Effects of Geometry, Landmarks and Orientation Strategies in the 'Drop-Off' Orientation Task[⋆]

David Peebles[1], Clare Davies[2], and Rodrigo Mora[3]

[1] Department of Behavioural Sciences, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK. `D.Peebles@hud.ac.uk`
[2] Research Labs C530, Ordnance Survey, Romsey Road, Southampton, SO16 4GU, UK. `clare.davies@ordnancesurvey.co.uk`
[3] Bartlett School of Graduate Studies, University College London, 1–19 Torrington Place, Gower Street, WC1E 6BT, UK. `r.vega@ucl.ac.uk`

Abstract. Previous work is reviewed and an experiment described to examine the spatial and strategic cognitive factors impacting on human orientation in the 'drop-off' static orientation scenario, where a person is matching a scene to a map to establish directional correspondence. The relative roles of salient landmarks and scene content and geometry, including space syntax isovist measures, are explored both in terms of general effects, individual differences between participant strategies, and the apparent cognitive processes involved. In general people tend to be distracted by salient 3D landmarks even when they know these will not be detectable on the map, but benefit from a salient 2D landmark whose geometry is present in both images. However, cluster analysis demonstrated clear variations in strategy and in the relative roles of the geometry and content of the scene. Results are discussed in the context of improving future geographic information content.

Keywords: orientation, landmark salience, space syntax, isovists, individual differences, visualisation, spatial cognition

1 Introduction

Part of the promise of applying a scientific approach to spatial information lies in its potential to enhance the information to improve future geographic data and mapping, together with the systems that manipulate it. If we are to improve geographic visualisations such as maps, whether on paper or screen, it is essential to understand the core tasks that users have to perform when interacting with geographic information. More importantly, as argued by other authors over the years (e.g., [1], [2]) we need to understand the role of the characteristics of the space itself, and of its representation, in influencing human performance on those core tasks.

One such core task is what has been called the "drop-off localisation problem" [3], [4]. In this situation a person is either viewing or is actually immersed within

---

[⋆] © Crown copyright 2007. Reproduced by permission of Ordnance Survey.

a scene, and has to look for the first time at a map of the area to try to match it to their orientation (and possibly their location). Of course, this problem can arise at a decision point within a navigation task, although the abrupt sense of 'drop off' (i.e. not having used the map already to progress to this point in space) would only be likely if the wayfinder was suddenly disoriented for some reason, for example if they had just reached an unfamiliar area after traversing a familiar one. There are also many non-wayfinding situations where it may arise. A few examples of this task scenario in an urban setting might include:

1. trying to identify a specific building or object which is not explicitly labelled on the map, e.g., to visit or study it, or in an emergency scenario;
2. trying to match a historic image (e.g., of an old street scene) to a modern-day map, or vice versa;
3. making planning decisions based partly on viewing the current visual landscape (or photographs of it), and partly on a drawn plan or model of a proposed development;
4. trying to judge relative distances and directions to unseen distant locations (whether or not one intends to navigate to them);
5. viewing a 'you-are-here' map signage within a space, where location is indicated but orientation is unclear [5], [6].

For this reason, the task can be seen as an important one to study and model as it may inform the development of geographic information content and visualisation to facilitate orientation across a wide range of uses, from emergency services to tourists and from urban planners to archaeologists. It may also be seen as a precursor to studying the more complex combined task of self-localisation, which inevitably includes simultaneous orientation to some extent.

The few previous studies of this process of orienting with a map in a drop-off scenario have taken an experimental approach, usually using a scene which is limited or simplified in some way. There are important differences, but also important similarities, among these previous studies, which we outline below. In the remaining sections of this paper we will describe the experimental approach we have taken, the results of an initial experiment attempting to focus on the role of the scene geometry in people's strategies to solve the task, and an analysis of those strategies to illustrate the individual differences that occur and the apparent role of different spatial metrics (mostly derived from the field of space syntax research) within each identified strategy.

2 Previous work

Studies[4] of drop-off orientation that involve physically matching a map to a scene (as opposed to viewing only one of them to infer what the other would

---

[4] For simplicity we have omitted studies that tested orientation processes in the absence of a map, and those where the matching took place from memory, e.g., location/direction judgements made after learning a map. However, it is recognised that these may also involve some common cognitive processes with the task of real-life orientation.

look like) have differed in a number of aspects, which makes their results difficult to collate into a single view of this task.

The first source of variation has been the type of space being matched. A common focus (e.g., [7], [8], [4]) has been on matching topographic maps to rural landscapes, where the primary focus is on the shapes of visible landforms. However there has also been an extensive body of work on orientation within aviation, where the scene view is partly from above rather than immersed within the landscape (e.g., [9], [10], [11], [12]). Meanwhile a different approach [13] required participants to judge their position relative to a single building. Other studies (e.g., [14], [15], [16], [17]) have asked either adults or children to match a map to an even smaller, room-sized space, or to images of it, or to a larger indoor space (e.g., a conference centre).

