A Process Model of Map-based Orientation in Urban Environments

David Peebles
Department of Behavioural Sciences,
University of Huddersfield,
Queensgate, Huddersfield HD1 3DH, UK

Clare Davies
Ordnance Survey Research C530,
Romsey Road, Southampton, SO16 4GU, UK

January 20, 2009

Abstract

This study describes how people identify the orientation of an urban viewpoint on a street map in terms of the processes and strategies they use and the information required. The analysis also attempts to identify the cognitive and environmental factors that influence each step in the process and specify how these factors affect performance.

Introduction

Orienting oneself in an environment is one of the key tasks people undertake with maps. In order to navigate through an environment with a map, one must first align one's direction of sight in the environment with a location and direction on the map. This type of orientation task is complex as it requires knowledge of (or the ability to infer) the 2D representational features of the map that correspond to the 3D visible features in the environment and for people to interact with both an *egocentric* (i.e., 'how it looks to me') and *allocentric* (i.e., 'how it looks from above') representation of the environment.

Understanding the cognitive and perceptual mechanisms underlying this task and the environmental and cartographic factors that affect people's ability to carry it out successfully is important for several reasons. First, it may be possible to identify constraints on performance imposed by the human cognition and perception systems (such as limited working memory capacity or visual attention mechanisms). In addition, it may be possible to characterise one or more key strategies that people adopt and to identify those that improve performance. Second, it may be possible to uncover features of the map that facilitate or hinder performance and relate these to psychological principles such as *visual salience* or *similarity.* These findings may then inform design decisions or suggest ways in which people may be advised to use maps more effectively.

Furthermore, at the start of this project it was hoped that if those features and principles could be captured via spatial measurements or analyses, it might be possible to predict points in space where the orientation task would be hardest and hence where some cartographic intervention, such as inclusion of specific landmarks on a map at these points, might help users to avoid errors. In the longer term, as automated generalisation becomes easier, a computational cognitive model could eventually combine with other generalisation processes to automate the production of mapping designed explicitly for use in orientation or other scene-matching tasks. From Ordnance Survey's perspective this could lead to a superior but cheaply produced data product, optimised for these key uses of mapping.

Aims of the project

The primary aim of this project was to investigate people's orientation strategies when using Ordnance Survey maps in a (simulated) urban environment. We sought to explain how people identify the orientation of an urban viewpoint on a street map in terms of the processes and strategies they use and the information required. The analysis also attempted to identify the cognitive and environmental factors that influence each step in the process and specify how these factors affect performance.

Because the study was primarily interested in orientation within urban settings, we also sought to investigate whether the concepts and analytical methods of *space syntax* theory may be applied to explain our findings. Space syntax is a conceptual and mathematical framework for analysing spaces and is particularly employed in the analysis of wayfinding for the design of urban spaces and buildings. A key concept in space syntax theory is the *isovist*, defined as the volume of space visible from a specific point in that space. A two dimensional isovist may also be defined as a horizontal section of the volume, perhaps best conceived as an eye-level slice through the visible space as a person rotates 360°. By analysing the properties of these 3D and 2D isovists, researchers have been able to quantify the accessibility and navigability of various spaces and to investigate the relationship between spatial layouts and aspects of human behaviour such as traffic flow and the incidence of crime. It has previously been suggested that isovist analysis may have relevance to urban orientation because the shape and size of the isovist may affect people's ability to deduce direction and position (Conroy-Dalton & Bafna, 2003). We aimed to test this proposal by applying isovist-based space syntax analyses to our task environments to determine whether they correlated in any significant way with our behavioural measures.

The investigation took the form of two experiments that recorded several behavioural measures, including 'end of task' measures such as task completion time and judgement accuracy and more sophisticated and rich 'during task' measures such as eye movements and verbal 'think aloud' reports. The completion time and judgement accuracy measures can be aggregated over multiple experiment participants to provide a global analysis of how people carry out the task. In contrast, the eye movement and verbal protocol data allowed us to generate a more detailed analysis of the cognitive, perceptual and strategic processes undertaken by individuals.

From an analysis of previous studies into map-based orientation, we developed a model of orienting behaviour that characterises the interaction between internal mental and environmental factors that determine success or failure in the map-based orientation task.

Models of orientation

There have been several previous attempts to study orientation behaviour in different environments using different graphical representations. These studies vary considerably in terms of the nature and complexity of the environment and map used, the precise nature of the task and the level of the subsequent analysis. Below we review three studies which, although quite dissimilar from the task and environment used in our study, provide useful theoretical insights into the information requirements and problem solving strategies employed that may form the basis of our analysis.

Pick, Heinrichs, Montello, Smith, Sullivan & Thompson (1995) conducted a study of a *drop-off localization* or "where am I?" task in a rural environment using a contour-only map to determine how the problem can be solved using only topographic information. From an analysis of verbal protocols, the authors proposed four main processes that their participants used to solve the problem: (a) Reconnaissance, in which features, attributes and relations on the terrain or map were identified for subsequent processing, (b) Feature matching, whereby features in the terrain were matched to features on the map and vice versa, (c) Viewpoint hypotheses, the generation, comparison and evaluation of hypotheses about the viewpoint, and (d) Conclusion, selecting one of the alternative hypotheses.

