was preserved. Table 9 illustrates the point for panel size. It can be seen that hit rates for 48 sheets were substantially higher than for 6 sheets at the three cut-offs shown, and median RTs were correspondingly lower. What is not apparent from the Table but is clear by comparing it with Table 2 is that the Search Task 2 produced lower hit rates (and of course slower responding) than Search Task 1. Terminal accuracy for poster scenes (the hit rate after 5 seconds viewing) in Block 1 was 96.7% and 99.1% for 6 and 48 sheets respectively, and for Block 2 it was 95.6% and 99.7%. These values compare favourably with those achieved in Search Task 1.

Table 9: Performance on search task 2 as a function of panel size: Hit rates and median search times (ms) with the eye-tracking hit rates for reference.

		6 SHEETS	48 SHEETS
BLOCK 1	N<1000	0.331	0.516
	N<1250	0.496	0.688
	N<1500	0.623	0.797
	Median RT	1252.3	1006.5
BLOCK2	N<1000	0.431	0.575
	N<1250	0.594	0.778
	N<1500	0.731	0.894
	Median RT	1099.5	927.2
EYE-TRACKING HIT RATE		0.507	0.696

The final table in this section may provide some further insight into why Search Task 2 may have been anomalous. The table summarizes correlation coefficients between the various measures of visibility derived from the two search tasks and the prominence judgement task. In evaluating these correlations, it should be noted that: (a) Search Task 1 was done by a different set of subjects than Search Task 2 and the Prominence task; (b) Search Task 2 and the Prominence task were done by the same subjects; (c) Prominence was always done before Task 2. The pattern of correlations may of course be the way the cookie crumbles (i.e., random), but it is striking (from the top half of the table) that Search Task 1 correlates at a somewhat higher level with the Prominence task than with Search Task 2 (with which one might expect the greater cognitive overlap). Less surprising, but suggestive of considerable commonality, are the sizeable correlations between Search Task 2 and the Prominence task (completed by the same subjects).

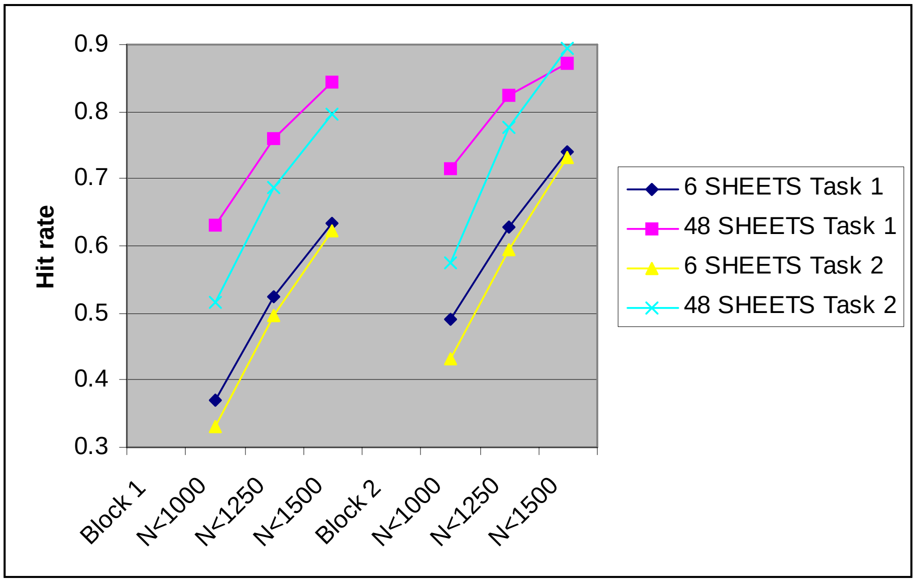
Table 10: Inter-task correlation coefficients for search hit rates (1.0 second cut-off) and prominence judgements (N=40 images)

		Task2 Block1	Task 2 Block 2		Prominence Block 1	Prominence Block 2
Task 1 Block 1		0.406**	0.505**		0.511**	0.582***
Task 1 Block 2		0.355*	0.433**		0.486**	0.484**
Task 2 Block 1					0.728***	0.721***
Task 2 Block 2					0.674***	0.776***

