An Approach to Pathfinding for Real-World Situations

by

Sarah Cook

**Submitted in accordance with the requirements
for the degree of Doctor of Philosophy.**

**The University of Leeds
School of Computing**

February 2018

The candidate confirms that the work submitted is his own and that the appropriate credit has been given where reference has been made to the work of others.

Abstract

People plan their routes through new environments every day, but what factors influence these wayfinding decisions? In a world increasingly dependent on electronic navigation assistance devices, finding a way of automatically selecting routes suitable for pedestrian travel is an important challenge. With a greater freedom of movement than vehicular transport, and different requirements, an alternative approach should be taken to find an answer for pedestrian journeys than those taken in cars. Although previous research has produced a number of pedestrian route recommendation systems, the majority of these are restricted to a single route type or user group. The aim of this research was to develop an approach to route suggestion which could recommend routes according to the type of journey (everyday, leisure or tourist) a person is making.

To achieve this aim, four areas of research were undertaken.

Firstly, six experiments containing 450 participants were used to investigate the preference of seven different environment and route attributes (length, turns, decision points, vegetation, land use, dwellings and points of interest) for two attribute categories (simplicity and attractiveness) and three journey types (everyday, leisure and tourist). These empirically determined preferences were then used to find the rank-orders of the attributes, by comparing more of them simultaneously than earlier studies, and found either new rankings (for attractiveness, leisure journeys and tourist journey) or extended those already known (everyday journeys).

Using these ranks and previously accepted relationships, an environment model was defined and built based on an annotated graph. This model can be built automatically from OpenStreetMap data, and is therefore simple enough to be applicable to many geographical areas, but it is detailed enough to allow route selection.

Algorithms based on an extended version of Dijkstra's shortest path algorithm were constructed. These used weighted minimum cost functions linked with attribute ranks, to select routes for different journey types. By avoiding the computational complexity of previous approaches, these algorithms could potentially be widely used in a variety of different platforms, and extended for different groups of users.

Finally, the routes suggested by the algorithms were compared to participant recommendations for 'simple' routes with five start/end points, and for each of the three journey types (everyday, leisure and tourist). These comparisons determined that only length is required to select simple and everyday routes, but that the multi-attribute cost functions developed for leisure and tourist journeys select routes that are similar to those chosen by the paricipants. This indicates that the algorithms' routes are appropriate for people to use in leisure and tourist journeys.

Acknowledgements

Declarations

Some parts of the work presented in this thesis have been published in the following article:

Cook, S. and Ruddle, R. A. Effect of Simplicity and Attractiveness on Route Selection for Different Journey Types. In C. Freksa, B. Nebel, M. Hegarty, and T. Barkowsky (Eds.), *Spatial Cognition IX* (Vol. 8684, pp. 190–205, 2014). Cham: Springer International Publishing.

Contents

List of Figures

xiv

List of Tables

Chapter 1

Introduction

People plan their routes through new environments every day, but what factors influence these wayfinding decisions? Although researchers have studied how humans navigate, what affects their success and ability to do this and the attributes which encourage people to walk, few studies have investigated which factors are important when selecting a route for pedestrian travel beyond the shortest path approach.

In a world increasingly dependent on electronic navigation assistance devices, finding a way of automatically selecting routes suitable for pedestrian travel is an important challenge. With a greater freedom of movement than offered by roads for vehicular transport, and different requirements, an alternative approach should be taken to find an answer for journeys on foot. Preference for pedestrian routes involve more complex and diverse attributes than purely the distance to be travelled, and previous studies have also indicated that different types of journey require different characteristics [6]. Further investigations, have also indicated the motivations for the selection of specific routes for a given purpose [83].

Although previous research has produced a number of route recommendation systems, the majority of these are restricted to a single route type or user group [36, 42, 95, 145, 152, amongst many others]. In addition, many systems use complex approaches such as genetic algorithms or fuzzy logic [5, 33], or require information (such as obstacles) which is not readily available [4, for example]. The processing requirements for these previous solutions means that most are not suitable for use in power-conscious environments (e.g. mobile devices), or they require complex databases and interfaces [101].