The verbal protocols also revealed that people who solved the problem successfully used four key strategies: (a) they focussed their initial reconnaissance on the terrain rather than the map, (b) they organised terrain features into configurations, (c) they focussed on terrain features close to the viewpoint rather than distant ones, and (d) they generated and evaluated multiple hypotheses. Two additional strategies used by some successful participants were: (e) to compare and test hypotheses using a disconfirmation procedure to avoid confirmation bias, and (f) to change viewpoint in order to access more terrain features and test hypotheses.

Warren, Rossano and Wear (1990) investigated the effects of feature salience and the alignment of the scene and map in a study in which participants viewed simplified geometric images of buildings and corresponding floor plans and were required to decide the location on the floor plan from which the image of the building was taken. Consistent with other studies (e.g., Gunzelmann and Anderson, 2006; Hintzman, O'Dell, & Arndt, 1981), they found that performance in the task improved with closer alignment between scene and map.

They proposed a model of the task that explained successful participants' performance in terms of an algorithm, the key steps of which were to (a) identify a feature or configuration of features on the building, (b) find the feature or configuration on the map and (c) hypothesise the point on the map representing the location from which the visible configuration would be viewed. The algorithm they proposed was an iterative process to reduce ambiguity in which additional features of the environment were sampled to successively narrow down the possible choices, using the most salient features of the scene first.

More recently Gunzelmann and his colleagues (Gunzelmann, 2008, Gunzelmann & Anderson, 2006; Gunzelmann, Anderson, & Douglass, 2004) have investigated map-based orientation using a paradigm in which participants were required to identify on a map a specific target cone from a number of similar cones scattered in a circular arena, viewed from outside the arena.

They developed a model of the task which explains human performance in their experiment as a hierarchical process to finding the target on the scene. In this process, a cluster containing the target is first identified and the number of objects and their location relative to the viewpoint (left, right, straight ahead) encoded. This involves the production of a verbal description of relative positions of the objects. Then the target within the cluster is identified in terms of its position from left or right and how close it is relative to the other objects in the group. The model then uses this verbal description to identify the object in the map as this contains sufficient information to do so.

Gunzelmann and Anderson's model is similar to that of Pick et al (1995) in that it characterises the process as one involving the identification of groupings or specific landmark features, and then locating the target relative to that. Their model is limited in that it does not include any mechanism for mental rotation. However it does capture the key characteristics of human performance in the task, such as improvements with map/scene alignment and the slowing of performance with increases in the number of nearby distractors.

A process model of orientation

Although none of these three studies investigates orientation behaviour in urban environments using realistic maps and the model of orientation each proposes is specific to their individual task, they all provide important insights into how people carry out such tasks in other settings. Of the three models discussed above, that of Pick et al. provides the most specific and detailed analysis of the strategic processes involved. The insights provided by Warren, Rossano and Wear are close to those of Pick et al. in that they suggest that orientation should be initiated by scanning the environment rather than the map. In addition, both suggest that orientation proceeds by matching features from the scene and map, producing one or more hypotheses and then testing them by sampling additional features and attempting to match them until sufficient evidence is accumulated to produce a decision. Of the three models, that of Gunzelmann is perhaps least applicable to the scenario we wish to investigate because the tasks are so different. In common with the other two models however, it suggests that people use a strategy involving the selection of salient features to match.

We have attempted to distil the key features of the three previous models into a single general model of orientation that may be applicable to the scenario we intend to investigate. Figure 1 shows a process diagram of our proposed model of the sequence of processes involved in carrying out the orientation task. The model consists of a number of key processes that are undertaken sequentially but which may be reiterated depending on the results of various decisions made during the course of the task. We now describe each of these steps in detail and outline the various environmental and cognitive factors that influence the processing in each.

Scan

The action of visually sampling the environment or map for individual features or configurations of features to serve as input for the matching process. Pick et al., (1995) found that successful participants in their study more often focussed their initial scan on the environment rather than the map. Their explanation for this is that, because the area covered by most maps is typically much larger than an individual viewable scene, many of the features on the map will be irrelevant to a specific localisation task. The viewable features in the scene however, are all potentially relevant to the current task and will have corresponding features on the map.

Select

The action of selecting individual features or configurations of features to be matched. Initially, which feature or configuration is selected and the time taken to do so are primarily affected by the perceptual salience of the various features (e.g., buildings, building clusters or street configurations) in the scene. Perceptual salience is typically characterised in terms of a feature's size, shape, colour or orientation relative to other nearby features with greater difference corresponding to greater salience and distinguishability. Problems related to the low salience or numerosity of features may be mitigated somewhat by mentally grouping features into configurations as this creates a more discriminable pattern to identify which reduces the ambiguity in the scene (Pick et al., 1995).

Initial feature selection may also, at least to some extent, be determined by the user's estimate of how each feature will be represented on the map, with distinctive shapes or subjectively easier predicted allocentric patterns being chosen first.