***significant with p<0.001
**significant with P<0.01
*significant with p<0.05

The following chart shows some hit rate results for Search Tasks 1 and 2, from the two presentation blocks for each task. The three cut-offs around the key 1-second point are shown for 6 sheet and 48 sheet panels (the upper pair of lines is for the 48 sheet panels in each case). It is clear that the relative and understandable advantage of 48 sheets over 6 sheets was sustained for both tasks and both blocks. Crucially it can be seen that, despite the presumed benefit of having seen the images twice before (during the Prominence task), the hit rates for Search Task 2 were below the parallel conditions for Search Task 1 for the first two cut-offs shown. However, the two sets tended to converge as the cut-off interval lengthened, the 48 sheet lines even managing to cross over at the 1.5-second cut-off.

Figure 3: Hit rates for cut-offs between 1.0 and 1.5 seconds as a function of Panel size for Search Tasks 1 and 2 compared for Blocks 1 and 2



Hit rate
0.9
0.8
0.7
0.6
0.5
0.4
0.3
Block 1
N<1000
N<1250
N<1500
Block 2
N<1000
N<1250
N<1500
6 SHEETS Task 1
48 SHEETS Task 1
6 SHEETS Task 2
48 SHEETS Task 2

OPERATIONAL CONSIDERATIONS

Both variations of the search procedure proved to be viable not just in terms of data collection but usability (by the subjects). It is useful to have confirmation, by the subject having to nominate the poster that evoked his or her response. This serves to reduce the possibility that the response was a mistake, and it indicates which object evoked the response. When there is more than one panel in a scene, the pointing response reduces any ambiguity about which was the one the subject noticed first. Similarly it is also clear when the subject mistakenly points to an object that is not a panel. Sometimes a subject may revert to guessing that a poster is present, possibly because something about the scene seems to indicate this likelihood; sometimes the response is over-prepared and is triggered incorrectly. All of these possibilities can be illuminated by the use of the pointing procedure. Refinements of the data collection software could be made to provide an on-line indication of whether an error was made; this was not available for the present studies but could be considered for implementation later. This would require the target region to be specified (e.g., by the screen coordinates of the corners of the target panel(s)). Meanwhile software was written to enable the acquisition of these coordinates, and the information is available for any further analysis of the pilot results.

The prominence judgement task also proved to be no obstacle for users, but the method of entering the responses (via the PC’s numeric pad) is improvable. The scale used posed no problems, though again there may better ways of labelling the scale, and of instructing the subjects. It seems worth retaining as a convenient method of tapping a viewer’s judgement of the contents of a scene in a relatively unpressured and reflective fashion. The core task also offers opportunities to assess different instructional emphases (e.g., “what catches your eye, judging the scene as a shopper/traveller/etc”), which may merit further investigation. For the present, the data from a modest sample of subjects were impressively stable and in concordance with the search and eye-tracking results.

CONCLUDING DISCUSSION

The results from the search and prominence judgement tasks when set alongside those from the Pedestrian Visibility Study are remarkably stable. There is a reassuring consistency between the two tasks, which were quite diverse in their demands on the subjects, and which both contrasted in this sense with the original eye-tracking task. The measures derived from the tasks themselves are also very different, yet they show a strong degree of empirical agreement. This agreement is demonstrated by their inter-correlations, and by the fact that they are similarly sensitive to key scene variables such as poster size and location. The consistency between the results from the two presentation blocks which participants underwent for the search and prominence judgement tasks suggests that practice effects may not affect the profile of the results even if it is likely to show absolute changes in the level of performance on the speeded tasks.

Because the test of practice effects was within a session (on the same day), it would be advisable to confirm this finding about practice by testing some respondents on successive occasions on separate days (since experience tends to show that performance changes between days or weeks are more substantial than changes within days). There would be useful practical benefits if longer-term practice effects were to be shown to be unimportant, since it would then be possible to countenance maintaining a regular respondent panel to assess poster scenes rather recruiting new respondents whenever new poster scenes or environments needed to be assessed. Having a pool of experienced representative respondents to call on would undoubtedly reduce the amount of "noise" in the data[1], and the overheads of recruitment and training would be reduced. The essential requirement for data collection via a respondent panel would be that the data continue to be as diagnostic relative to panel and environment conditions as any "gold standard" technique such as eye-tracking; the present evidence is certainly supportive in this respect.

