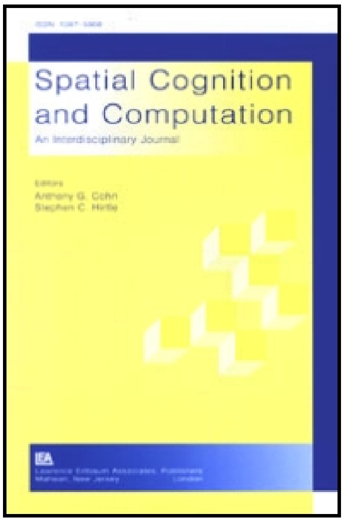

Spatial Cognition & Computation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653698

Spaces or Scenes: Map-based Orientation in Urban Environments

Clare Davies[a]; David Peebles[b]
[a] Research C530, Ordnance Survey, Southampton, UK [b] Department of Behavioural and Social Sciences, University of Huddersfield, Queensgate, Huddersfield, UK

Online publication date: 03 June 2010

To cite this Article Davies, Clare and Peebles, David(2010) 'Spaces or Scenes: Map-based Orientation in Urban Environments', Spatial Cognition & Computation, 10: 2, 135 — 156
To link to this Article: DOI: 10.1080/13875861003759289
URL: http://dx.doi.org/10.1080/13875861003759289

Spatial Cognition & Computation, 10:135–156, 2010

ISSN: 1387-5868 print/1542-7633 online
DOI: 10.1080/13875861003759289

Spaces or Scenes: Map-based Orientation in Urban Environments

Clare Davies[1] and David Peebles[2]
[1]Research C530, Ordnance Survey, Southampton, UK
[2]Department of Behavioural and Social Sciences, University of Huddersfield, Queensgate, Huddersfield, UK

Abstract: Two experiments examined people's strategies when orienting with a map in outdoor scenes within unfamiliar urban environments. We investigated how the 3D visual scene and the 2D layout geometry influenced people's choices of features when matching the scene and the map, and studied the problems they encountered when doing so. Results support previous evidence that in geographically realistic contexts, visible salient landmarks bias people away from using optimal geometry-matching strategies. This implies that prediction of orientation difficulty merely from analysing the spatial layout (e.g., with space syntax isovist measures) may be highly problematic. Implications for future map design are discussed.

Keywords: orientation, map reading, isovists, landmarks, visual salience, cartography, urban, wayfinding

1. INTRODUCTION

As anyone knows who has exited an unfamiliar subway station or studied a you-are-here map, local spatial awareness involves more than merely knowing where you are. Your orientation—which way you are facing—is also crucial. But how do you align the map with the scene to establish which direction is which? Which cues do you draw on? What cognitive processes do you follow?

It may well be that solving this problem in a complex urban scenario is quite different from orientation experiments in simpler and smaller laboratory environments. Cues are richer and more varied. The whole scene may need to be aligned to the map, rather than merely spotting a single target. The purpose

Correspondence concerning this article should be addressed to Clare Davies, Research C530, Ordnance Survey, Romsey Road, Southampton SO16 4GU, UK. E-mail: ccdavies@primex.co.uk

of the orientation is more ecologically valid: usually as part of ongoing personal navigation, but not always—sometimes one is matching the map to photographs or video footage, e.g., in historical research or police crime detection. Therefore this paper considers the map-matching orientation task in the context of real urban environments, and with a view to matching the scene taken as a whole to the map rather than a specific target object within it. We feel that this is a more common and critical task in ground-based urban scenarios.

Why is this valuable? As well as helping us to understand and potentially model the cognitive processes in their own right, we may start to understand how far the processes depend upon the 2D geometry (i.e., the local geographic layout) represented on the map, and how far they may be instead influenced by the visual 3D scene. If the former was important to the task then we may be able to use simple and even semi-automated cartographic interventions to help map users. These could involve spatially analysing the mapped layout (using either GIS or *space syntax*-type techniques: see e.g., Wiener et al., 2007), and adding more landmarks to the map in places where orientation should be particularly hard.

However, if the task strategies and performance depend more heavily on the visual 3D scene, this would imply that adding orienting landmarks to maps would need to involve selecting not just any landmark, but those which were most visually salient *within the scene* (Winter, 2003). The two experiments we report here tested whether features in the 3D visual scene do tend to draw many people's attention, and whether as a consequence people often ignore useful details of the more informative 2D spatial layout, to the detriment of successful orientation.

2. BACKGROUND

Previous research has looked at orientation in real-world environments. Some studies (e.g., Griffin & Lock, 1979; Eley, 1988; Pick et al., 1995) asked participants to match topographic maps to rural landscapes, examining the role of terrain shape. Aviation-inspired studies have taken this further and studied orientation from a partly overhead perspective (e.g., Aretz, 1989; Harwood, 1989; Wickens & Prevett, 1995; Gunzelmann et al., 2004). One line of research (Warren et al., 1990) asked people to judge their orientation relative to a single building from the outside, but a more common theme (Presson, 1982; Blades & Spencer, 1990; Hagen & Giorgi, 1993; Meilinger et al., 2007a) is to ask people to orientate inside a building or even a single room. While the distances, viewing angles, degrees of realism and available visual cues all differ greatly among these environments, nevertheless they could all be seen as types of 'vista' space, defined by Montello (1993) as a scale of space that can be completely viewed but not touched without locomotion.

These studies have usually not tried to assess the relative contribution of the 2D and 3D aspects of the environment to people's task performance. Nevertheless, some clues as to the relevance of these aspects have emerged. Across different environments there seems to be consistent evidence of some role for prominent *landmarks* (both single features and groupings of features) in people's choice of strategy for solving the task, rather than abstracting the geometry of the scene layout. If a unique landmark exists both in the scene and the map, then matching it between the two can provide an orienting shortcut that saves the observer from having to abstract, rotate and match less salient geometric layout shapes or features—it is rather like having a 'north' arrow painted on the ground in the scene. This would suggest, but does not necessarily imply, that the 3D visual scene would influence the choice of salient landmark, rather than merely configurational considerations.