While visually sampling the scene, the user is generating a structural description of the various features relative to each other and the user's viewing position. Pick et al., (1995) found that successful participants in their study tended to organise features into configurations. This

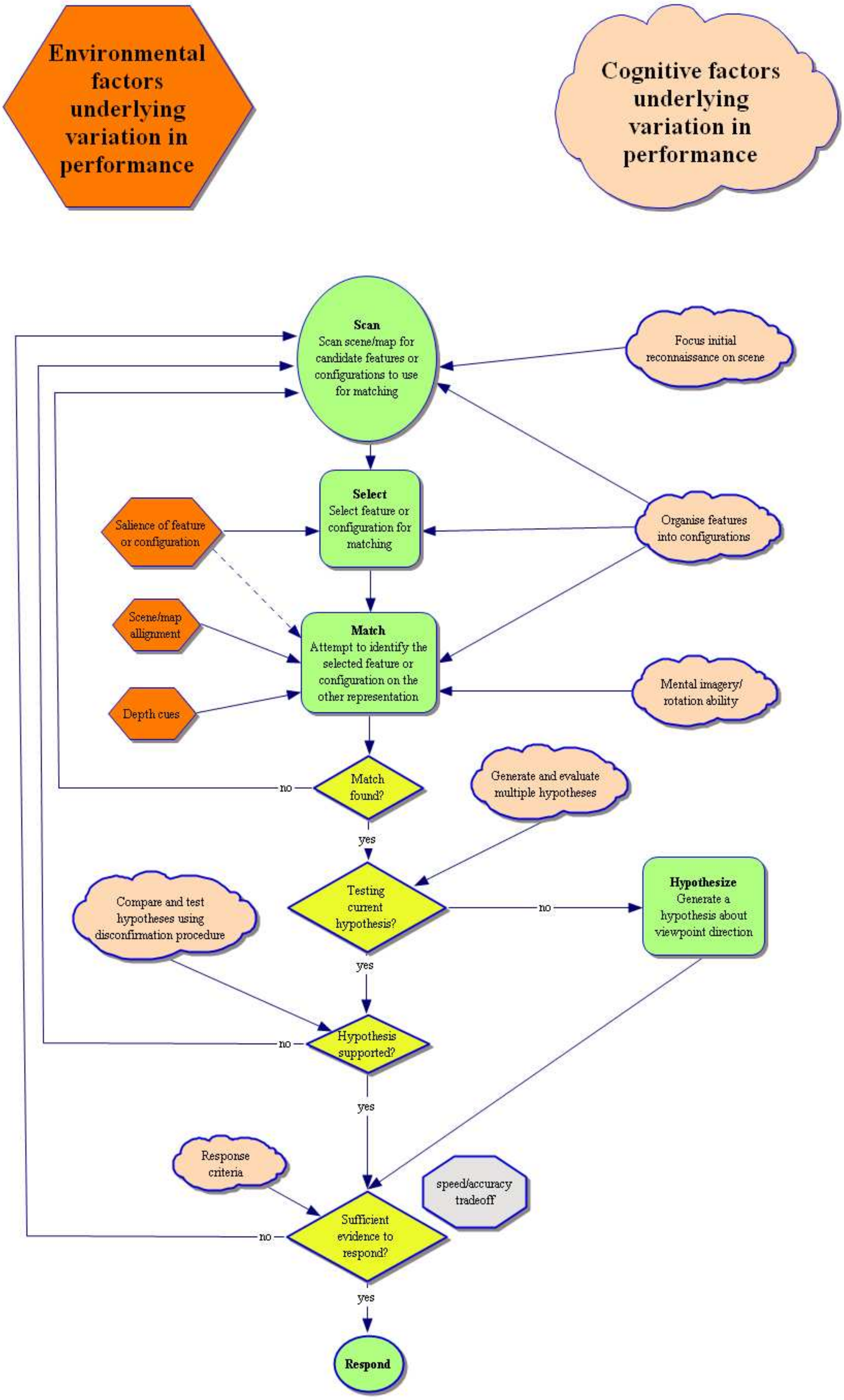

Figure 1: A processes model of orientation

decreases ambiguity by increasing the number of features involved in the matching process and adds the further constraint of encoding spatial relationships between features (e.g., "tall building with a smaller building on its left and a road running behind it"). Pick et al., (1995) note that such configurations are most effective when constructed along a line of sight as they form a linear order that constrains the search for the corresponding features on the map. For example, in Figure 2, the user may identify a small building in front of a larger rectangular building as the primary configuration to search for on the map while also encoding that s/he is looking across a path or road with a large open space in front of the small building etc.

Match

The matching process is the most cognitively demanding stage of the task and involves identifying correspondences between the two representations, one of which is in working memory, the other being viewed as a current visual stimulus (i.e., scene or map). The cognitive representation may be a mental image, a verbal description (such as the one above) or a combination of the two. If a mental image is involved matching may take place using a mental rotation process whereas if a purely verbal description is employed, the process involves attempting to identify one or more features that satisfy the description from an allocentric perspective (cf. Gunzelmann, 2008, Gunzelmann & Anderson, 2006). Other factors known to affect matching performance are the alignment between the scene view and the map (Gunzelmann & Anderson, 2006) and, from the results of Experiment 1, the number and quality of depth cues in the environment.