Although there was a remarkably high level of agreement of the results between the original eye-tracking study, the Search task 1 and the Prominence judgement task, some caution is required in the light of some aspects of the findings from Search task 2. This task was always performed after two blocks of the Prominence task; this was not ideal in terms of experimental design but the decision to explore Search Task 2 was almost post hoc, the task being constructed after a viewing of the other two tasks at an early stage of the study for Search Task 1. In any event the fact that the Prominence task had preceded it is a possible reason Search Task 2 should have produced data that were somewhat out of register with the rest. Search task 2 provided the third and fourth occasions on which the subjects saw the 56 scenes, and they would no doubt have been familiar by this stage with the setting, the scenes and their contents, and the general focus of the research. Figure 10 shows that the general profile of the hit rates (as a function of cut-off and panel size) was actually quite similar for both blocks of the two search tasks, yet the level of performance was initially noticeably lower for Search Task 2 but rapidly rising to the level for Search Task 1. So the experience of having made perceptual judgements of the same stimulus material that was to be scrutinized in the search task does not seem to have facilitated the latter. There is also the possibility that the subjects who did the first search task were abler than those who did the second; although there is no way of being certain about this, it seems unlikely given the rather different performance profile for the second search task, and it is here that a viable explanation can begin.

Quite subtle differences between the two search tasks seem likely to have been responsible for the differing results. In the first search task the subjects needed only to locate a single poster (even if it was from memory). To deal with the demands of the second task, which

[1] It is worth noting that, although the report is not burdened with the detail thereof, standard deviations of the search time measures decreased markedly between Blocks 1 and 2.

specified as a possible second stage a prolonged pointing task (specifying up to four posters), the subjects may have adopted a strategy that suited the extra activity, resulting in prolonged initial inspection as they continued to scan for further posters. If this inspection was overlapped with the decision to respond, search would at least initially have been delayed, and perhaps initiated at some intermediate point when a complement of posters had been located for the second stage of the task, reflected by the eventual convergence of the hit rates on those of the first search task. Moreover the intended priority to be attached to the target-located response might well have been undermined by the fact that the response did not erase the display, which would otherwise have effectively punctuated the search activity. It is possible that further detailed analysis, of single vs. multiple poster scenes would throw light on this hypothesis.

It should be remembered that in the final analysis, the results of Search Task 2 are merely a watered down version of the results of the other two tasks, and they do not contradict the overall picture or the detailed trends established by data from those two tasks. This minor perturbation of the overall account has necessarily taken space, possibly more than it will eventually be shown to warrant.

The fact is that the pilot study has produced results that enable a very positive set of conclusions about the search and prominence tasks to be made relative to the criteria for assessing methods for the Postar inclusivity study in the context of visibility measurement (see Appendix 1). With regard to the two principal methods used in this investigation, the foregoing supports the following conclusions relative to the criteria (desirable values for each as shown in parentheses):

- **Validity re poster viewing (HIGH)**
  HIGH - convergent evidence from the eye-tracking study

- **Labour-intensiveness (LOW)**
  MEDIUM - Data collection easy, but data processing laborious for the time being (until automated)

- **Delivery (FAST)**
  FAST – potentially, once portable

- **Equipment availability/ease of procurement (HIGH)**
  HIGH – laptop/touch-screen/data capture software

- **Equipment cost (LOW)**
  LOW – laptop, but cost higher if touch-screen used

- **Data yield (HIGH)**
  MEDIUM – less than eye-tracking

- **Acceptable discrimination between poster types and environments (HIGH)**
  HIGH – as discriminative as eye-tracking data; see evidence above

- **Generality re environments (HIGH)**
  NOT FULLY TESTED – high according to evidence so far, but more environments to be studied

- **Suitability for visibility modelling requirements (HIGH)**

HIGH – inasmuch as it provides hit-rate and related measures

- **Convergence with the relevant POSTAR visibility model (HIGH)**
  HIGH – given the demonstrated agreement between hit rates from the eye-tracking and search tasks

REFERENCES

Antes, J.R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, **103**, 62-70.