The second variation lies in the task: some studies (e.g., the field study of Pick et al [4], [18]) have actually immersed participants in a real or a virtual environment and asked them to match it to a map, whereas most other studies have relied on a laboratory simulation where the participant views a static image of a scene. Perhaps more fundamentally, while some of the above studies asked participants to localise (locate) themselves on the map as well as to orientate, others asked only for one or the other—either by marking the participant's position on the map, or by asking for such a mark without asking for direction of view. Arguably, in the latter case, orientation as well as localisation is implied (because the localising task requires one's position relative to nearby objects to be established), but the reverse is not the case. However, since orientation without localisation is often the case in real-world scenarios (e.g., emerging from a subway station, or matching the map to a photograph taken at a named location), this focus has the advantage of narrowing the task and hence the cognitive processes under study (reducing noise in the experimental data) without completely losing ecological validity.

Most of the above studies have focused on response time as the dependent variable indicating task difficulty and performance. However, the average response time is likely to vary greatly with the complexity of the environment and map, from a few seconds in the case of simple displays [19] to (apparently) whole minutes in a field study [4]. Furthermore, the level of accuracy obtained in people's responses also varies greatly, with a typical score of around 50% correct in Pick et al's laboratory study [4], whereas Gunzelmann and Anderson's participants reached near-perfect performance [19].

Despite these many differences of task, focus and outcome, some general conclusions can be drawn about the factors affecting orientation performance. First, there is almost always an effect of the alignment of the map relative to the forward view of the observer. Many studies, both with navigation and with static orientation, have shown that the mental rotation necessitated by map misalignment can have systematic effects on performance (e.g., [19], [13]). Meanwhile, familiarity with the map through its use in previous tasks appears to improve performance if those tasks involved a focus on its geographic content and frames of reference [20].

Another common finding appears to be the role of prominent landmarks and groups 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—rather like having a 'north' arrow painted on the ground in the scene. This tendency to shortcut the matching task by finding a unique landmark to match instead, has been argued to be a late-developing strategy in human orientation in general [21]. If matching the whole geometry is the default for young children and mature rats, as Hermer and Spelke have famously claimed [21], then it is perhaps surprising to see a role for a landmark (by which in this context we mean any feature whose relative location can be used to aid the matching process) creeping into the drop-off map-matching task, even in extremely sparse scenarios with fairly simple geometric layouts [19]. People's apparent use of such features, often apparently via some kind of propositional description of approximate relative location, may be related to the findings of an apparent tendency to code object location in both an exact and an inexact way in spatial memory [22]. The suggestion that the inexact description of relative location (depending on a landmark or other simplifying cue) may be to some extent a verbal strategy may help explain Hermer and Spelke's failure to find landmark use in their task in very young children or in rats. Although a linguistic explanation has been disputed by some authors [23], it does seem that the use of a landmark is an approximate strategy which functions somewhat as if using a verbal description—relying on an inexact representation of location rather than an exact spatial calculation.

If a landmark is often chosen as a shortcut to matching, what is likely to be chosen? The most systematic study of what makes a landmark more or less suitable, albeit in the context of wayfinding rather than static orientation, is probably that of Winter [24]. This showed that if we assume a landmark is a feature of the scene that is somehow salient to the viewer, this may be a complex mix of visibility (attracting visual attention) and structural salience (relevance to the task). With drop-off static orientation involving a map, however, we may surmise that a key aspect should be the appearance of the landmark on that map.

If a 3D landmark is not labelled on the map such as to make it recognisable and hence easily matchable, and assuming that its most salient feature (e.g., height or unusual roof shape) is not shown in the planimetric 2D view that most maps represent, then it cannot be used effectively for matching. Any attempt to do so is likely to impair performance, either by slowing it or by creating misinterpretations (wrong answers), or both. However, if a salient 2D shape is distinctive and the map is of sufficient detail to reflect that shape, use of the landmark for matching should greatly reduce response latencies and improve accuracy. We therefore tested the relative effects of two- and three-dimensionally salient landmarks in the study we report below.

Previously, apart from observing some role for landmarks such as distinctive clusters of features, no studies have attempted to examine systematically the role of the spatial geometry itself in predicting the ease or difficulty of orientation with a map. In addition, and perhaps surprisingly, few studies have tested orientation (as opposed to navigation) performance within the type of space where the task perhaps most commonly occurs: urban street scenes. Arguably, while the problems of interpreting topography in rural landscapes are well established (e.g., [7]), the opportunity to improve larger-scale urban street mapping to facilitate orientation has been neglected. Furthermore, if landmark matching is indeed an optimal strategy for orientation in any environment, then understanding how this works could help improve all types and scales of geographic spatial representation.

The experiment described below was designed to address these issues by investigating orientation strategies where the scene people viewed contained only the 2D ground layout and the 3D building shapes. The scene images were generated using a 3D model of a UK city (Southampton). All irrelevant details that could distract from the use of an optimal strategy for matching to the map were removed. An example scene and map used in the experiment are shown in Figure 1, together with the corresponding Southampton street location. The map itself contained no name labels or other indicators to distinguish the geographic features. The only remaining salience cues for items within the scene were size (both in terms of ground area and height), shape (again in terms of both roof line and ground layout), and colour (since the same colour scheme was used for both the scene and map, to emphasise the similarity of their 2D geometry and to facilitate its use in matching). In these scenes, therefore, choosing a single 3D item to match to the map was unlikely to be successful, since its 2D geometry was usually ambiguous, although distinctive 2D features and the overall ground layout were not.