Hypothesize

The cyclic process of scanning the two representations attempting to identify matches between their features continues until the user identifies a correspondence. At this point a hypothesis is made about the viewpoint direction on the map that would imply the correspondence.

Once a hypothesis has been produced, participants have to decide whether to test it with additional evidence or not. This decision will depend on an individual's judgement whether the currently accumulated evidence is sufficient to justify making a decision based on some threshold or criterion. It will also depend therefore on the emphasis an individual places on response accuracy or speed. This typically results in a tradeoff between the two, with more accurate judgements taking a longer time to produce.

If the user decides to test the hypothesis this will be achieved by comparing additional features in the scene and map in the light of the expectations produced by the hypothesis. This involves going through the *scan*, *select*, and *match* steps again to either confirm or disconfirm the hypothesis. Pick et al. (1995) found that all of the successful participants in their experiment generated and evaluated multiple hypotheses and that a number of unsuccessful participants evaluated their hypotheses by attempting to confirm them with additional evidence rather than disconfirming them. This bias towards confirmation often led people to "explain away" unsupporting evidence rather than using it to reject the hypothesis.

Environmental factors underlying performance variance

According to this model, the primary environmental factor underlying task performance is the *salience* of features or configurations in the environment or map. This term covers a number of factors that affect the discriminability of features including visual similarity and numerosity. If there are many similar features or configurations in the environment, it is more difficult to select or identify one compared to environments containing relatively few or individually distinct ones.

This factor is encapsulated in the idea of the *landmark*, a particular feature or configuration that is either unique in the environment or so visually distinct as to provide a readily available point of reference.

Highly salient features can facilitate a relatively rapid and accurate orientation strategy if a landmark can be identified in both the scene and map. People can use the landmark as an orientation indicator and can match other items in the scene according to their position relative to it. Previous studies of orientation tasks (e.g., Gunzelmann & Anderson, 2006; Pick et al, 1995; Warren, Rossano, & Wear, 1990) have suggested that people do tend to use this landmark-based strategy, as opposed to an alternative *geometry-based* strategy of constructing a mental representation of the allocentric 2D ground layout from the 3D geometry of the scene from the observer's egocentric view. Geometry-based strategies have previously been suggested as explanations for orientation behaviour (e.g., Hermer & Spelke, 1994).

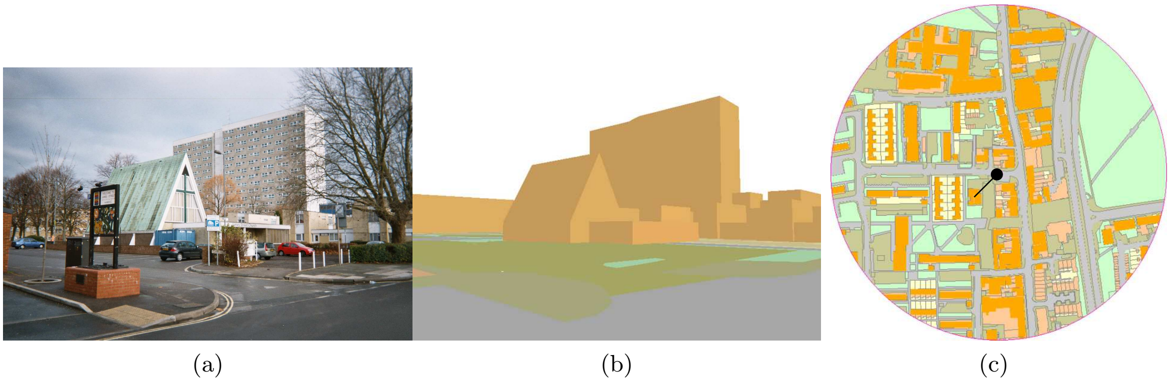
(a) (b) (c)

Figure 2: Example street location (a) scene (b) and corresponding map (c) from Experiment 1

Experiment 1

One possible criticism of the above model is that it is derived from studies which have used environments which were either extremely sparse, or limited to a view of a single building, both of which may promote a salient-feature matching strategy. In more dense, homogeneous, complex (and arguably more real-world) environments such as urban landscapes, this strategy may not be so easy and therefore people may be forced to adopt alternative strategies, for example studying the geometry of the scene and deriving from it a mental representation of the 2D shapes of the ground layout, as would be seen if viewed from above.

Unless the urban landscape is based on a block system (like many cities in the US), the 2D geometry of a city is almost always uniquely specified from any point in a typical urban space: it is rarely completely symmetrical and hence tends not to be ambiguous in terms of orientation.

We therefore decided to investigate orientation strategies where the scene people viewed was an image taken from a 3D model of a UK city, Southampton, with only the 2D ground layout and the 3D building shapes being shown. The scenes were shorn of irrelevant detail that would encourage the identification of specific landmarks for matching to the map, and the map in turn contained no name labels or other indicators to differentiate buildings and other objects. 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 individual item and attempting to match it to the map was unlikely to be successful, since the 2D geometry of one item would not be unambiguous (but the overall ground layout and relative positions of items would be).

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 other 49 participants were encouraged to perform the task as quickly and accurately as possible.