BARBER, P.J. and DICKENSON, A. (1997). Factors affecting visibility distances for poster panels. Technical Report: POSTAR Ltd.

Buswell, G.T. (1935). *How people look at pictures*. Chicago: University of Chicago Press.

Henderson, J.M., Weeks, P.A. and Hollingworth, A. (1999). Eye movements during scene viewing: Effects of semantic consistency. *Journal of Experimental Psychology: Human Perception and Performance*. **00**, 000-000.

Loftus, G.R. and Mackworth, N.H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*. **4**, 565-572.

Mackworth, N.H., and Morandi, A.J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, **2**, 547-552.

Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.

APPENDIX 1: TOWARDS A GENERIC ARCHITECTURE FOR VISIBILITY

Paul Barber
School of Psychology
Birkbeck College

A number of possible methods are under consideration for the purpose of measuring visibility across outdoor environments.

They vary in terms of several criteria/factors including the following (desirable values for each as shown in parentheses):

- **Validity re poster viewing (HIGH)**
- **Labour-intensiveness (LOW)**
- **Delivery (FAST)**
- **Equipment availability/ease of procurement (HIGH)**
- **Equipment cost (LOW)**
- **Data yield (HIGH)**
- **Acceptable discrimination between poster types and environments (HIGH)**
- **Generality re environments (HIGH)**
- **Suitability for visibility modelling requirements (HIGH)**
- **Convergence with the relevant POSTAR visibility model (HIGH)**

APPROACH 1

- *Eye tracking (as in POSTAR driver visibility study)*
- *Use still images*
- *Range of environments (roadside, tube, bus, taxi, station, mall, etc)*
- *Examine effects of displacement from scene centre, poster size, site locale*
- *Examine effects of poster content if desirable to estimate the separate contribution of content to visibility*
- *Can be developed with enhanced technology to include video imaging of scenes.*

Eye-tracking with still images of relevant scenes from the target environments, i.e., same methodology as the original POSTAR visibility study, but with new images representing the new environments (Tube, taxi, etc).

The same measures would be available directly from the eye tracking data, but hit rates (i.e., a measure of "eyes on panel" summed over the full extent of the observer's passage towards the poster) are estimated from the visibility model, using the eye tracking data as a time-sliced sample along that trajectory. If the trajectory is known, modelling is relatively straightforward. If it is unknown, or if there are many alternative trajectories, modelling will be much more complicated, and may need to be radically different. Some special cases may be simpler, for example, when the observer is stationary in the proximity of a poster, in which case the summation would be over time with distance fixed (viewing angle may vary, as the eye/head turn, adding a new dimension of variability to take into account).

NOTES

1) As yet have there is no POSTAR pedestrian visibility model, though there are data for the purpose. Convergence as above would be with either the existing driver (as far as it has been developed) or the (impending) pedestrian model.

2) The pedestrian visibility study used "walkthroughs" as well as random sequences from the relevant environments.
3) The scenes for pedestrian visibility were classified as in the driver study (i.e, using POSTAR's classification into arterial - residential – shopping) but they also represented "commercial" environments (e.g., railway stations).

APPROACH 2

- *Record observer gaze direction by human observer*
- *Surveillance/security camera or embedded camera on poster*
- *Accuracy not known*
- *Ethical aspects*

The method here replaces the eye-tracking equipment by a human observer. Assessing the direction of a person's gaze is one of the perceptual skills supporting social interaction. It is clear that under certain circumstances most of us can do this without effort. Approach 2 would capitalize on this ability, by asking a panel of judges (more than one to check for reliability/agreement) to assess what pedestrians are looking at when in the vicinity of a poster, and specifically whether or not, and possibly for how long, they look at the poster.

Gaze direction could be assessed using video recordings of passers-by at poster sites, with cameras positioned appropriately. A trial of the method could potentially be done using existing security videos if they are of adequate quality. A possible variant of the method would be to mount one or more such cameras on a panel and to ask the judge(s) to assess when passers-by look at the panel/camera. Obviously the camera would need to be hidden. Because people's behaviour is being recorded without their awareness, ethical questions are raised, and would need to be resolved.

The accuracy of gaze direction judgements has been studied for direct interpersonal interactions and there is some data on its accuracy in video communication (e.g., video-conferencing). Whether judges are accurate when judging direction of gaze at target objects as would be required for this application does not seem to have been researched.