The overarching goal of this research is to create a single route recommendation system that can be used to suggest appropriate routes for different types of journey. This system should use route and environment attributes which are readily available from existing maps, or could be crowd sourced from local knowledge, and are applicable in many geographic areas. The algorithms to suggest routes should be simple enough to be run on many different platforms, and efficient enough to be embedded in power conscious environments such as mobile devices. These algorithms should also be straightforward enough to incorporate simple approaches for personalising the routes suggested, using variables such as maximum route length.

Although the collection of crowd sourced data is beyond the scope of this project, as is the testing of the developed algorithm on multiple platforms, the aim of this thesis is to develop an approach to route suggestion, which by combining many predefined subjective and objective attributes, can recommend routes according to the type of journey a person is making. It will be based on a simple algorithm and associated methods, giving a solution which is computationally straightforward to implement in many different environments. It will also use only attributes which can be gathered easily from maps or satellite images. In order to achieve this aim four areas of research have to be undertaken.

1. Empirically establish how different environment and route attributes influence pedestrian preference.

2. Create an environment model which is simple enough to be applied to many geographical areas, but detailed enough to allow route selection.

3. Develop simple and efficient route selection algorithms.

4. Empirically evaluate the routes suggested, and therefore the performance of each algorithm.
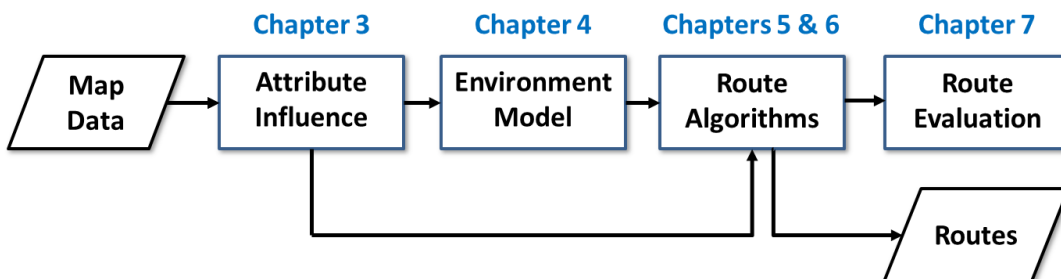
Figure 1.1: The route recommendation system.

Figure 1.1 shows how these components are used to build a route recommendation system, and the chapters which correspond to each component.

1.1 Contributions

There are three main contributions of this research:

1. This research has established previously unknown rank orders for the influence of seven environment and route attributes (length, turns, decision points, vegetation, land use, dwellings and points of interest), on 'attractive' route selection and for leisure and tourist journeys. In addition, this research has extended the previously known rank order of these attributes for everyday journeys. These ranks were found empirically by running a set of six experiments with 450 participants, which compared more attributes simultaneously than earlier studies.

2. This research has also developed new multi-attribute algorithms, based on Dijkstra's algorithm, which suggest routes that people would find attractive, and for tourist and leisure journeys. These algorithms were formulated using the developed environment model and rank-ordering, taken directly from experiments. By avoiding the complexity of previous approaches, these algorithms could be widely used across a variety of environments, and easily extended for different groups of users.

3. Two of the tourist and leisure algorithms that were developed have been shown to select routes that were similar to those suggested by participants in a user experiment for tourist and leisure journeys, respectively.

1.2 Thesis Outline

This chapter has so far briefly discussed the goal of this research, the problem it aims to address and the drawbacks of previous approaches. In addition it has detailed the contributions made in achieving this goal. Chapter 2 discusses previous research, both in the route recommendation field and other related research areas. It gives details about existing systems, the attributes and journey types to be used in this research and the environment representation chosen to be the basis of the system. It also examines route selection problems which may be applicable to the present research, and investigates the algorithms available for solving these problems along with their advantages, drawbacks and performance.

The design and results of a study to rank seven environment and route attributes (length, turns, decision points, vegetation, land use, dwellings and points of interest) according to specific route type are given in Chapter 3. This user study consisted of six experiments which applied a stated preference approach to determine participant route

preference, and analysed this preference to establish the order of influence of the selected attributes for two attribute categories (simplicity and attractiveness) and three journey types (everyday, leisure and tourist).

In Chapter 4, these attributes are combined with data taken from OpenStreetMap to produce a suitable model of the University of Leeds campus, the environment chosen for this research. This model takes the form of an annotated graph, and details are given of an investigation which examined different approaches for representing some of the more subjective attributes to find an appropriate model.