3 Method

3.1 Design and Participants

Forty-nine students and members of staff from the University of Huddersfield took part in the experiment. All participants saw the entire set of stimuli in random order. An additional five participants carried out the experiment while having their eye movements and verbal protocols recorded to enable qualitative assessment of their apparent strategies in solving the task. The 49 participants in the main study were encouraged to perform the task as quickly and accurately as possible.

3.2 Materials

The computer based experiment was carried out using PC computers with 17 inch displays and the eye movement and verbal protocol study was conducted using a Tobii 1750 remote desktop eye tracker with a 17 inch display.

Twenty-five scenes from various locations in the city of Southampton, UK were generated using a buildings-only 3D model overlaid on OS MasterMap® Topography Layer and draped on an OS Land-Form PROFILE® terrain model to provide a realistic and accurate representation of height information. The corresponding maps were circular sections of OS MasterMap® Topography Layer at 1:1250 scale. The scenes were selected from photographs of the actual street locations in Southampton (see e.g., Figure 1a) in order to allow subsequent replication of the experiment with the photographs and were chosen to represent a wide range of building shapes, degrees of salience and distinctiveness, together with a range of urban features such as green spaces and road patterns. The colour schemes of the scenes and maps were matched in order to remove any unnecessary distracting information and to facilitate orientation using spatial geometry. Although this procedure reduced the possibility of participants using anything other than the visible geometry, sometimes this still entailed the presence of a landmark (only a completely uniform scene would have no variations in salience).

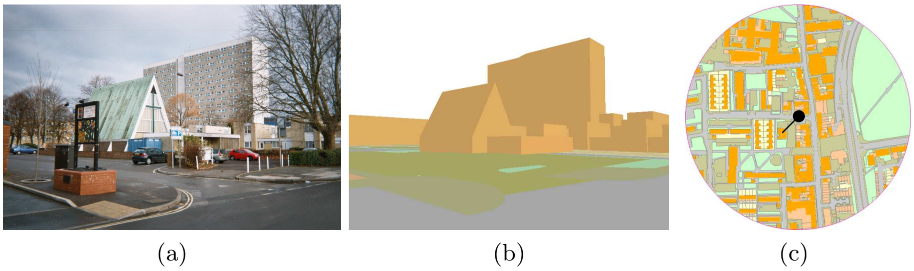

(a) (b) (c)

Fig. 1: Street location (a) for example scene (b) and corresponding map (c), used in the experiment. © Crown copyright 2007. Reproduced by permission of Ordnance Survey.

We hypothesised that a prominent 3D landmark, if its 2D geometry was not especially unique or salient, might distract participants into either slowing their decision or making an incorrect one. An obvious 2D cue in the scene, however, which would be reflected as such on the map, ought to help performance, but in these scenes a 2D shape only tended to be unambiguously salient when it formed part of the foreground layout geometry (e.g., an unusually shaped lawn or pathway). Therefore, we coded the scenes according to occasions when a single object or feature was particularly separable and hence individually salient, either in 3D or in 2D. Thus roughly a quarter of the scenes were deemed to contain each of four possible scene types: an obvious 2D foreground landmark, a prominent 3D landmark, both, or neither. The stimuli were also controlled for alignment by ensuring that the map alignment angle ranged over a roughly even spread from −180 to +180 degrees, independently of other aspects of the scenes.

The scene-map pairs were presented sequentially on a 17-inch computer monitor, using specially-programmed software written in tcl/tk which also recorded response locations and times. Each trial presentation, including the five initial practice trials, was separated from the next by a blank screen with a button on it which the participant had to click to move on. This gave all participants a chance to break if needed. It also allowed us to check the eyetracker calibration, where used, by observing the gaze trace on the button between trials.

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 and told that their task was to work out in which direction they must be facing on the map to see the scene. A black dot in the centre of each map indicated the location of the observer. When the mouse cursor was moved over the map, a short black line of fixed length was drawn from the centre of the dot toward the tip of the cursor (see e.g., Figure 1c). This rotated around the dot in either direction as the mouse was moved around, to follow its position. Participants had to click on the map when they believed they had aligned the pointer towards the centre of the scene on the left of the screen. Participants were asked to respond as rapidly and as accurately as possible. They were told that the maps were all at the same scale, and that they should avoid the natural assumption that the 'upwards' direction on the map indicates 'forward' in the environment [25].

There were five practice trials and twenty experiment trials in total. When the participant responded by clicking on the map the angle of the response from the vertical was recorded, as well as the response time from the onset of the stimulus. Participants in the eye movement and verbal protocol study were asked to talk through each trial as they attempted to solve the problem, in particular to say what they were looking at, how they were thinking through the problem, and why and how they were choosing a particular direction.