APPROACH 3

- *Head-mounted eye camera worn as observer moves through real environment*
- *Considerable yield of complex data, high processing load*
- *Long-term solution*

A head-mounted eye-camera could be acquired or possibly hired, to be used in conjunction with a scene camera to monitor what observers look at as they move /are moved about relevant environments. The plan would be to record from each of a number of observers (one at a time) wearing the kit in a variety of environments for a set duration. They could be completely free to select their own route between a starting and finishing location (passing poster sites as haphazardly as that would imply), or they could be asked to follow a prescribed route (being exposed to a series of selected poster sites). Equipment for a realistic visibility study would need to be comfortable, portable, easy to calibrate, and capable of delivering accurate identification of gaze targets in a wide range of poster environments, and to function in a wide range of illumination conditions. It is very likely that technical support would need to be provided during the recording stage, so the wearer would have a technically qualified companion at all times, together with whatever computer kit was necessary for the transfer of data from the eye tracker. It would also be necessary for the equipment to be unobtrusive – clearly it would be counter-productive to mark out the wearer as a visitor from outer space.

Technological developments mean that eye tracking facilities are more accessible and they are becoming increasingly familiar and may soon even attain the final accolade of being ubiquitous. This does not mean that the yield of such technologies matches up with user requirements. They produce a vast amount of data which has to be sifted, and while there have been highly beneficial software improvements, when the target of an observer's gaze is as fuzzily defined as it is with a scene camera, and in environments are as heterogeneous as those of the poster industry, the problem of the labour-intensive nature of the data-processing end of the system will remain for some time to come. A long-term solution for the poster industry might entail mounting signalling devices on posters, which would provide information about the topography of the panel to be integrated with information from the observer's eyes about his/her gaze location. Intelligent software capable of recognising posters from the scene camera data would also enhance the applicability of head-mounted eye tracking equipment. Such solutions seem very much long-term approaches, which would require substantial amounts of funding. They are mentioned here for completeness, and to illustrate their present limitations. The strength of such approaches lies in their potentially high validity relative to poster visibility measurement.

APPROACH 4

- *Virtual reality*
- *Ever-present possibility but currently yield more virtual than real*
- *Technical problems associated with scenario construction and simulating different motion characteristics (driver, pedestrian)*
- *After-effects of exposure may be counter-indicative*

Virtual reality is another approach that comes readily to mind as a means of presenting poster images to viewers, with concurrent eye tracking to monitor what they look at. The method has its advocates, and it is widely used in training high-level skills. There are reservations about the method too, notably concerning the after-effects of exposure to a virtual environment, which may be sufficiently long-lasting to warrant caution as to the representativeness of an observer's visual experience. How easy this would be to adapt to the different poster audiences (including drivers controlling a car and pedestrians) is not known. The quality of visual presentations engineered by virtual methodologies is not always convincing. A possible problem for the short-term future for an application in which, for example, a cine-technique is married to a simulation platform, is the limitation on how many scenarios could be produced to enable a viable yield of visibility data. This is quite apart from the problem of how to equate the manner in which the cine version of the scenario was recorded for different types of observer (e.g., driver vs. pedestrian). The problem of representativeness of the real world experience besets all techniques (even the most direct and immediate of them, i.e., Approach 3 as explained above).

APPROACH 5

- *Poster awareness task to tap eyes on panel aspect*
- *Simulation of non-voluntary nature of poster viewing by assigning poster detection low priority status*
- *Use simple speeded poster-unrelated task as primary task*
- *Distraction awareness task as secondary task*

- *Adjust viewing times to reflect viewer approach speed (e.g., driver, pedestrian, tube traveller)*
- *Obtain hit rate type of measure from poster task and adjust via eye tracking data (initially using POSTAR driver data and scenes for the adjustment factor)*
- *Derive aggregate visibility scores by modelling the accumulation of momentary hit rates increments along the viewer's passage (or set of possible passages) towards the poster (as in the POSTAR visibility model)*

This will probably be the most controversial approach but it has significant advantages and should not be rejected out of hand. It may be appropriate to trial it to see what it delivers. A pilot study for this method would the easiest and quickest to deliver of all the advance studies that would be necessary to decide among the five "approaches".