Chapter 5 and Chapter 6 describe the investigation performed to establish appropriate algorithms for route selection (Section 5.3), a suitable way to represent single attributes as costs in this process (Section 5.3.1 and Section 5.3.2) and finally to combine these attribute costs to construct algorithms capable of suggesting routes for different journey types (Chapter 6). Chapter 5 also discusses the development of a tool used to investigate the performance of each algorithm and select a test set of points, and Chapter 6 describes the metrics used to establish the 'best' weights for each multi-attribute cost function.

A second user study is described in Chapter 7, which evaluated the performance of the route suggestion algorithms. It first discusses the design of the study to gather participant route suggestions for each attribute category and journey type. The resulting routes were then analysed, and the attribute characteristics were compared to the results found in Chapter 3. Chapter 7 also uses the collected route suggestions to evaluate the algorithm selected routes. Finally, Chapter 8 draws overall conclusions from the research, and gives suggestions for future work.

Chapter 2

Background

In the field of personal navigation, many different commercial systems are becoming available, from mobile apps to specialised Personal Navigation Devices (PNDs). One popular market research report [21, see summary] indicated that there were 180 million dedicated car navigation systems in use globally at the end of 2013, and that the number of monthly users of mobile navigation apps was approximately equal to this. Of the dedicated personal navigation devices, three main players - Garmin [75], TomTom [215] and MiTAC [154] - hold 75% of the global market, but these specialised systems face increasing threats from the mobile phone and tablet navigation apps that are now becoming more prevalent.

With a lack of commercial system reviews being available in research publications websites [14, 148, 194, 229, for example] must be used to consider performance. However the majority of reviews within these are based purely on cost and usability, giving very little consideration to the actual routes they produce. Most existing commercial pedestrian route recommendation systems suggest routes using the shortest path approach [75, 84], with switching from vehicular journeys to those on foot only affecting the inclusion of pedestrian-only areas and designated trackways. Even systems which suggest more than one possible solution (such as Google Maps Navigation [84]) are reliant on discovering two or three of the shortest routes available. However, humans rarely rely on ththe length of the route alone [83]. In contrast, vehicle SatNav systems offer alternative heuristics based on fuel usage, aesthetics or other similar attributes [75, 215]. However, although a scenic route may be desirable for pedestrian travel, even this strategy may need to be

adapted as the 'scenic' quality for vehicular travel is restricted to the view from a moving vehicle window (see [197] for an explanation of this). One exception to this is Walkit [230], which does offer different types of routes to be selected, but even then the choice is restricted to 'low pollution' or 'less busy' routes which aim for more walking to be done on off-road walkways.

Although shortest path routes (including the variations suggested above) may be the most appropriate for getting quickly from one point to another, they are unlikely to be the most appropriate for those wanting to walk for leisure (say for a stroll on a nice day) [134], or for tourists looking to explore the sights of a city [205]. This implies that the existing 'one size fits all' approach to commercial pedestrian route selection is restricting the types of journeys for which it can be used. The present research aims to extend the possible journey types covered, by investigating which attributes of the route and surrounding environment are considered important for different journeys. This information will then be used to develop a suitable environment model, and form the basis for algorithms which can select routes for different types of journey.

The following sections review previous research, which provides a foundation for the present research. First we discuss the attributes that are considered important for pedestrian wayfinding, and the expected importance of these for different types of journeys. We then examine different approaches to route suggestion offered by previous reseach into navigational aids, along with their limitations. Next a number of possible wayfinding problems to tackle will be established, and suitable environment models discussed. Finally, algorithms to solve the wayfinding problems will be examined, their individual advantages and disadvantages compared, and a suitable approach will be selected.

2.1 How Environment and Route Attributes Affect Route Choice

Many factors have been shown to affect the choice of pedestrian routes. An empirical study examining travel between two spatially distinct locations [83] discovered nine criteria that are commonly used when planning routes in unknown areas. Further criteria linked to user interest [197] and perception [191, 228, 236] can be added to this list, along with those concerning salient features [6] and ambient conditions [129]. Examples of some of the criteria which have been suggested are shown in Table 2.1, but this is by no means exhaustive.