Experiment 1 was designed to examine this possibility for typical (British) urban scenes, with participants freely choosing a strategy to match a photograph of the scene to a specific direction on a corresponding circular map. The main aim of the experiment was to examine people's errors qualitatively as well as quantitatively, to identify the apparent cues and strategies they adopted based on the types of errors made in different types of scenes.

3. EXPERIMENT 1: ERROR TYPES WHEN MATCHING PHOTOGRAPHS

3.1. Method

Design and Participants. Thirty-five students and members of staff (aged 20–60 years, 6 males) from the University of Huddersfield took part in the experiment. All participants saw the entire set of 40 stimuli in random order (following five randomly ordered practice stimuli).

3.2. Materials

The stimuli were 40 (+5 practice) scenes and corresponding maps from various locations in the city centre and inner-city residential areas of the English south coastal city of Southampton (over 300 km from Huddersfield). The maps were derived (removing labels and some other minor features) from circular sections of Ordnance Survey's OS MasterMap® Topography Layer large-scale mapping, which is accurate to 1:1250 scale in urban areas; each map showed a 200 m radius around the photographer's location. The stimuli were displayed (see Figure 1) and responses recorded using a specially programmed tcl application, running on Windows PCs with 17-inch (~42 cm) displays.

Figure 1. Example map and scene, Experiment 1.

A dot in the centre of the map indicated the location of the photographer. When the mouse cursor was over the map, a short black line of fixed length appeared, pointing from the centre of the dot toward the tip of the cursor. This rotated around the dot as the mouse cursor was moved over the map, so that it always pointed towards it. To indicate the scene's orientation, participants had to click the mouse after moving it to a position where the black line pointed to what they believed to be the centre of the pictured scene.

Scenes and maps were selected to represent a wide range of building shapes and types, degrees of salience and distinctiveness (judged informally by eye according to the degree of variation in heights and building styles—from uniform Victorian terraces to scenes with a prominent tower or church spire), together with a range of urban features such as open green spaces and differently proportioned streets. The stimuli were also controlled for alignment so that the correct response ranged across the full 0–360 degree circle, and there were approximately equal numbers of broadly north- and south-facing scenes. Unlike most orientation experiments, however, there was no attempt to limit the number of possible angles from 0 to fixed 30- or 45-degree increments; instead the bearing could be any number of degrees from north (as is more typical in real life, particularly in nongrid cities).

3.3. Procedure

Participants were introduced to the experiment through the following scenario: "Imagine that you are standing in the street in an unfamiliar town, holding a map. You know where on the map you are standing, but you need to find out which way you are facing." They were then shown an example scene/map pair (on paper) and told that their task was to work out in which direction they must be facing on the map in order to see the scene. Participants were instructed how to make a response, asked to respond as quickly and as accurately as possible, and told that the maps were all the same scale. They were also warned against the natural assumption that the 'upwards' direction on the map necessarily indicated 'forward' in the environment (since piloting had suggested this clarification was helpful).

There were five practice trials and forty experiment trials in total; each scene was shown once in randomised order (reordered for each participant). The angle (bearing) of the response was recorded, from 0 degrees (pointing directly north, i.e., to the top of the map) to ±180 degrees (pointing directly south, to the bottom). Response times were also recorded in seconds (i.e., time from first appearance of stimulus to recording of the mouse click).

As an additional measure to try to account for any individual differences, after participants had finished the experiment they were asked to complete the Santa Barbara Sense of Direction Scale (SBSOD; Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002). This measure requires participants to rate their level of agreement on a 7-point scale to 15 statements concerning their spatial abilities and preferences. Previous studies have shown the SBSOD to predict objective measures of environmental spatial abilities, including orientation, quite highly (Hegarty et al., 2002). We also added one further question, asking participants to rate on a ten-point scale their experience of actually matching maps to real-world scenes.

4. RESULTS

Responses were scored as correct if the angle of the response line fell within 15 degrees of the true angle in either direction (i.e., a 30-degree arc), i.e., within half of the ~60-degree arc subtended by the scene itself. This was done because a simple "number of degrees from target" accuracy measure would not meaningfully describe people's errors in the urban scenarios used in this experiment. For example, consider standing at a crossroads looking down one street. A 90-, 180- or 270-degree error are all equally wrong, but all much more logical and likely than say a 120- or 300-degree one.

Thus, in the absence of biasing cues the expected response distribution, rather than one normal curve peaking at zero degrees, would instead be four separate smaller curves. This reflects the qualitative difference between minor inaccuracies (given that participants had to remember to match up the scene's estimated centre line, not the street's distant vanishing point) and choosing a completely wrong direction (street, path or building). Only the latter type of error was of interest to us (since it would lead to a wrong decision in a real-life scenario, whereas slight inaccuracy would not), so we took a categorical approach to the accuracy data.

Participants' performance on the task varied, with mean accuracy across participants being 53.5% with a minimum of 27.5% and maximum of 82.5%. Similarly, task completion times ranged from a minimum of 9.87s to a maximum of 43.22s, mean 25.98s. Performance also varied markedly between scenes, with nine scene/graph pairs eliciting an accuracy of less than 30% but 10 pairs being accurately processed over 70% of the time.

For the SBSOD questionnaire the mean score was 4.12 (SD 1.15) with a minimum of 1.87 and a maximum of 6.07. Score on the SBSOD correlated significantly with accuracy in the task ($r_s = .471$, $p < .005$) but not with response time ($r_s = -.064$, $p = .71$). Participants' scores on the additional experience question also correlated significantly with task accuracy ($r_s = .559$, $p < .001$) and with scores on the 15 SBSOD questions ($r_s = .538$, $p = .001$), but not with response time ($r_s = .135$, $p = .44$). This demonstrates the similarity of our task to that of real-world orientation and navigation requirements, while also showing the difficulty of teasing out aptitude from experience in causing individual performance differences.

For this experiment the main focus of our analysis was an attempt to identify the primary factors that appeared to create common patterns of error. It will be noted from the examples in Figures 2–4 that, rather than a single normal (Gaussian) distribution, incorrect responses typically clustered around specific candidate feature types that were in some way similar to the target location. For instance, if the scene included looking down one of the streets that connected at an intersection, incorrect responses were clustered around the angle of each of the alternative streets leading from it. However, there were always enough cues in every scene to allow participants to choose between them if the whole scene was considered carefully; this was evidenced by the average correct performance across scenes being 53.2% (well above chance).