The method hinges on an assumption that probably lacks face validity for at least some of those presently signed up for the current project. The method involves the observer reporting in some way the presence or absence of a poster panel in a natural scene, presented on a computer screen, the speed and accuracy of responding being the indices of the attention value of the poster site. It should be noted that there is no direct evidence that the method does not discriminate between poster locations or panel sizes. Indeed it would be surprising if it did not. Clearly it may well not discriminate as desired between different poster environments or different viewer roles (driver/passenger, pedestrian, tube or bus traveller). However, the latter problem may be dealt with in the implementation of the method (in effect by imposing different exposure durations to reflect the different OTS durations) or at the modelling stage (by aggregating over different total approach times, i.e., OTS durations).

A number of variations of the method have been developed, each seeming to increase the face validity of the method. For the moment a broad outline of the raw method is described. Possibly this is foolhardy because it stands the least chance of being accepted. However, it is the overall rationale that needs to be evaluated, and rejection of the method should not be based on the voluntary nature of the behaviour which is tapped.

The observer will view a series of images, a proportion of which will contain a poster at one of a set of prescribed locations. Panel size and site environment will be varied. Prior to the onset of the display, the observer will fixate a symbol at a marked location at the agreed centre of the display (possibly the road centre, the scene's vanishing point, or some agreed normative central location). A scene (from the poster or control/baseline condition) will follow. The pre-fixation symbol will be replaced by a small letter, which the observer must classify (e.g. vowel or consonant). This is the task that is assigned top priority in terms of speed and accuracy. The ancillary task, the accuracy of which will be noted, is to report whether or not the scene contains one of a number of possible distractions (e.g., person approaching on a collision course, animal, road sign, poster, etc). Accuracy of spotting posters will be used to index poster visibility. The second task is to be treated as having low priority for the observer, and efforts will be made to camouflage the intent of the study. The display will be for a duration which as indicated above may vary (an alternative is to allow free viewing but to impose a deadline for responding).

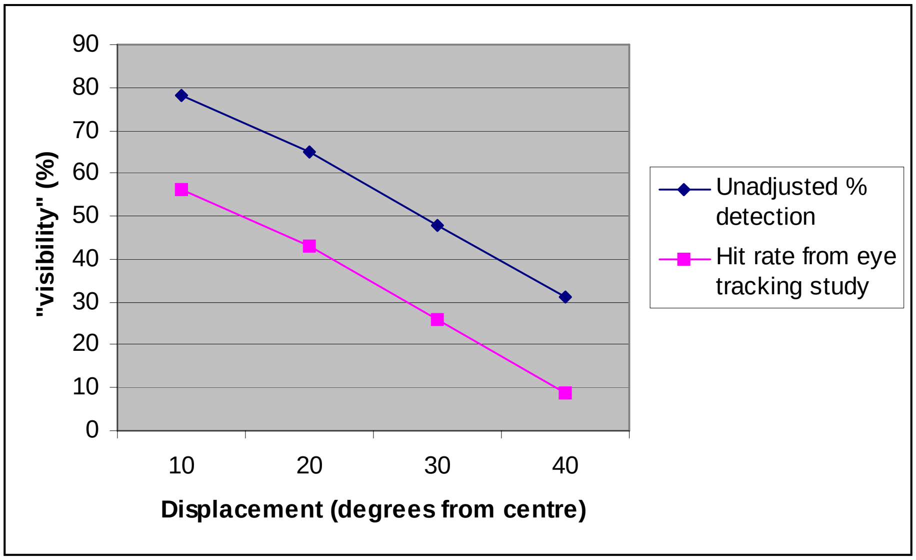
The task would be undemanding for the observers (a good deal less than eye tracking) and the data acquisition would be quick and summary data would be swiftly available. Getting hold of images, or commissioning their acquisition would possibly be the most time-consuming stage of the study. Preparation of software to run the study would be straightforward. A pilot study could be run with students or volunteers supplied by participating companies. A research assistant would be needed for a couple of months.