Space Syntax measures Space syntax measures of urban spaces, although originally focused mainly on understanding paths through it in terms of axial lines, has in recent years also focused on the concept of an *isovist*—the 2D shape that is visible from standing at a particular point in space (and rotating one's body through 360 degrees). Most metrics that can be used to describe an isovist were proposed some years ago [26] although some additional ones have been more recently proposed by other authors (e.g., [27], [28]). A review of the potential of such measures [29] suggested that they may have relevance in helping people to orientate, since the shape and size of the space may make it easier or harder to deduce direction and position. Accordingly, for the twenty scenes used in the main experiment, and looking at both the usual 360° isovist (which of course was visible on the map but not in the scene) as well as on the 60° section of it that was visible within the scene as well, we calculated various metrics as suggested in the space syntax literature. Figure 2 shows (on a simplified, buildings-only version of the map) the 360° and 60° isovists for the scene shown earlier in Figure 1.

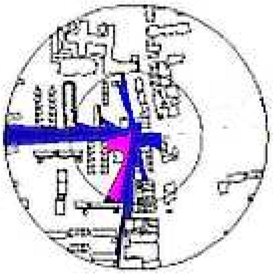


Fig. 2: Isovist at ground level for the scene in Figure 1, showing both the full 360° version (dark blue) and the 60° segment (lighter pink) visible in the scene image.

The isovist measures that we took for our analyses included all of those that have been identified in the space syntax literature which we felt could have some conceivable role in people's cognition of the scene and map. These included the area and perimeter length of the isovist, its minimum and maximum radius, and these additional measures of its geometry:

1. *Occlusivity*: the extent to which some features of the local environment are hidden by others within the scene.
2. *Compactness*: nearness of the isovist shape to a circle.
3. *Jaggedness*: tending to be inversely related to compactness, this indicates the complexity of the isovist shape (e.g., an isovist from a crossroads in a highly built-up area may be shaped like a long thin cross).
4. *Drift magnitude*: distance of the viewer's location from the centre of gravity of the isovist shape. (Broadly, this and drift angle indicate the level of asymmetry in the isovist; one might expect that an asymmetrical isovist is easier to match unambiguously to a map if the isovist shape is used at all by participants.)
5. *Drift angle*: the angular distance in degrees of the viewer's location from the centre of gravity of the isovist shape. Measured relative to a horizontal line (east).

In addition to these, measures were also calculated that considered the content of the scene. These included the extent to which the isovist perimeter was defined by buildings (as opposed to the edge of the map—we restricted the isovist to the scene that the map depicted, i.e. within the 400m-diameter circle).

Other such measures included the proportion of the scene's 2D area that consisted of surface features, since the sidewalks, streets, vegetation and occasional unclassified areas were all distinguished on the map as well as in the scene.

Finally, the above measures were all taken both for the overall 360° isovist, and for the 60° angle subtended by the scene (which is typical of a photograph from a normal camera lens). If a participant was focusing on aspects within the geometry of the scene, either initially or after identifying the broad scene orientation on the map, then it was felt that these versions of the measures might be more relevant than the overall isovist. On the other hand, the overall isovist measures might logically be expected to prove more significant for placing the overall scene geometry within the map's circular area.

4 Results

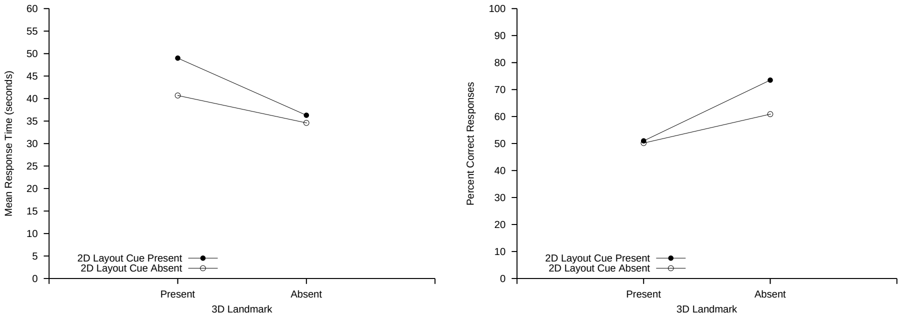
4.1 General

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. within the 30 degree range that it bisected, cf. [13]), at the point when the participant clicked the mouse. Given that the scenes tended to subtend about 60 degrees of visual angle in total, which is also typical of a photograph taken with a normal camera, this meant that the participants had got within 'half a scene' of the exact line.

In general, participants were 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 s (SD = 9.76). A Spearman's rho test produced a moderate but non-significant negative correlation between the probability of a correct response and latency, $r_s = -.410$, $p = .072$ indicating that differences do not result from a speed-accuracy tradeoff but suggesting that both measures tended to indicate similarly the relative difficulty of the task for a particular stimulus.