We analysed these errors by identifying clusters of responses for each scene (i.e., regions of the map containing more than one response) and counting the number of responses in each cluster. Then, looking across the range of scenes, the experimenters analysed features of the map associated

Figure 2. Scene and map 31 from Experiment 1 illustrating correct response (north-east line) and incorrect responses (other lines) caused by scene object salience.

Figure 3. Scene and map 38 from Experiment 1 illustrating correct response (south line) and incorrect responses (other lines) caused by missed ground-level cues.

with the responses and attempted to identify possible causes of the error. A set of nine error types were thus identified that captured all of the data (e.g., a relevant cue in the scene obscured by an unmapped object, a relevant ground-level cue has been missed, etc.).

The two experimenters then independently coded all of the erroneous responses according to these patterns (error types), recording which cues appeared to be ambiguous (i.e., non-discriminating) between the chosen and the correct directions, and which cues were unambiguously different and hence must have been ignored. Multiple causes for each error were recorded if necessary. The codings were then filtered to retain only those for which level of agreement reached 65% or above. The four most frequent patterns where agreement was 65% or above are described next.

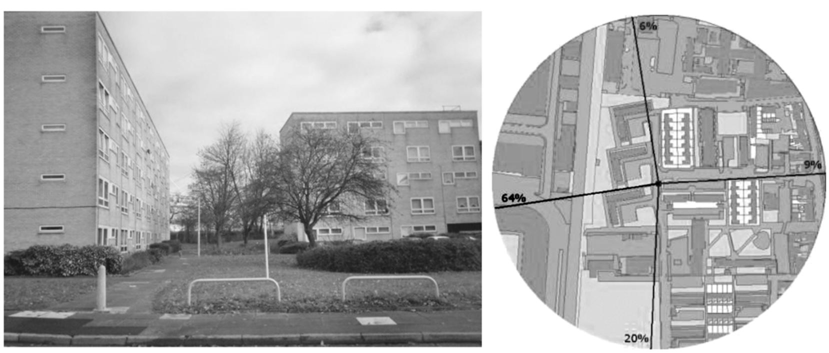

Figure 4. Scene and map 7 from Experiment 1 illustrating correct response (west line) and incorrect responses (other lines) caused by misperceived object distance.

1. Scene Object Salience (SOS)

This occurred when there was a very visually salient object in the scene (e.g., a tall or distinctive building) that was not unambiguous on the map. This was implicated, on average, in 33.2% of erroneous responses in the 10 scenes where it was seen to occur (interrater agreement: 97%).

An example of this factor is shown in Figure 2. This suggests that the five main erroneous clusters (red lines), totalling 36% of the responses, resulted from participants noting the salient blue building in the scene, but failing to identify it correctly on the map. The erroneous responses suggest that alternative buildings were identified instead, and the response was then aligned to this assumption.

2. Missed Ground-level Cue (MGC)

This occurred when there was an unambiguous cue at ground level (e.g., a traffic calming obstacle, distinctively shaped lawn or path) that was apparently ignored. This may be due to strategic factors whereby the participant chose to focus on buildings or other features, ignoring the greater distinctiveness of the ground-level layout. This was implicated, on average, in 25.1% of erroneous responses in 32 scenes (inter-rater agreement: 80%). An example of this is shown in Figure 3. Here the three main erroneous responses (totalling 40% of responses) identified a road on the map but did not note the pattern of pathways on either side of the road (nor the left-turning road immediately in front of the viewing point).

3. Misperceived Object Distance (MOD)

This occurred when the incorrect response direction would be more accurate if a key object in the scene was either nearer or further away than it actually was. This was implicated, on average, in 22.4% of erroneous responses in 22 scenes (inter-rater agreement: 71%). Examples of this factor (totalling 29% of the responses for this scene) are shown by the east- and south-pointing lines in Figure 4. These suggest that people identified the view as lying between two buildings, but then misjudged the distance of at least one of the buildings from the viewpoint. (Arguably, two of these errors are also an example of error type 2 here, illustrating that the error types were not mutually exclusive.)

4. Left/Right Reversal (LRR)

This occurred when the incorrect response would be more accurate if the scene was left/right reversed to its mirror image. This was implicated, on average, in 17.2% of erroneous responses in 26 scenes (interrater agreement:

Figure 5. Scene and map 1 from Experiment 1 illustrating correct response (north-west line) and incorrect response (south-east line) caused by left/right reversal.

65%). An example of this factor is also shown by the north-pointing line in Figure 5 (6% of the responses for this scene). The reader will see that the arrangement of objects in this direction is roughly the mirror image of the actual scene configuration: the viewer is actually closer to the left side of the road in the scene, and distant buildings are also more salient on the left, but if the reverse was true then the alternative response would make more sense.

5. DISCUSSION

These error patterns, although not accounting for all of the errors in many of the scenes, nonetheless point towards a set of different cognitive issues. The tendency to pick on a distinctive (but not easily matchable) 3D landmark is particularly interesting, as it implies a strong role for the visual salience of the scene in people's orientation strategies in this task. In this sense it may depart from the body of work suggesting a stronger role for geometry than for landmarks in orientation within lab experiments (e.g., Hermer & Spelke, 1994), although there are many potential reasons for this given the very different task scenario.

The second error pattern, missed ground-level cues, may possibly result from participants' inexperience of maps at the scale we used. The most widely used maps are typically at least 1:10,000 scale. At this scale the details of urban street furniture, sidewalks, grass verges and paths through parks will usually have been removed or simplified to minimise clutter. The larger-scale mapping that we used in our experiment is typically only frequently used by professionals (e.g., urban planners), and no such users were included in this sample. Many participants therefore may not have expected to be able to use ground-level cues as part of their map-matching strategy, particularly where the task scenario was deliberately reminiscent of real-world orientation.