Materials

The stimuli were 25 scenes and corresponding maps from various locations in the city of Southampton, UK. The scene images 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 maps were circular sections of OS MasterMap® Topography Layer at 1:1250 scale. A black dot in the centre of the 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 Figure 2c). This rotated around the dot as the mouse was moved around the map so that it always pointed toward the mouse cursor.

Scenes and maps were selected 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 stimuli were also controlled for alignment so that the correct response ranged across the full 0–360 degree circle and there were roughly equal numbers of roughly north- and south-facing scenes.

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. Participants were instructed how to make a response, asked to respond as rapidly and as accurately as possible, and told that the maps were all the same scale and that they should avoid the natural assumption that the 'upwards' direction on the map indicates 'forward' in the environment.

There were five practice trials and 20 experiment trials in total. The degree of the response was recorded, from 0° pointing directly to the top of the map to 180° pointing directly to the bottom, with the sign of the angle indicating left (negative) or right (positive).

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.

Results

The results of the experiment are described in detail in published reports (Davies, Mora, & Peebles, 2006; Davies & Peebles, 2007; Peebles, Davies, & Mora, 2007) and so we simply provide a summary here.

In line with previous studies (e.g., Warren et al., 1990), 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° range that it bisected), at the point when the participant clicked the mouse. 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 40.94s ($SD = 9.76$). A Spearman's rho test produced a moderate but non-significant negative correlation between the probability of a correct response and latency, $\rho_s = -.410$, $p = .072$ 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.

Environmental factors affecting performance

In order to test whether performance was influenced by the presence of salient 3D landmarks or 2D geometry, the scenes were coded according to the presence or absence of these features. Both 3D landmarks and 2D geometry had a statistically significant effect on the number of errors participants made. The presence of an obvious 2D cue decreased error rates but only when a salient 3D cue was not present, in which case errors always greatly increased.

In terms of solution time, although both 3D and 2D cues slowed participants down, the presence of a 2D landmark increased response times slightly except when a 3D landmark was also present, in which case performance slowed down even more.

These findings 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.

Space syntax analysis

Two 2D isovists were computed for each of the maps in the experiment (examples of which are shown in Figure 3), the first for the full 360° visible space (drawn dark blue in Figure 3), the second for the 60° segment that was visible in the scene image (drawn lighter pink in Figure 3). The 60° isovist was created to determine whether it was sufficient because participants were specifically focusing on aspects within the geometry of the scene rather tat of the entire map.

We then computed a number of metrics[1] for the isovists that are commonly used in space syntax analyses which we identified as having a potential role in people's performance of the task. These included the area and perimeter length of the isovist, its minimum and maximum radius, and five additional measures of its geometry: *occlusivity* (the extent to which some features of the local environment are hidden by others), *compactness* (the nearness of the isovist shape to a circle), *jaggedness* (the complexity of the isovist's shape), *drift magnitude* and *drift angle* (the distance and angular distance respectively of the viewer's location from the isovist's centre of gravity). The last two measures indicate the level of asymmetry in the isovist, with the assumption that more asymmetrical isovists being easier to match unambiguously to a map.

---

[1] These metrics were computed by Rodrigo Mora at University College London as part of his PhD thesis work, funded by Ordnance Survey.

A multiple regression analysis applied to the error data provided no evidence to support the proposition that any of the space syntax measures had an effect on participants' performance across the sample as a whole.

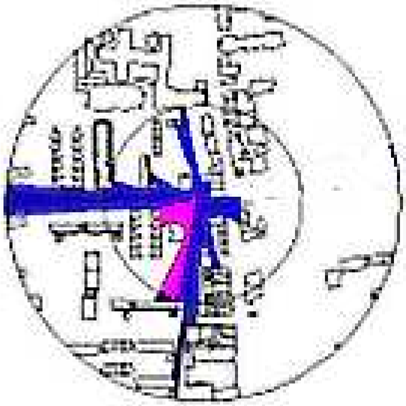

Figure 3: Isovist at ground level for the scene in Figure 2 showing the full 360° version (dark blue) and the 60° segment (lighter pink) visible in the scene image.

Cognitive factors affecting performance

Having identified the effects on performance of some of the environmental factors identified in the model and eliminated the possibility of others, we then looked to determine whether cognitive factors, specifically individual differences in strategy, could be associated with any of the spatial metrics of the scenes. The results of this analysis are presented in full in Peebles et al., (2007) and are summarised below. Four main strategies were identified. The first, identified in 41% of participants, involved selecting a single feature to match, most likely from the 2D ground layout, since they were faster with more open spaces. These participants were usually distracted into trying to use a salient 3D landmark when they were present and focused little on the isovist shape. The second strategy, identified in 29% of participants, involved abstracting the 2D isovist geometry and then matching it to the map. As with the first category, these participants were distracted into trying to use salient 3D landmarks when it was present. The third strategy, identified in 18% of participants, involved focussing on the street pattern in built-up areas. This was enhanced when streets were lined with buildings making their shapes more salient. This strategy required mental rotation to match the scene to the map and was therefore worsened by scene/map misalignment. The final strategy, identified in 12% of participants, involved the use of both ground layout patterns and abstracting the overall isovist geometry. (the latter being easier when the isovist was smaller and more clearly defined by buildings). These participants were highly distracted by attempt to match foreground 2D cues when it was available however.