- The set of scenes would include scenes from the POSTAR driver visibility study, for which eye tracking data on hit rates exist. The correlation between these baseline data and the accuracy data for the same scenes for the poster awareness task would be computed. It is crucial that there is an acceptable level of agreement between them. Whatever level is obtained for the hit rate measures directly derived from the poster awareness task, a regression adjustment would be possible based on the eye tracking data, to bring the poster awareness data into register with eye tracking hit rates. The same adjustment would be used to produce hit rate data for the different environments.
- The visibility modelling exercise would be completed by using the adjusted hit rate data for the various tasks, aggregating over the agreed OTS durations, which would in turn be estimated from the times taken on average for observers to complete their approaches to the target poster in a scene.
- A representative set of scenes for a given environment/medium would be required and a pooled visibility index for a medium could be obtained by averaging over all scenes for that environment.

APPROACH 6

This is a more direct and simpler version of Approach 5. There is one task for the observer, namely to detect whether or not the scene contains a poster, and to respond accordingly as fast as possible (but taking care also to respond accurately). Half (say) of the scenes contain a single poster (size and location are carefully controlled). A set amount of time is allowed for the task before time-out is called, and the observer's detection time is noted. An unadjusted visibility index would be obtained for each poster type/environment. Suppose, for the sake of illustration, that the total exposure duration is 5 seconds. Given an interval as long as this, nearly 100% of the observers will be capable of correctly detecting the poster in a given scene. Clearly the number of observers who have detected the poster in that scene within the first quarter-second of the interval will be considerably less than 100%. This percentage will clearly increase as the length of this "deadline" interval increases. A cut-off can be imposed at any convenient point between 0 and 5 seconds, and its unadjusted visibility score can be derived as the detection percentage for that cut-off. A poster panel that is very prominent for whatever reason (e.g., low surrounding clutter) will get a high score on this measure. Indeed the method will order panels by their visibility (and observers will very likely need to fixate the panel to respond, i.e., the eyes on panel requirement will be met). But, it will be objected, this requires deliberate searching for a poster.

To deal with this important and potentially fatal objection, consider the following graph. It shows fictitious data from a detection experiment and an eye tracking study using identical images.



It is unlikely that such perfect agreement would be achieved empirically, however, it can be seen that if the two methods converge satisfactorily on the same functional relationships with all relevant poster panel factors (like size, displacement and environment), then a simple regression-based adjustment can bring the two into agreement. The viability of the approach could be assessed relatively easily using a sample of the existing images from the POSTAR visibility research. Assessing the agreement across the POSTAR environments in particular would be the major objective of an initial study. There seems no obvious *a priori* reason why there should not be adequate agreement across the poster factors of interest to POSTAR and its associates.

It should be noted that the eventual common currency could be set via agreement about which deadline to use for the detection task. New environments can be incorporated relatively straightforwardly using the detection task with an appropriate battery of additional images. This aspect of the method would need to be developed with particular care so as to ensure its robustness relative to existing environments, and so the test images would probably need to include a common set of baseline scenes to calibrate the data for a new environment.

COMMENTS ON METHOD

What might be difficult to engineer/agree would be the contents of the poster sites, particularly for those involving a live journey. A particularly ineffective poster on a highly prominent panel would distort the result for that site. This was one of the reasons why the POSTAR visibility studies have relied on a multiple instances approach, i.e., it was decided necessary to have several representatives of each panel type. That is, for example, a given observer was exposed to multiple instances of a 48 sheet at a specified angular displacement from the edge of the road. The same was done across a range of displacements, and this was repeated for each of the standard poster environments (Shopping, Residential and Arterial). This was done for 6, 48 and 96 sheet panels. This achieved the goal of aggregating the experiences of observers over a range of executions, and provided a wide geographic/ecological sampling of poster sites.

A solution that has from time to time been mooted is to design material specially for the purpose of assessing panel visibility, which would be neutral (presumably this would mean something like “average”) in content and appearance. Even if there were such a thing as an agreed neutral poster to be displayed at every site to provide a common visibility baseline, this would be bizarre from the point of view of industry practice.