Nevertheless, ground-level layout is more likely to be correctly matched (to a map that does show it) than is a distinctive 3D landmark that may have a far less distinctive 2D ground plan. In any case, ground-layout matching is sometimes important in real-world orientation even with more generalised maps, particularly in more rural areas and urban open spaces, and any scenes where other cues are too uniform to be of use (e.g., a set of identical nondescript and unlabelled buildings). Furthermore, the overall ground-level geometry is, as discussed further below, the only aspect of the scene which is safely predictable from viewing or analysing a map, without requiring knowledge of what the scene looks like in reality. Error distributions showed that participants had apparently already constrained their responses to fit the cruder 2D scene structure, in terms of the general street directions, and they apparently did not *always* ignore the more detailed 2D cues. Therefore it seemed worthwhile to look further at the strength of their relative effect in a more controlled scenario in Experiment 2.

The variation in performance among scenes is worth noting. Obviously with an allowed error margin of ± 15 degrees, i.e., a 30-degree arc, a chance performance level should be 360/30 or 8%. However, since we deliberately used a range of urban scene types, some scenes were much more constrained than others in the number of realistic response choices (e.g., a crossroads has only 4, effectively raising chance levels to 25% if we assume that all participants could take the basic layout into account; open space has many more so expected chance levels are still close to 8%). Sure enough, inspecting the data showed that the scenes with the lowest performance levels were indeed those in open spaces such as parks and squares.

Similarly, although all scenes contained sufficient cues for the task to be solvable if enough detail was taken into account, obviously they varied in the amount of this that had to be taken into account to successfully eliminate incorrect choices. However, again we have no easy way of measuring this variation. Later, in the General Discussion, we will say more about two potential types of measure—spatial 'isovist' measures and visual salience models—which may have theoretical potential as predictors of performance on this task.

Obviously two potential components of the map-matching task (depending on participants' choice of strategy) are mental rotation and the transformation between the 2D and 3D geometry. The third and fourth error patterns above suggest that these may be a challenge for a minority of participants. However, it may not be that they must be trying and failing to make these geometric transformations successfully. Instead, they could be avoiding doing them altogether by simply trying to match single objects or groups of objects between the scene and the map, and ignoring the fact that their chosen solution violates the geometric configuration in terms of either relative depth or left-right asymmetry. Experiment 2 allowed us to more closely assess these potential differences in problem-solving strategy.

6. EXPERIMENT 2: 2D SPACE VERSUS 3D VISUAL SALIENCE

For realism, Experiment 1 deliberately used unedited outdoor photographs in order to closely replicate a real-world orientation scenario. These photographs often include many obtrusive unmapped objects, such as parked cars and trees. In addition, as in everyday life, the symbolic nature of the map contrasts with the realistic appearance of the scene, making it harder to match objects between the two stimuli. Although this closely replicated typical real-world scenarios, it also potentially left people's responses at the mercy of uncontrolled distractors in the scenes, such as visually salient colours or unusual objects, and allowed some matchable objects to be hidden from view. This possibly biased the results for some scenes in unpredictable ways.

Therefore, Experiment 2 replicated the study under more controlled conditions where irrelevant scene features and random visual salience issues could be eliminated, in order to assess the extent to which the differences among scenes might actually be due to their variable *spatial* layouts—and how much was still down to aspects of the 3D visual scene. To achieve this we simplified the scene images by using ground-level snapshot images of a 3D model of Southampton rather than real photographs. This had the additional advantage of moving away from potentially limiting real-world-based assumptions by participants about what might be feasible strategies for the orienting task.

In Experiment 1, the most commonly observed type of error appeared to place an emphasis more on the commonly salient visual features of each scene than on its relationship to the map. Controlling the scenes more rigorously would test if these errors disappear or reduce when the scenes vary less randomly in visual salience, and to observe the effect of explictly manipulating the presence or absence of obvious 3D and ground-level 2D cues. Therefore we explicitly manipulated the presence or absence of a strongly salient 3D cue and/or distinctive 2D ground layout cue, using a selection of the scenes previously used in Experiment 1 (with a fresh set of participants).

Since our overall goal was to understand participants' problem-solving strategies, in Experiment 2 we also collected verbal protocol data from a few additional participants to help us get a qualitative understanding of how people approached the task.

7. METHOD

7.1. Design and Participants

Forty-nine students and members of staff (aged 30–60, 19 males) from the University of Huddersfield took part in the experiment. As before, all participants saw the entire set of stimuli in random order. An additional five participants carried out the experiment while having their verbal protocols recorded to enable qualitative assessment of their apparent strategies in solving the task. Participants were encouraged to perform the task as quickly and accurately as possible.

7.2. Materials and Procedure

The experiment design was exactly as in Experiment 1 (except that the SBSOD was omitted), using the same 5 practice scenes and 20 of the same main stimulus scenes, but in all trials the previous photograph was replaced with a simplified scene image. These were generated using a buildings-only 3D model (a research prototype developed at Ordnance Survey), overlaid on the same mapping (OS MasterMap Topography Layer) as used for the maps, and draped on an OS Land-Form PROFILE® terrain model to provide a realistic and accurate representation of the 3D landscape and its buildings. Thus each scene image matched the map in its content, 2D layout and approximate color scheme, but still exactly matched the (this time unused) photograph in its basic 3D geometry (minus extraneous unmapped detail such as trees, cars, windows, street signs and fences). An example of a scene is shown in Figure 6; this scene corresponds to the photo shown in Figure 4.

As before, the map contained no name labels or other indicators to differentiate among buildings or streets. Therefore the only remaining relevant cues for items within the scene were size and shape (in terms of ground area layout) and also colour (since, to avoid confusion and distraction, a similar colour scheme was used for both the scene and map). In these scenes, therefore, choosing a single individual item (e.g., a building) and attempting

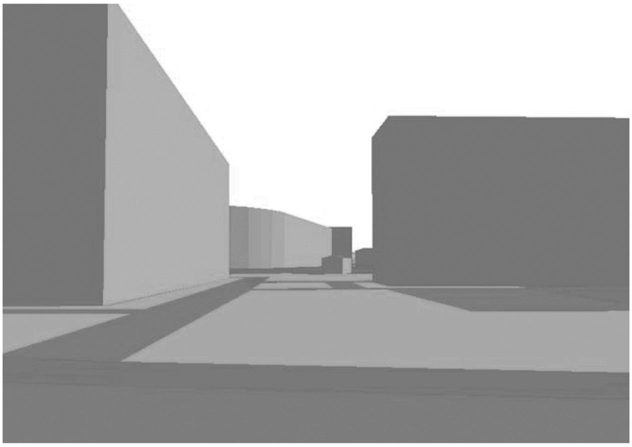

Figure 6. Scene 7 from Experiment 2, corresponding to photo 7 in Experiment 1 (shown in Figure 4). The map in Figure 4 was used in both experiments.