In order to test whether performance was influenced by the presence of salient 3D landmarks, the scenes were coded according to the presence or absence of such a landmark. Ten scenes included at least one. Similarly, scenes were also coded according to the presence of distinctive 2D ground layout information in the foreground of the scene, which would facilitate identification on the grounds of 2D layout. Nine of the 20 scenes included such 2D features, e.g., an extensive and irregularly shaped strip of lawn or pavement in the foreground. For example Figure 1 shows a scene that includes both a salient 3D landmark and distinctive ground layout cues.

The mean response time and percentage of correct responses for the stimuli categorised by the presence or absence of 2D and 3D cues are presented in Figure 3. Separate 2×2 repeated-measures ANOVAs were performed on participants' error rates and response times, with presence or absence of 2D and 3D cues as the two within-subjects factors. For errors, there was a highly significant effect of presence of salient 3D landmarks, $\mathrm{F}(1, 48) = 40.35$, $\mathrm{p} < .0001$, and a much smaller but still significant effect of presence of distinctive 2D ground layout, $F(1, 48) = 5.47$, $p < .05$. There was also a significant interaction between them, $F(1, 48) = 5.26$, $p < .05$. The directions of these effects showed that while presence of an obvious 2D cue was able to decrease error rates, this was only in the absence of a salient 3D cue which always greatly increased them.



(a) (b)

Fig. 3: Mean response time (a) and percentage of correct responses (b) for stimuli categorised by the presence or absence of 2D and 3D cues

The analysis of response times, however, showed that both 3D and 2D cues seemed to slow participants down — again much more so for 3D, $F(1, 48) = 29.7$, $p < .0001$ than for 2D, $F(1, 48) = 9.28$, $p < .005$. There was again a mild interaction, $F(1, 48) = 4.37$, $p < .05$, which indicated that the presence of both a 2D and a 3D cue had the most marked effect of all on response times; the presence of a 2D landmark made only a small difference except when a 3D landmark was also present.

Some caution should be expressed with the above analyses since both the response time and error data showed minor deviations from normality; however, the main effects were also checked using non-parametric Wilcoxon signed-rank tests, which showed the same strong significance patterns (but could not, of course, test the interaction effects).

This finding was independently confirmed by qualitative verbal protocol and eye movement analysis of the five additional participants. By far the most commonly reported feature used for solving the problem was ‘buildings’, and the eye movement patterns in the scenes with the most salient 3D landmarks (e.g., large skyscrapers or church steeples) tended to strongly focus around those landmarks.

4.2 Map alignment

Previous studies where a map is matched to a scene have tended to find a distinctive ‘M’ shape pattern in the effect of map alignment with observer position (e.g., [19], [30]). Performance typically is better not only at 0 degrees (where 'up' on the map exactly corresponds to the forward direction within the scene), but also at 90, 180 and 270 (i.e. -90) degrees. It seems that mental rotation to these cardinal directions is easier than with more oblique angles. Although there was a modest effect of map alignment on response time in the current study, the M shape pattern was considerably less well defined than those found in other studies. This is possibly due to the fact that alignment angle was varied semi-randomly rather than at fixed points (such as 0, 45, 90, etc.) in this study and because of the possibility that landmark-based orientation strategies sometimes make mental rotation unnecessary (cf. [19]).

4.3 Space Syntax measures

A multiple regression analysis incorporating the space syntax measures outlined above was applied to the error data to determine whether they were able to account for the differences in number of correct responses to each scene (described in further detail in [31]). It was found, however, that the space syntax measures appeared to show little clear predictive power for the overall performance across participants. This apparent lack of a clear consistent role for the scene geometry prompted a systematic analysis of individual differences, discussed below.

4.4 Individual differences

In order to see whether individual differences in strategy could be linked to spatial metrics of the different scenes, the following analysis of the response time data was undertaken. First, for each participant and across the 20 scenes, correlation coefficients were calculated (using non-parametric Spearman correlations due to non-normality of some variables) to indicate the size of effect on performance of the various spatial metrics. These coefficients were then used in a cluster analysis, to examine apparent groupings of participants in terms of which spatial variables appeared to most influence their performance. The cluster method was Ward's linkage [32], since this minimises the variance within groups and thus would most clearly highlight similarities among participant strategies. Squared Euclidean distance (E^2) was used as the similarity measure, as generally recommended in the literature for this clustering method. The Duda-Hart stopping rule [33] indicated that four clusters was the optimum solution in terms of distinguishing clear groups. These are indicated by the horizontal line in the dendogram shown in Figure 4.

Table 1 shows the four identified clusters, the cluster sizes (i.e. number of participants in each group), the key spatial variables whose correlation patterns appeared to strongly distinguish that group's performance (with the mean Spearman r coefficient of the correlation with response time — so a positive correlation meant slower times), and the apparent broad strategies that were thus implied.