MODELLING FOR GENERIC VISIBILITY

The POSTAR visibility model was restricted to drivers and passengers, but is due to be extended to include pedestrians. Visibility is aggregated over the entire approach distance for an observer moving towards and finally passing a roadside poster at a constant speed (taken as the average vehicular speed in the relevant environment). A simple generalization of this approach is possible for any approach pathway that is more or less linear, and that can be assumed to be traversed at a fixed speed. What changes is the interval over which momentary contributions to visibility are aggregated. Approach pathways are not usually known, they may be haphazard, or there may be several alternative paths. This aspect of visibility modelling is the most challenging. It will be necessary to agree acceptable assumptions about average OTS durations for different environments and poster architectures, and to agree a way of representing approach pathways.

pjb: 30-10-2002

APPENDIX 2

INSTRUCTIONS

All Tasks:

THANK YOU VERY MUCH FOR AGREEING TO TAKE PART IN THE EXPERIMENT. REMEMBER , YOU CAN DECIDE TO WITHDRAW AT ANY POINT SO THERE IS NO PRESSURE ON YOU.

THIS IS WHAT WE WANT YOU TO DO. PLEASE READ THIS CAREFULLY AND ASK ANY QUESTIONS BEFORE YOU BEGIN.

Search Time Task 1

We are doing a series of experiments on the visibility of posters in various conditions. The task today is to search as rapidly as you can to spot a poster displayed on the screen. You will see a variety of photographic scenes of outdoor environments, like roads, stations, pedestrian areas and so forth. One or more advertising posters may appear in each scene. But while there may be more than one, there may also be none at all. You can think of your task then as searching for a poster, deciding whether or not there is one, and making a quick accurate response each time.

If you decide that a poster is present then press the "YES" key (number 1 on the number-pad). If you decide that there is no poster present press the "NO" key (number 2 on the number-pad).

Press the key as quickly and as accurately as you can. Then touch the screen where the poster appeared. If you decided that you did not see a poster in the picture then touch the screen at the bottom where it will say "[No Poster]". If you happen to see more than one then touch the screen at the position of the first poster you saw.

Speed is important. However, don't worry if you make a mistake, just try to be as quick and as accurate as you can manage.

Search Time Task 2

We are doing a series of experiments on the visibility of posters in various conditions. The task today is to search as rapidly as you can to spot a poster displayed on the screen. You will see a variety of photographic scenes of outdoor environments, like roads, stations, pedestrian areas and so forth. One or more advertising posters may appear in each scene. But while there may be more than one, there may also be none at all. You can think of your task then as searching for a poster, deciding whether or not there is one, and making a quick accurate response each time.

If you decide that a poster is present then press the “YES” key (number 1 on the number-pad). If you decide that there is no poster present press the “NO” key (number 2 on the number-pad).

Press the key as quickly and as accurately as you can.

Then touch the screen on the first poster you saw and then touch any others in the picture as quickly as you can (up to a possible total of 4). When you have touched all the posters on the screen then touch the box that appears in the bottom right hand side of the screen that says “End”. If you decide that you did not see a poster in the picture then touch the “End” box straight away.

Speed is important. However, don’t worry if you make a mistake, just try to be as quick and as accurate as you can manage.

<u>Prominence Judgement Task</u>

The task is to judge the ease with which you can spot a poster displayed on the screen. You will see a variety of photographic scenes of outdoor environments, like roads, stations, pedestrian areas and so forth. One or more advertising posters may appear in each scene. There may be more than one, but there may be none at all. You can think of your task then as to signalling the prominence of the poster, how well it stands out in the scene.

After each display I would like you to response by giving a number between 0 and 100. If the poster is completely obvious, and it takes no effort at all to spot, then you should assign a very high number. A large poster in the middle of an otherwise blank screen would get 100. So any poster that pops out at you with little or no effort on your part would get a high number, and if it is a bit less prominent, then a lower number would be needed. The more you have to do to find and make it out, the lower the number you should give. If on balance you judge that there was no poster but are not completely sure then use a low number, but if you are quite sure that there was no poster, then use an even lower number. To anchor that end of the scale, you could consider that a blank screen with no poster would get 0. Try to be consistent with your use of numbers but don't worry about it. This scale might help you decide.

100	95	80	65	50	35	20	5	0
Poster on blank screen	Poster obvious and without effort	Poster clear but not quite obvious	Poster eventually clear but not immediately	Poster seen but only with real effort	Poster probably present but unsure	Poster probably not present but unsure	Poster definitely not present	Blank screen