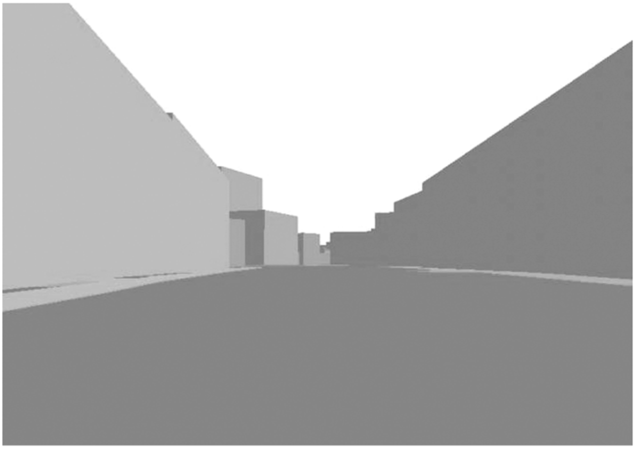

Figure 7. Scene 1 from Experiment 2 (matching the photograph and map in Figure 5).

to match it to the map was more obviously unlikely to be successful. The appearance of one item would usually not be unambiguous, but the overall ground layout and relative positions of items would be.

The 20 scenes used in this experiment were selected to fulfil the same criteria as previously in terms of range of scene type, alignment etc. In addition, they were coded for presence/absence of an obvious 3D landmark such as a tall or unusually shaped building, and for presence/absence of an obvious foreground 2D layout cue such as an asymmetric lawn, criss-crossing paths in a park, traffic-calming chicane etc. 5 scenes included both types of cue; 5 had only 3D; 4 had only 2D; 6 had neither. Figure 6 shows a scene with both types; Figure 7 shows one with neither.

8. RESULTS

As before, participants were usually able to perform the task reasonably accurately, with the proportion of correct responses for each stimulus ranging from .07 to .72 ($M = .56$, $SD = .17$). The mean response time for the 20 experiment trials was 40.94 seconds ($SD = 9.76$).

Clusters of participant errors for each scene were rated in the same way by both experimenters according to the same criteria as in Experiment 1; in each scene, the same general error clusters were seen as previously.

8.1. 2D versus 3D Cue Effects

A two-way factorial ANOVA examined the effects of 2D/3D presence/absence on error rates and on response times, across the 20 scenes and 49 participants.

For errors, 2D ground cue presence had a small but significant effect of decreasing error rates ($F(1, 48) = 5.47$, $p < .05$), but 3D cue presence had a much stronger effect that *increased* them ($F(1, 48) = 40.35$, $p < .0001$). There was also an interaction effect ($F(1, 48) = 5.26$, $p < .05$): the slight performance improvement with a 2D cue only occurred in the absence of a 3D one, which apparently so completely distracted people from using the more effective 2D strategy that the 2D cue presence made no difference.

Perhaps more surprisingly, the response-time analysis suggested that performance actually slowed down when either cue was present, whether 2D ($F(1, 48) = 9.28$, $p < .005$) or again more strongly for 3D ($F(1, 48) = 29.7$, $p < .0001$), and most strongly of all for both (interaction $F(1, 48) = 4.37$, $p < .05$). The slowing effect of the presence of a 2D cue was small except where a 3D cue was also present.

In other words, in the absence of any strongly salient 3D landmarks, the presence of a strong 2D ground layout cue helps at least some participants to improve accuracy, albeit at the expense of a little speed. Adding a 3D cue as well may confuse people as to which cue to use, slowing them down more than when either cue is present alone, yet they typically opt for using the 3D cue in these scenes and thus make as many errors as when the 3D cue appears on its own.

The main effects of these analyses were also checked using nonparametric Wilcoxon signed-rank tests, since both response times and error rates showed minor deviations from normal distributions. The tests showed the same patterns of relative F-ratio size and significance.

8.2. Verbal Protocols

As the verbal protocol data were taken for only five participants and their utterances were relatively few in number, we were unable to subject them to rigorous quantitative analysis but rather used them to gain some additional qualitative insight into the strategies they used. The verbal protocols showed similar ambivalence about 2D versus 3D cues. One participant said very little during the experiment despite continual prompting. Of the other four, all reported depending on the buildings in the scenes for their overall strategy across the experiment; only two mentioned additional cues which for one were "road and color" and for the other "grass." Examining the four participants' comments scene-by-scene, although there were only 11 explicit mentions of building height or 3D shape, buildings were mentioned roughly twice as often (56 times) as either roads (27), grass/green areas (25) or other 2D cues (26). Since these participants were able, unlike the main group, to take their time and think about suitable strategies, the continued bias towards buildings even where stronger ground-level cues were present is quite striking.

Although the participants often appeared to realise that only the buildings' 2D shapes would be of use in matching to the map, their eye movements still often appeared to dwell on an obviously salient (e.g., extra tall) 3D cue where present, even if they did not mention it in their comments. This could help to explain the lengthened response times among the main participant group when 3D cues were present: they may have a distractor effect even when rejected as a solution strategy.

9. DISCUSSION

In this experiment, where the scenes looked far less realistic, the same types of errors were nevertheless observed. Further, although it was much more obvious with these stimuli that the 2D ground layout cues were more valuable than building heights or roof shapes, the presence of such distinctive 3D cues both slowed and worsened performance by encouraging an unfeasible strategy, with people apparently disregarding the 2D matching strategy at such times even though (even slower) response times suggested that they were aware of the potential of both. If 3D cues are to have such a strong effect even where far more effective strategies are made so much more obvious, it is difficult not to conclude that in many situations the 3D visual salience of the scene in front of the observer will be more influential on their successful orientation than the surrounding 2D geographic layout.