It can be seen that the groups overlap in strategy use, but apparently differ in the most common 'default' strategy (as inferred from the aspects of the space that appear to affect their performance). Groups I and II were affected by the

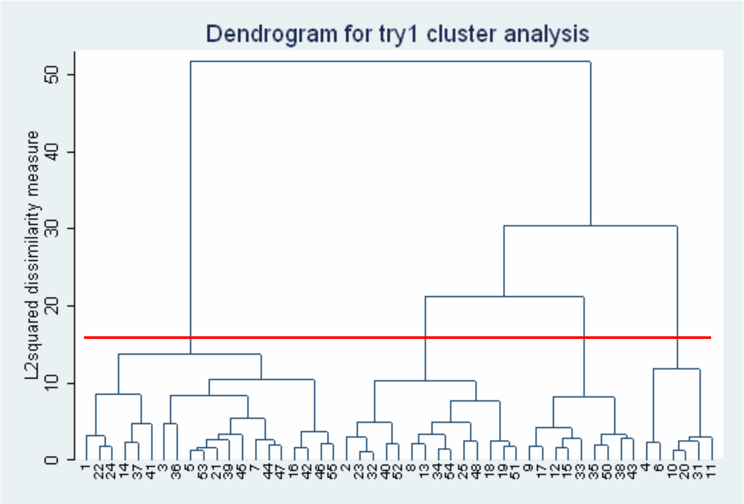


Fig. 4: Dendogram showing the clustering of participants' responses according to the four primary spatial variables described in Table 1. The horizontal line shows where the Duda-Hart stopping rule indicated the optimum number of clusters.

presence of a 3D landmark more than groups III and IV. Group III appeared to use the 'optimum' strategy of focusing on the overall scene geometry and rotating it to match the map; surprisingly, none of the other groups appeared to be slower with greater map misalignment.

The table shows that Group III's response times were facilitated by a strongly jagged scene geometry (which in this environment usually implies a road junction with more than one connecting street), and by a scene where few objects were obscured by others (i.e. a built-up scene whose visible perimeter was largely formed by buildings rather than the obscured spaces behind them). They performed more slowly where the angle of map misalignment was greater, and where there was more vegetation in the scene (probably implying a more open space with more scattered buildings).

One-way ANOVAs to compare the clusters on errors and response times found no overall difference for response times, but did show significant differences in error levels between the groups, $F(3, 45) = 3.00$, $p < .05$). Post hoc contrasts showed that this was due to group III performing significantly better than Groups II or IV (with Group I's performance falling somewhere in between). On average Group III were correct 72% of the time, compared with 63% for Group I, 49% for Group II and 47% for Group IV. This suggests that using the 'correct' strategy of matching overall geometry and performing mental rotation did produce optimum performance on the task, but was only adopted by a minority of participants. Yet the largest group of participants—around 40%

in Group I—did not perform significantly worse in general than this 'optimal' strategy group.

Table 1: Identified clusters, contributing correlation coefficients, and the corresponding apparent strategies by participants.

Group	N	Key spatial variable effects	Inferred strategies
I	20	Overall isovist area (−0.18); 3D landmark presence (0.29); No mean correlations above 0.3	Little focus on isovist shape. Probably picking a single feature to match: possibly from ground layout, since faster with more open spaces but distracted into trying to use salient 3D landmarks when present.
II	14	Within-scene isovist occlusivity (0.34); Perimeter length (0.36); Drift magnitude (0.28) & area (0.36); 3D landmark presence (0.29).	Abstracting the 2D isovist geometry and then matching it to the map but distracted into trying to use salient 3D landmarks when present.
III	9	Map alignment (i.e. angular bearing of scene centreline from north, 0.26); Overall isovist compactness (0.28) & jaggedness (−0.28); vegetation extent (0.27); proportion of isovist perimeter formed by buildings (−0.36).	Focus on street pattern in built-up areas (enhanced when streets are lined with buildings making their shapes more salient); hence dependence on mental rotation to match to map (worsened by map misalignment)
IV	6	Presence of strong 2D foreground cues (0.51); overall isovist area (0.30), occlusivity (0.37) and perimeter length (0.35); extent of visible area showing as footpaths (sidewalks etc., 0.26), as streets (−0.30), and as undefined ground cover (−0.32); proportion of isovist perimeter formed by buildings (−0.33).	Use of both ground layout patterns and abstracting the overall isovist geometry (easier when isovist smaller and more clearly defined by buildings), but highly distracted by attempt to match foreground 2D cues when available.

5 Discussion

The results of the experiment are consistent with previous studies in showing that, when possible, people tend to match a single salient landmark between a 2D and 3D representation of a scene, and particularly to pick on a landmark with a distinctive 3D (but not 2D) shape despite the absence of that shape in the 2D map. This is particularly noteworthy given that this strategy was discouraged by the nature of the stimuli. In the scenes used in the present study, as in the studies by Gunzelmann and Anderson [19], the 2D shapes and colours were directly matchable between the scene and the map (though they would not be in real-world scenes or photographs), and all distracting salient cues were removed other than the 3D geometry. Despite this, participants still made errors through attending to the latter rather than the more reliable 2D geometry, most likely due to the particular visual salience of landmarks in the scenes (cf. Winter [24]).