Discussion

The various implications of the experiment are discussed in detail in the published reports listed above. A summary of the main implications is presented below.

- The results of of the experiment are consistent with previous studies in that it found that people typically attempt to identify a salient landmark, particularly one with a distinctive 3D 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 used in the experiment, because the 2D shapes and colours in the scenes were directly matchable to features in 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.
- Almost half of participants did show a reliance on single salient landmarks 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 did attempt to abstract the 2D isovist geometry, the (simpler) street pattern, or the patterns made by ground layout features.
- Although the analysis 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.
- 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.

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. 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.

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 discernible 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, 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 may be a while yet before new imagery/video data capture techniques permit efficient automated collection and update of a national street-corner landmark database.

One question remains concerning the nature of the stimuli used in the experiment. Although the scenes were deliberately used to remove all distracting salient cues apart from the 3D geometry in order to provide a relatively a 'pure' 3D space for the study matching, this level of abstraction, together with the matched colours between the scene and map may have resulted in orientation behaviour that would not necessarily be found with real-world scenes or photographs. To investigate this further we conducted a second follow up experiment using photographs of the scenes instead of abstract 3D models.

Experiment 2

Design and Participants

Thirty-five students and members of staff from the University of Huddersfield took part in the experiment. The design of the experiment was exactly the same as Experiment 1 in all but two details. The first was that the scenes participants saw were actual photographs rather than abstract 3D models. The second difference was that participants were presented with twenty additional scenes (from the same areas around Southampton) after seeing the original scenes from the first experiment.

Procedure

The experiment was conducted in the same way as the first. As an additional measure in this experiment, after participants had finished the experiment, they were asked to complete the *The Santa Barbara Sense of Direction Scale* (SBSOD; Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002), shown in Appendix A. This measure requires participants to rate their level of agreement on a seven point scale to 15 statements concerning their spatial abilities. Previous studies have shown the SBSOD to predict objective measures of environmental spatial abilities quite highly (Hegarty et al., 2002). We also added one further question that asked participants to rate on a ten point scale their experience of matching maps to real-world scenes (also shown in Appendix A).

Results

As with Experiment 1, responses were scored as correct if the angle of the response line fell within 15 degrees of the true angle in either direction. As the data from Experiment 2 have not been published at this point in time, the proportion of correct responses and mean response times for scenes are displayed in Table 1. The pattern of responses was similar to that found in Experiment 1, with participants in general performing the task reasonably accurately, with the proportion of correct responses for each stimulus ranging from .09 to 1.0.

For this experiment the main focus of our analysis was an attempt to identify the primary factors underlying errors. As with Experiment 1, incorrect responses typically clustered around specific candidate feature types more or less related to the target location. We analysed these errors by counting the responses for each scene and for clusters containing more than one response, identifying a number of potential (non mutually exclusive) causes of the errors that were applicable across the range of stimuli. We then independently coded these erroneous responses according to the potential causes, retaining only those in which we achieved a level of agreement at 65% or above. The four main factors we identified are listed below.

Table 1: Proportion of Correct Responses and Mean Response Time (in seconds) for scenes, Experiment 2. Standard Deviations are in Brackets.

Original Scenes			New Scenes		
Scene No.	Correct Response	Response Time	Scene No.	Correct Response	Response Time
1	0.74 (0.44)	32.18 (26.10)	21	0.83 (0.38)	16.21 (6.36)
2	0.74 (0.44)	23.48 (15.22)	22	0.34 (0.48)	20.91 (18.25)
3	0.57 (0.50)	32.27 (17.86)	23	0.43 (0.50)	28.14 (15.81)
4	0.37 (0.49)	24.78 (12.75)	24	0.26 (0.44)	30.25 (26.94)
5	0.60 (0.50)	21.05 (14.85)	25	0.80 (0.41)	18.16 (12.17)
6	0.63 (0.49)	25.16 (14.05)	26	0.77 (0.43)	15.91 (6.99)
7	0.60 (0.50)	37.03 (23.04)	27	0.57 (0.50)	22.09 (13.76)
8	0.69 (0.47)	26.78 (17.64)	28	0.77 (0.43)	11.04 (6.96)
9	0.89 (0.32)	29.78 (19.03)	29	0.71 (0.46)	25.13 (15.41)
10	0.51 (0.51)	35.13 (22.93)	30	1.00 (0.00)	9.97 (5.76)
11	0.29 (0.46)	29.63 (14.54)	31	0.66 (0.48)	23.17 (11.62)
12	0.40 (0.50)	36.12 (26.33)	32	0.29 (0.46)	21.10 (15.00)
13	0.34 (0.48)	32.13 (19.67)	33	0.14 (0.36)	21.59 (11.91)
14	0.20 (0.41)	29.55 (22.81)	34	0.89 (0.32)	18.22 (11.56)
15	0.11 (0.32)	32.17 (26.21)	35	0.80 (0.41)	18.44 (11.15)
16	0.29 (0.46)	34.51 (27.84)	36	0.54 (0.51)	28.84 (25.33)
17	0.11 (0.32)	28.91 (16.68)	37	0.69 (0.47)	23.19 (13.28)
18	0.31 (0.47)	27.59 (16.12)	38	0.63 (0.49)	28.56 (20.81)
19	0.09 (0.28)	48.07 (40.49)	39	0.57 (0.50)	26.63 (16.90)
20	0.66 (0.48)	28.04 (31.02)	40	0.57 (0.50)	17.50 (10.19)