9.1. Cross-Experiment Analysis

Analysing across both experiments (for the 20 scenes that were used in both) allows us to verify and estimate the sizes of the main effects across a larger sample of people, and also to investigate potential differences in people's performance between real-world photographs and the simplified computer-generated model images.[1] We might hypothesise that the presence of real-world depth cues in the Experiment 1 photographs, rather than the smooth surfaces of the computer model, might aid performance. This might be particularly so when assessing the 2D ground layout, as opposed to relying on single-landmark matching—although better depth information could also help with the latter. Alternatively we might hypothesise that real-world distractors, such as unmapped objects and irrelevant detail, and the potential for occlusion of key scene cues by such items, could worsen performance compared to the 'ideal' scene represented in the computer images of Experiment 2.

Accordingly, analyses of variance (ANOVA) were performed on the combined data for those 20 scenes, with the same two repeated-measures factors as above (presence/absence of salient 3D landmark cue; presence/absence of distinctive 2D ground-layout cue) and an additional between-subjects factor of image type (photograph versus computer model).

For response times, image type had a moderate and significant effect ($F(1, 328) = 25.75$, $p < 0.0001$, $\eta^2 = 0.07$), but had no significant interactions with the 2D and 3D effects (reported below). Responses were on average 25% quicker in Experiment 1 with the photographs (mean $RT = 30$ s) than with the computer-generated scenes in Experiment 2 (mean $RT = 40$ s).

For accuracy, a similar ANOVA again showed a moderate and significant effect of image type ($F(1, 328) = 20.58$, $p < 0.0001$, $\eta^2 = 0.06$), but again no interaction with the 2D or 3D cue effects. Participants made errors 58% of the time in Experiment 1 with the photographs, but only 44% of the time in Experiment 2 with the computer-generated images. This suggests a speed-accuracy trade-off between the two experiments: people responded more quickly and intuitively to the photographs, but at the expense of accurately matching their content to the map.

Across the two experiments, where a 3D landmark cue was present participants responded significantly more slowly (40 s as opposed to 32 s; $F(1, 328) = 18.95$, $p < 0.0001$, $\eta^2 = 0.05$). Where there was a distinctive 2D ground layout cue, participants were slightly slower (39 s as opposed to 34 s; $F(1, 328) = 7.14$, $p < 0.01$, $\eta^2 = 0.02$). There was also a weakly significant interaction between these effects ($F(1, 328) = 3.51$, $p = 0.06$, $\eta^2 = 0.01$). This showed that while people responded more quickly when no 3D landmark was available, their responses in this situation were not affected by presence of distinctive 2D ground layout cues (32 s without 2D, 33 s with). Yet where a 3D landmark was present, the additional presence of 2D cues slowed people down even further (37 s without 2D, 43 s with). This suggests that participants did notice 2D cues even when an obvious 3D landmark did appear, and that they may have struggled to choose which cues to match (or tried to match both, matched the 3D landmark wrongly as usual, and hence faced a conflict with the solution suggested by the 2D cues).

Across the two experiments, where a 3D landmark cue was present participants made significantly more errors (56% with as opposed to 42% without; $F(1, 328) = 36.72$, $p < 0.0001$, $\eta^2 = 0.10$). A distinctive 2D ground layout cue made little significant different to error rates (46% with as opposed to 51% without, $p > 0.10$), but there was a significant interaction between 2D and 3D cue presence ($F(1, 328) = 12.45$, $p < 0.001$, $\eta^2 = 0.04$). When no salient 3D cue was present, and only then, the presence of a 2D cue significantly decreased people's error rates from 48% to only 33%. With a 3D cue present, error rates were 55% without a 2D cue and 56% with it—effectively identical performance.

This suggests that 3D landmark cues were used wherever present, to the almost complete exclusion of detailed 2D ground layout cues. This seems to have occurred even though responses were still constrained by the overall 2D scene structure, and even though the more detailed ground layout cues did attract some attention (but apparently only served to slow people down).

This combined data from the two experiments was also used to check for two other effects of potential interest, taking advantage of the greater statistical power of the total sample of participants. First, separate ANOVAs examined whether trials where the correct response lay in the 'northern' (top) half of the map, where less rotation was required, produced better performance than those where it lay in the 'southern' (bottom) half. There was indeed a moderate and significant effect over the two experiments, for both response times ($F(1, 164) = 11.07$, $p < 0.01$; $\eta^2 = 0.06$) and accuracy ($F(1, 164) = 19.41$, $p < 0.0001$; $\eta^2 = 0.11$). Response times were slower, and errors greater, when the correct response lay in the southern half of the map.

Second, a correlation analysis was run across all response-time and accuracy data from the combined sample, to check for any evidence of an overall speed-accuracy trade-off. There did appear to be some evidence for this across the total sample ($r = -0.28$, $n = 84$, $p = 0.01$; effect size $(r^2) = 0.08$). This is as we might expect in this task: the more carefully cues are selected, and the more cues that are checked between scene and map, the longer it takes to respond, but with a decreasing probability of error (see Pick et al., 1995).

10. GENERAL DISCUSSION

Like Experiment 1, Experiment 2 demonstrated the apparent power of salient 3D visual features in the scene to dominate many people's thinking, with the finer details of the 2D spatial layout apparently often ignored to the detriment of successful orientation. (The other types of error observed in Experiment 1—distance errors and left/right confusions—were also seen in Experiment 2, but again to a lesser extent. It may be assumed that if useful, well-labelled landmarks were added to a map, the frequency of these types of error would decrease along with the 3D/2D ones.)

To truly assess the relative power of the space and the visual scene in influencing orientation performance, we would need to go beyond this factorial "yes/no" design (i.e., 3D/2D cue presence/absence), since real-world scenes do not fall neatly into such categories. However, it is obviously not trivial to assess the proportion of variance accounted for by spatial geometry (i.e., geography) versus visual salience.