A clue as to a potential reason for people's sometime preference for inappropriate landmark use, rather than sticking to the more reliable 2D geometry, lies in the finding that the presence of a strong 2D cue (whose visual salience would perhaps push participants towards its use) seemed to actually slow people down. It seems reasonable to assume that the process of extracting an overhead 2D geometric configuration from the 3D scene, and then carrying this over to rotate and match to the 2D layout of the map, may sometimes create more cognitive load than finding an alternative such as matching a single feature or taking an approximate, broad account of the approximate layout (e.g., just being aware that one is 'looking down the road').

As well as the obvious implications for understanding human cognition of large-scale spaces, this may also help to explain the public popularity of 'bird's eye' urban maps that show the buildings from an oblique angle rather than from overhead [34]. It also implies that if large-scale maps were to be designed explicitly to aid their use in orientation, it would not be sufficient merely to include orienting landmarks at places where the 2D geometry was an ambiguous cue, since it may not be used efficiently even when unambiguous.

Analysing individual differences via the cluster analysis provides a different perspective however. Here the different aspects of the spatial geometry and features are shown to be relevant to specific strategies for solving the task. Almost half of participants did show a reliance on single salient landmarks, as implied by our overall analysis and by previous studies (e.g., [4]), and a further quarter of the sample would be distracted into this strategy when a salient 3D landmark was offered. However, just over half of participants actually did appear to show some efforts to abstract the 2D isovist geometry, the (simpler) street pattern, or the patterns made by ground layout features. Also, although the overall analysis had suggested that the presence of a 2D foreground landmark generally improved performance, the individual differences analysis showed that it actually slowed down a small minority of participants (possibly because it detracted from their preferred strategy of abstracting a more general sense of the ground layout and/or isovist geometry). Meanwhile, although the minority of participants who adopted the optimal geometry-matching-and-mental-rotation strategy did perform best, it was not significantly better than the largest group of participants who would apparently be helped by being able to reliably reference and match a single feature between the map and the scene—in other words, some kind of landmark. If, say, a church in real life was marked with a church symbol on the correct street corner on the map, then matching would become trivial and highly accurate for this largest group of participants.

It seems, therefore, that the potential role of space syntax measures in interpreting cognitive tasks of this nature is one that is highly dependent on problem-solving strategies, rather than as an overall predictor of task difficulty. It also appears that even when the abstraction of the 2D geometry is the only reliable cue for solving the task, as was the case in the present experiment, a majority of people will still attempt to rely on salient landmark cues within that geometry, whether or not they are discernable from the map.

This confirms the value of landmarks in aiding orientation with maps, but also warns us that the match between the landmark's appearance in the real world and on the map must be unambiguous and rapid if errors are to be avoided. With less congruent representations, e.g., a photograph or actual real-world scene where colours and shapes will usually differ between the scene and the map (and where the map is likely to be smaller-scale and hence subject to greater cartographic generalisation), this is likely to be a greater challenge, although abstraction of the 2D geometry will also be more difficult due to the presence of street furniture, vegetation, cars and other objects. For this and other obvious reasons, increasing congruence by adding 3D realistic landmark representations to the map would not necessarily be the best solution: as decades of cartographic research suggest, along with more recent studies [35], a symbol merely representing the category of object (e.g., church or pub) may be recognised more quickly than an attempt at a photorealistic image of it. In any case, since appearances of real-world objects often change, it would probably be unrealistic to suggest that a mapping agency collect and maintain photorealistic images of the landmarks found on thousands of street corners.

Further experiments will investigate the strategies used under these more realistic circumstances, and the implications for the design of suitable map representations.

6 Acknowledgements

The authors wish to thank Claire Cannon and Jon Gould for their help in analysing and interpreting the data, Alasdair Turner at UCL for help in extracting the space syntax measures, Isabel Sargent, Jon Horgan and Dave Capstick (creators of the 3D building model), Guy Heathcote and Tim Martin for help in using the model and mapping to create the stimuli, the experiment participants for their time and cooperation, Glenn Gunzelmann and Glen Hart for insights and inspiration, and Ordnance Survey of Great Britain for funding and supporting this work.

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

1. Robinson, A., Petchenik, B.: The Nature of Maps: Essays towards understanding maps and mapping. University of Chicago Press, Chicago (1976)