Scene Object Salience (SOS)

This occurs when there is a very visually salient object in the scene (e.g., a tall or distinctive building) that is not unambiguous on the map. This was implicated, on average, in 33.2% of erroneous responses in 10 scenes (inter-rater agreement: 97%). An example of this factor is shown in Figure 4b which reveals that the five main erroneous responses (red lines) totalling 36% of the responses resulted from participants identifying the salient blue building in the scene but failing to identify it correctly on the map. All of the erroneous responses indicated that alternative buildings had been identified instead.

Missed Ground-level Cue (MGC)

This occurs when there is an unambiguous cue at ground level in the scene (e.g., a traffic calming chicane, lawn or pavement) that is ignored. This may be due to strategic factors whereby participant chose to focus on higher level building features or external factors relating to the salience of the ground level features in the map. This was implicated, on average, in 25.1% of erroneous responses in 32 scenes (inter-rater agreement: 80%). An example of this factor is shown in Figure 4d which reveals that the three main erroneous responses (totalling 40% of the responses) identified a road on the map but did not notice the pattern of pathways on either side of the road (or the left turning road immediately in front of the viewing point).

(a) Scene 31 (b) Map 31

(c) Scene 38 (d) Map 38

(e) Scene 7 (f) Map 7

Figure 4: Scene and maps from Experiment 2 illustrating correct (black lines) and incorrect (red lines) responses. Errors on Map 31 are caused by *scene object salience* whereas those on Map 38 are caused by *missed ground-level cues*. Map 7 illustrates two errors. The north pointing line is caused by *left/right reversal* while the other two are caused by *misperceived object distance*.

Misperceived Object Distance (MOD)

This occurs when the incorrect response direction would be more accurate if an object in the scene is nearer or further away. This was implicated, on average, in 22.4% of erroneous responses in 22 scenes (inter-rater agreement: 71%). Examples of this factor are shown by the east and south pointing red lines in Figure 4f (totalling 29% of the responses) which reveal that people identified the view as lying between two buildings but misjudged the distances of at least one of the buildings from the viewpoint.

Left/Right Reversal (LRR)

This occurs when the incorrect response would be more accurate if the scene is left/right reversed. This was implicated, on average, in 17.2% of erroneous responses in 26 scenes (inter-rater agreement: 65%). An example of this factor is also shown by the north pointing red line in Figure 4f (6% of the responses), resulting from the fact that the object configuration in this direction is the mirror image of what is viewed in the scene.

Implications of these results

It is important to determine whether these four typical error types fit in to the process model proposed earlier and, if they do, where in the process they occur. Revisiting the model diagram in Figure 1, it would seem that all of the factors have their effect early on in the process, specifically in the *select* and *match* stages. The *Scene Object Salience* factor is intrinsically environmental in nature and affects the selection of features for processing. The *Missed Ground-level Cue* factor also has its effects at this stage but may be related to the scene, map or strategic choice, or some combination of all three. It may be possible to reduce this type of error by increasing the salience of such ground level cues, perhaps by making their colour more distinct. This would not only increase their visual salience but may also increase the likelihood that users may choose to use them when orienting themselves. The *Misperceived Object Distance* and *Left/Right Reversal* factors are more cognitive in nature and it is unlikely to be possible to reduce these errors by making a modification to the map.

A key implication of this research is that variation in performance in urban orientation can be adequately accounted for in terms of a relatively small number of basic cognitive and environmental factors in the early stages of the task. Other spatial measures such as the isovists of space syntax analysis were less useful in predicting orientation performance across the whole task (although it may be the case that isovist analysis would be relevant to certain stages, e.g., to the ease of finding suitable features to match, or to the frequency of left/right errors).

The process model developed during this project therefore provides a more detailed account of this task than previous models and points to the key stages and processes where human performance breaks down. This level of detail in a process model is about at the limit of what is currently possible to implement. In the original plan for the project we had envisioned modelling the task within a computational *cognitive architecture* such as ACT-R (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004). Such architectures embody general theories of cognition, perception and motor control as computer programming systems allowing researchers to implement cognitive models as runnable computer programs. An ACT-R model of an orientation task has been produced (Gunzelmann, 2008; Gunzelmann, & Anderson, 2006; Gunzelmann et al., 2004) and was one of the inspirations for this project. During the course of the project however, we came to the realisation that the task environment that we were investigating was much more complex than that in the earlier study and that the mechanisms and representations in ACT-R are currently not sophisticated enough to capture the key environmental and cognitive factors we were interested in. Specifically, ACT-R is currently unable to interact with a complex environment and does not include the ability to model people's fallible spatial processes (e.g., mental rotation) realistically. Our study has shown however, for example by revealing participants' left-right errors, that such fallibility does need to be represented in any predictive model of the orientation task.