For a start, such an assessment would depend on good measures being available for at least one of these two factors. Yet both areas are currently still poorly understood. Although claims are made for *space syntax* as a predictor of human responses (Conroy, Dalton, & Bafna, 2003), and for visual salience in predicting at least eye movements and potentially more (e.g., Itti & Koch, 2001; Henderson, 2003), in neither case is it yet clear how best to measure them. Although space syntax measures (specifically, measures of *isovists* i.e., the visible 2D space from a given point) have a potentially important role in understanding behavioral interactions with environments (e.g., Franz & Wiener, 2008; Wiener, Franz, Rossmanith, Reichelt, Mallot, & Bülthoff, 2007), space syntax currently still has a number of limitations (e.g., Montello, 2007) and offers a variety of poorly-understood spatial isovist measures that could act as predictors, with only some showing limited promise in studies so far (e.g., Dara-Abrams, 2007; Meilinger et al., 2007b; Peebles et al., 2007). Similarly, in the visual attention domain studies are still refining the factors that may predict eye movements (e.g., Tatler & Vincent, 2008), and in any case it is not clear whether behaviours beyond early-stage eye movements can be effectively predicted from salience measures (e.g., Einhauser & Konig, 2003).

The reader will recall that our primary interest was in the feasibility of being able to predict orientation difficulty based on spatial analysis (which could be done using existing mapping data), as opposed to scene content (which would require image analysis of real-world photographic or video data for every individual scene—whatever an ‘individual scene’ is in continuous real-world space). Therefore we have attempted to apply isovist and other spatial measures to the results of Experiment 2. This work has been reported separately (Peebles et al., 2007). To summarise briefly, the findings suggested that the role of these spatial measures is possibly limited.[2] They were of most value in distinguishing among groups of participants who appeared to take different strategies to solving the task, at least for some of the scenes. In other words, where the combination of scene types and participant individual differences encourages a greater reliance on the spatial layout, isovist measures may help to account for this.

Yet, for many scene types and for many more people, the visual scene itself seems to be a greater influence on orientation performance. This difference was clearly demonstrated in our analysis by the contrast between the small group of participants who appeared to be the most influenced by isovist measures (12% of participants) and the largest and least influenced group (41% of participants). In the analysis, the strongest predictors of performance by the former group were ‘occlusivity’ (the extent to which some features of the local environment are hidden by others within the scene) and the presence of strong 2D cues, whereas for the latter group there were no isovist-related correlations over 0.2 and the strongest predictor was presence of a 3D landmark ($r_s = 0.29$).

Perhaps this is unsurprising: after all, successful navigation without maps may favour a focus on—and memory of—visually salient landmarks in the environment. Maps have only been with us for a few thousand years, and navigational maps for far fewer; early maps in our own civilisation almost always showed towns, buildings, bridges and woodlands in oblique perspective to aid recognition (see, e.g., Delano-Smith and Kain, 1999).[3] The strict 2D plan view is a relatively recent development, and seems to require explicit skill development for its effective use in orientation.[4]

Further work is continuing into ways of measuring visual salience and scene content in order to predict performance, but ultimately this may only be useful in a future scenario where every scene in which orientation might be required (e.g., subway exits, public spaces, street intersections) could be captured and analysed from ground-level photography or video data.

11. CONCLUSIONS

We have shown that the design of you-are-here and other orientation-enhancing maps probably cannot be improved merely by using the 2D geographic layout to predict where orientation will be hardest, and then placing some kind of orienting symbol or landmark at such locations. Rather, the 3D visual scene in front of the wayfinder or observer will also make a difference to the ease with which they match it to a map. Of course, with permanent static you-are-here maps whose location is predetermined (usually where there is suitable open space and where tourists/wayfinders are most likely to need help), the scene around them can be recorded and analysed to determine the most salient 3D landmark, and an image of this can be placed appropriately on the map to aid orientation as well as self-location.

Map-based orientation also occurs in many other scenarios, such as in mobile navigation devices or traditional portable mapping, or in scenarios where someone has to match multiple scenes to a map without necessarily being present within those scenes (as in use of mobile closed-circuit television images for crime analysis, or identifying old photographs in historic research). For such applications, multiple scenes may have to be assessed from multiple viewpoints in order to decide both on the likely level of orientation difficulty, and the degree and type of intervention required to help. Elsewhere, work has already begun on assessing the relative visual salience of landmarks (e.g., Klippel & Winter, 2005). To assess when and whether they are required for orientation at all—since cluttering maps with extra orienting landmarks is not always desirable—will require further developments in image analysis, and in our understanding of how visual salience operates and influences visual cognition.

ACKNOWLEDGMENTS

This work was funded by Ordnance Survey of Great Britain. Portions of the data from Experiment 2 were previously presented in Peebles, Davies, and Mora (2007). The authors would like to thank: the 84 participants; Rodrigo Mora for discussions and the isovist measurement calculations; Jon Gould and Claire Cannon for help with qualitative data analysis; Guy Heathcote, Tim Martin, Jon Horgan, David Capstick and Isabel Sargent for help with the stimuli; Glenn Gunzelmann and Glen Hart for helpful discussions; the three paper reviewers for very useful suggestions. This article has been prepared for information purposes only. It is not designed to constitute definitive advice

on the topics covered and any reliance placed on the contents of this article is at the sole risk of the reader.

REFERENCES

Aretz, A. (1989). Spatial cognition and navigation. In *Perspectives: Proceedings of the Human Factors Society 33rd annual meeting* (pp. 8–12). Santa Monica CA, USA: Human Factors Society.

Blades, M., & Spencer, C. (1990). The development of 3- to 6-year-olds' map using ability: The relative importance of landmarks and map alignment. *Journal of Genetic Psychology, 151,* 181–194.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (Second ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conroy Dalton, R., & Bafna, S. (2003). The syntactical image of the city: A reciprocal definition of spatial elements and spatial syntaxes. Paper No. 59. In *4th International Space Syntax Symposium, London, UK, 17–19 June 2003*. London: University College London.