2. Montello, D.R., Freundschuh, S.: Cognition of geographic information. In McMaster, R.B., Usery, E.L., eds.: A research agenda for geographic information science. CRC Press, Boca Raton, FL (2005) 61–91
3. Thompson, W.B., Pick, H.L., Bennett, B.H., Heinrichs, M.R., Savitt, S.L., Smith, K.: Map based localization: The 'drop-off' problem. In: Proceedings of the DARPA Image Understanding Workshop. (1990) 706–719
4. Pick, H.L., Heinrichs, M.R., Montello, D.R., Smith, K., Sullivan, C.N.: Topographic map reading. In Hancock, P.A., Flach, J., Caird, J., Vicente, K., eds.: Local applications of the ecological approach to human-machine systems. Volume 2. Lawrence Erlbaum, Hillsdale, NJ (1995) 255–284
5. Levine, M.: You-are-here maps: Psychological considerations. Environment and Behavior **14** (1982) 221–237
6. Klippel, A., Freksa, C., Winter, S.: You-are-here maps in emergencies – the danger of getting lost. Journal of Spatial Science **51** (2006) 117–131
7. Griffin, T.L.C., Lock, B.F.: The perceptual problem of contour interpretation. The Cartographic Journal **16** (1979) 61–71
8. Eley, M.G.: Determining the shapes of land surfaces from topographical maps. Ergonomics **31** (1988) 355–376
9. Aretz, A.: Spatial cognition and navigation. In: Perspectives: Proceedings of the 33rd annual meeting of the Human Factors Society, Santa Monica, CA, Human Factors Society (1989) 8–12
10. Harwood, K.: Cognitive perspectives on map displays for helicopter flight. In: Perspectives: Proceedings of the 33rd annual meeting of the Human Factors Society, Santa Monica, CA, Human Factors Society (1989) 13–17
11. Wickens, C.D., Prevett, T.T.: Exploring the dimensions of egocentricity in aircraft navigation displays. Journal of Experimental Psychology: Applied **1** (1995) 110–135
12. Gunzelmann, G., Anderson, J.R., Douglass, S.: Orientation tasks with multiple views of space: Strategies and performance. Spatial Cognition and Computation **4** (2004) 209–256
13. Warren, D.H., Rossano, M.J., Wear, T.D.: Perception of map-environment correspondence: The roles of features and alignment. Ecological Psychology **2** (1990) 131–150
14. Blades, M., Spencer, C.: The development of 3- to 6-year-olds' map using ability: The relative importance of landmarks and map alignment. Journal of Genetic Psychology **151** (1990) 181–194
15. Hagen, M.A., Giorgi, R.: Where's the camera? Ecological Psychology **5** (1993) 65–84
16. Presson, C.C.: The development of map-reading skills. Child Development **53** (1982) 196–199
17. Meilinger, T., Hölscher, C., Büchner, S.J., Brösamle, M.: How much information do you need? Schematic maps in wayfinding and self localization. In: Spatial Cognition '06, Bremen, Germany, Sept 2006., Berlin, Springer (2007)
18. Bryant, K.J.: Geographical spatial orientation ability within real-world and simulated large-scale environments. Multivariate Behavioral Research **26** (1991) 109–136
19. Gunzelmann, G., Anderson, J.R.: Location matters: Why target location impacts performance in orientation tasks. Memory and Cognition **34** (2006) 41–59
20. Davies, C.: When is a map not a map? Task and language in spatial interpretations with digital map displays. Applied Cognitive Psychology **16** (2002) 273–285

21. Hermer, L., Spelke, E.S.: A geometric process for spatial reorientation in young children. Nature **370** (1994) 57–59
22. Lansdale, M.: Modelling memory for absolute location. Psychological Review **105** (1998) 351–378
23. Cheng, K., Newcombe, N.S.: Is there a geometric module for spatial orientation? squaring theory and evidence. Psychonomic Bulletin & Review **12** (2005) 1–23
24. Winter, S.: Route adaptive selection of salient features. In Kuhn, W., Worboys, M., Timpf, S., Hartmanis, J., van Leeuwen, J., eds.: COSIT 2003, Spatial Information Theory: FOundations of geographic information science, Berlin, Springer (2003) 349–361
25. Shepard, R.N., Hurwitz, S.: Upward direction, mental rotation, and discrimination of left and right turns in maps. Cognition **18** (1984) 161–193
26. Benedikt, M.L.: To take hold of space: Isovist and isovist fields. Environment And Planning B: Planning & Design **6** (1979) 47–65
27. Conroy-Dalton, R.: Spatial navigation in immersive virtual environments. PhD thesis, University of London, London (2001)
28. Wiener, J.M., Franz, G.: Isovists as means to predict spatial experience and behaviour. In: Spatial Cognition IV: Reasoning, Action, Interaction, Berlin, Springer (2004) 42–57
29. Conroy-Dalton, R., Bafna, S.: The syntactical image of the city: A reciprocal definition of spatial elements and spatial syntaxes. In: 4th International Space Syntax Symposium, London, UK. (2003)
30. Hintzman, D.L., O'Dell, C.S., Arndt, D.R.: Orientation in cognitive maps. Cognitive Psychology **13** (1981) 149–206
31. Davies, C., Mora, R., Peebles, D.: Isovists for orientation: Can space syntax help us predict directional confusion? In: Proceedings of the Space Syntax and Spatial Cognition workshop, Spatial Cognition 2006, Bremen, Germany (2006)
32. Ward, J.H.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association **58** (1963) 236–244
33. Duda, R.O., Hart, P.E.: Pattern classification and scene analysis. Wiley, New York (1973)
34. Gombrich, E.H.: The image and the eye: Further studies in the psychology of pictorial representation. Phaidon Press, London (1982)
35. Smallman, H.S., John, M.S., Oonk, H.M., Cowen, M.B.: When beauty is only skin deep: 3D realistic icons are harder to identify than conventional 2D military symbols. In: Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA, Human Factors and Ergonomics Society (2000) 480–483