A secondary aim of the project was to determine whether the spatial measures provided by space syntax analysis may be incorporated into our model to allow us to predict orientation performance at different points in a given urban area from the its 2D mapping. The results of our study indicate however that such spatial measures may be less important than visual measures of the scene itself (which of course are not currently easy for Ordnance Survey to acquire), although future work with video capture and 3D modelling may enable it.

Conclusions

To conclude, this programme of work has greatly furthered our understanding of the nuances of the orientation task, and the sources of people's fallibility when making errors in it. In particular, we have demonstrated people's tendency to seize on a visually salient object in the scene, even where the 2D geometry of the scene would be a less ambiguous and error-prone cue to match to the map. This implies that the best cartographic innovation to help people with this task would be independent of the 2D geometry of any given wayfinding point. Instead, it would involve providing a symbol at every such point that could be instantly and unambiguously matched to the most salient landmark within the visible scene.

However, we have also investigated the feasibility of a predictive computational model of people's performance on the orientation task. This could allow us to identify the most problematic decision points, so as to allow for a more selective provision of landmarks wherever cartographic space/clutter was at a premium, such as on maps for mobile devices. Through our analysis of common error types, we have established that such a model would again need to take its inputs from the visual scene rather than from spatial isovist measures derived from 2D map data—although these do have some predictive relevance whenever people use a geometric strategy to solve the task. Through our process model we have shown what would need to be specified in a computational implementation.

At present it remains technologically infeasible both to build such a model with current cognitive architectures (because they do not simulate people's fallible spatial processes), and to collect and maintain a database of ground-level scene images for the whole country to use as inputs to it (and perhaps to actually show to wayfinders within the map or navigation system, and in other applications). However, as technology advances this may become feasible and desirable within the next three to five years.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y . (2004). An integrated theory of the mind. *Psychological Review, 111.* 1036–1060.

Conroy-Dalton, R., & Bafna, S. (2003). The syntactical image of the city: A reciprocal definition of spatial elements and spatial syntaxes. In *4th International Space Syntax Symposium*, London, UK.

Davies, C., Mora, R. & Peebles, D. (2006). Isovists for Orientation: Can space syntax help us predict directional confusion? Proceedings of the 'Space Syntax and Spatial Cognition' workshop, Spatial Cognition 2006, Bremen, Germany, 24 September 2006.

Davies, C., & Peebles, D. (2007). Strategies for orientation: The role of 3D landmark salience and map alignment. In Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum.

Gunzelmann, G. (2008). Strategy generalization across orientation tasks: Testing a computational cognitive model. *Cognitive Science*, *32(5)*, 835–861.

Gunzelmann, G., & Anderson, J. R. (2006). Location matters: Why target location impacts performance in orientation tasks. *Memory & Cognition*, *34(1)*, 41–59.

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

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.

Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology*, *13*, 149–206.

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.* 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. Flach, J. Caird, & K. Vicente (Eds.), *Local Applications of the Ecological Approach to Human-Machine Systems*, *Vol. 2*, pp. 255–284. Hillsdale, NJ: Lawrence Erlbaum.

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

Appendix A. The Santa Barbara Sense of Direction Scale

This questionnaire consists of several statements about your spatial and navigational abilities, preferences, and experiences. After each statement, you should circle a number to indicate your level of agreement with the statement. Circle “1” if you strongly agree that the statement applies to you, “7” if you strongly disagree, or some number in between if your agreement is intermediate. Circle “4” if you neither agree nor disagree.

1. I am very good at giving directions.
   strongly agree 1 2 3 4 5 6 7 strongly disagree

2. I have a poor memory for where I left things.
   strongly agree 1 2 3 4 5 6 7 strongly disagree

3. I am very good at judging distances.
   strongly agree 1 2 3 4 5 6 7 strongly disagree

4. My “sense of direction” is very good.
   strongly agree 1 2 3 4 5 6 7 strongly disagree

5. I tend to think of my environment in terms of cardinal directions (N, S, E, W).
   strongly agree 1 2 3 4 5 6 7 strongly disagree

6. I very easily get lost in a new city.
   strongly agree 1 2 3 4 5 6 7 strongly disagree

7. I enjoy reading maps.
   strongly agree 1 2 3 4 5 6 7 strongly disagree

8. I have trouble understanding directions.
   strongly agree 1 2 3 4 5 6 7 strongly disagree

9. I am very good at reading maps.
   strongly agree 1 2 3 4 5 6 7 strongly disagree

10. I don’t remember routes very well while riding as a passenger in a car.
    strongly agree 1 2 3 4 5 6 7 strongly disagree

11. I don’t enjoy giving directions.
    strongly agree 1 2 3 4 5 6 7 strongly disagree

12. It’s not important to me to know where I am.
    strongly agree 1 2 3 4 5 6 7 strongly disagree

13. I usually let someone else do the navigational planning for long trips.
    strongly agree 1 2 3 4 5 6 7 strongly disagree

14. I can usually remember a new route after I have traveled it only once.
    strongly agree 1 2 3 4 5 6 7 strongly disagree

15. I don’t have a very good “mental map” of my environment.
    strongly agree 1 2 3 4 5 6 7 strongly disagree

Finally, on a scale of 1 to 10 (where 1 is the least experience), how much experience have you had of matching maps to real-world scenes, e.g., when navigating?