Dara-Abrams, D. (2007). Architecture of mind and world: how urban form influences spatial cognition. In C. Hölscher, R. Conroy Dalton, & A. Turner (Eds.), *Space Syntax and Spatial Cognition* (pp. 93–111). Bremen, Germany: Universität Bremen (SFB/TR 8).

Delano-Smith, C., & Kain, R. J. P. (1999). *English Maps: A history*. London: The British Library.

Einhauser, W., & Konig, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*, 1089–1097.

Eley, M. G. (1988). Determining the shapes of land surfaces from topographical maps. *Ergonomics*, *31*, 355–376.

Franz, G., & Wiener, J. M. (2008). From space syntax to space semantics: A behaviourally and perceptually oriented methodology for the efficient description of the geometry and topology of environments. *Environment & Planning B: Planning and Design*, *35*(4), 574–592.

Griffin, T. L. C., & Lock, B. F. (1979). The perceptual problem of contour interpretation. *The Cartographic Journal*, *16*, 61–71.

Gunzelmann, G., Anderson, J. R., & Douglass, S. (2004). Orientation tasks with multiple views of space: Strategies and performance. *Spatial Cognition and Computation, 4,* 207–253.

Hagen, M. A., & Giorgi, R. (1993). Where's the camera. *Ecological Psychology, 5,* 65–84.

Harwood, K. (1989). Cognitive perspectives on map displays for helicopter flight. In *Perspectives: Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 13–17). Santa Monica CA: Human Factors Society.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30,* 425–447.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7,* 498–504.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature, 370,* 57–59.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Klippel, A., & Winter, S. (2005). Structural Salience of Landmarks for Route Directions. In A. G. Cohn & D. M. Mark (Eds.), *Conference on Spatial Information Theory—COSIT 2005* (pp. 347–362). Lecture Notes in Computer Science Vol. 3693. Berlin: Springer.

Meilinger, T., Hölscher, C., Büchner, S. J., & Brösamle, M. (2007a). How much information do you need? Schematic maps in wayfinding and self localization. In: T. Barkowsky, M. Knauff, G. Ligozat, & D. R. Montello (Eds.), *Spatial Cognition V: Reasoning, Action, Interaction*. International Conference Spatial Cognition 2006 (pp. 381–400). Lecture Notes in Artificial Intelligence, Vol. 4387. Berlin: Springer.

Meilinger, T., Franz, G., & Bülthoff, H. H. (2007b). From isovists via mental representations to behaviour: first steps toward closing the causal chain. In C. Hölscher, R. Conroy Dalton, & A. Turner (Eds.), *Proceedings of Space Syntax and Spatial Cognition Workshop (Spatial Cognition '06),* 1–16. (pp. 65–80). Bremen, Germany: Universität Bremen (SFB/TR 8).

Montello, D. (1993). Scale and multiple psychologies of space. In A. U. Frank, & I. Campari (Eds.), *Spatial Information Theory: A Theoretical Basis for GIS*. Proceedings of COSIT '93, pp. 312–321. Lecture Notes in Computer Science Vol. 716. Berlin: Springer-Verlag.

Montello, D. R. (2007). The contribution of space syntax to a comprehensive theory of environmental psychology. Unpublished invited paper presented at: 6th International Space Syntax Symposium, Istanbul, 2007.

Peebles, D., Davies, C., & Mora, R. (2007). Effects of geometry, landmarks and orientation strategies in the 'drop-off' orientation task. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds.), Spatial Information Theory, Proceedings of COSIT 2007 (pp. 390–405). Lecture Notes in Computer Science Vol. 4736. Berlin: Springer.

Pick, H. L., Heinrichs, M. R., Montello, D. R., Smith, K., Sullivan, C. N., & Thompson, W. B. (1995). Topographic map reading. In P. A. Hancock, J. M. Flach, J. Caird, & K. J. Vicente (Eds.), *Local applications of the ecological approach to human-machine systems, Volume 2* (pp. 255–284). Hillsdale, NJ: Lawrence Erlbaum Associates.

Presson, C. C. (1982). The development of map-reading skills. *Child Development, 53,* 196–199.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2,* 1–18.

Warren, D. H., Rossano, M. T., & Wear, T. D. (1990). Perception of map-environment correspondence: The role of features and alignment. *Ecological Psychology, 2,* 131–150.

Wickens, C. D., & Prevett, T. T. (1995). Exploring the dimensions of egocentricity in aircraft navigation displays. *Journal of Experimental Psychology-Applied, 1,* 110–135.

Wiener, J. M., Franz, G., Rossmanith, N., Reichelt, A., Mallot, H. A., & Bülthoff, H. H. (2007). Isovist analysis captures properties of space relevant for locomotion and experience. *Perception, 36*(7), 1066–1083.

Winter, S. (2003). Route adaptive selection of salient features. In W. Kuhn, M. Worboys, S. Timpf, J. Hartmanis, & J. van Leeuwen (Eds.), *International Conference, COSIT 2003, Spatial Information Theory: Foundations of geographic information science* (pp. 349–361). Lecture Notes in Computer Science Vol. 2825. Berlin: Springer.

NOTES

1. In this section, except where otherwise stated, effect sizes have been calculated from ANOVA tables using the 'partial eta squared' (η^2) method: the sum of squares for the effect is added to the residual (error) sum of squares. Effect size is the ratio of the effect sum of squares to that total. We interpreted this measure according the conventional criteria of a small effect when $\eta^2 < .06$, a moderate effect when $.06 \leq \eta^2 < .15$ and a large effect when $\eta^2 \geq .15$ (Cohen, 1988).
2. A broad range of isovist measures from the space syntax literature was included, and these were calculated both for the full 360-degree isovist and for the 60-degree visible scene shown in the experiment stimuli (although little difference was found). Further details can be found in Peebles et al (2007).
3. Thanks to Glenn Gunzelmann for this observation.
4. For example, some unpublished data from our own work, administering the Experiment 1 stimuli in paper-and-pencil form to a sample of longstanding Ordnance Survey field surveyors who daily match similar large-scale maps to outdoor scenes, suggested extremely high performance levels with very few errors and rapid overall performance (individual trial response times were not collected).