Although all of these criteria are known to influence navigational decisions, not all

Physical		Non-Physical
Shortest Distance[1]	Longest/Shortest Leg First[1]	First Noticed[1]
Least Time[1]	Number of Landmarks[3]	Safety[5]
Least Effort[2]	Fewest/Many Turns[1]	Accessibility[7]
Many Curves[1]	Number of Decision Points[4]	Weather[5]
Aesthetics[1]	Points of Interest[6]	Time of Day[5]

Table 2.1: Examples of Attributes Affecting Pedestrian Route Choice.
[1] [83],[2] [236], [3] [6], [4] [191], [5] [129], [6] [197] and [7] [228]

.

are appropriate for testing as part of this thesis. Some can be combined and measured directly from maps such as distance, implied using tested assumptions such as time taken, whilst others cannot be easily represented by physical features for example wind or safety. Still more would be expected to work only in conjunction with other attributes, such as the reduction in preference for parkland areas at night or for streets which are busy at lunchtime, which require knowledge not easily gained from a map (for examples of how some these affect preference see [83] and [129]). The present research will focus on those affecting physical attributes which can be gathered from maps, as specified in Table 2.2. Of the attributes shown, turns will be further defined as changes of direction at decision points rather than the general angle of overall travel (considered to be curves in Table 2.1). These attributes will be more clearly defined in the following sections, including breaking aesthetics down into some of its constituents, and how they can be used to form specific approaches for different journey types will be examined.

Longest / Shortest Leg First	Number of Landmarks	Fewest / Many Turns	Number of Decision Points
Length (distance)	Aesthetics	Points of Interest	

Table 2.2: Selected Pedestrian Attributes

2.1.1 Simplicity: Attributes Which Affect the Ability to Navigate

Simplicity in the context of wayfinding is how easy a route is to navigate, and is associated with the layout of a route or environment, and the presence of cues such as landmarks. Length is known to have a large effect on simplicity and, with all other things being equal, longer routes will require a larger cognitive load to navigate [98, 182, amongst many others]. In fact, the effects of length on simplicity are so well understood that they will not be investigated by the present research, with only the influence of length on other journey types being examined.

Turns (changes of direction at decision points) feature in a list of reasons given for choosing a route [83], and a route with a smaller number of turns is also preferred when choosing between two alternatives on a map [12]. In addition decision points with no turns have been shown to affect perceived distance [191] and the likelihood of correctly traversing a route [163, 164], regardless of participants' familiarity with it.

The influence of landmarks on wayfinding has been the focus of much previous research [142, 199, and many more]. It is now generally accepted that they are a key, usually visual, cue to help people determine the right direction at decision points [150]. They increase legibility [142], decrease wayfinding errors and reduce the time taken to learn new routes [110, 189].

Unlike the attributes discussed so far, the influence of initial leg length on wayfinding success is less clear. Also known as the Initial Segment Strategy [12, 102], the result of this attribute is a preference for different paths between two points dependent on the direction of travel [12]. The asymmetry in movement occurs by delaying changes of direction until the end of the route, in an attempt to minimise cognitive effort [12, 83].

2.1.2 Attractiveness: Attributes Which Affect Scene Preference

Attractiveness, or how 'scenic' a route is, is associated with the areas surrounding a route, and the views visible when walking along it. The first criteria in this group, aesthetics (see Table 2.2), can be subdivided into a number of attributes such as vegetation, land use, cleanliness, maintenance and architecture [180]. Of these, cleanliness and maintenance are considered to be outside the scope of the present research due to having only a small effect on participant preference [180], and difficulties in representing them adequately in this form of testing. To consider the influence of the remaining factors, we must look to the preference for them in everyday life.

Vegetation is known to affect house prices, reduce stress and anxiety, and may even be therapeutic [203], as well as being associated with a variety of recreational activities. In general, plant life has a positive effect on satisfaction with urban scenes, and trees in particular are valued for their shade and ability to increase visual complexity, a key factor in aesthetics assessment. Vegetation is listed as fourth on a list of scenic qualities which make roadsides more pleasing [120], and a previous study [222] indicates that its presence increases the preference for one urban view over another.

In addition to the effects of vegetation, specific land use such as nearby parks or countryside is also known to allow properties to command higher prices [203]. Natural views including urban parks and farmland are preferred to residential ones which are in

turn preferred to views of commercial areas [222], and green space encourages walking for recreation as well as everyday travel [180]. In contrast, the prevalence of signs and wires in commercial areas make them by far the less desirable areas [203].

Evidence for the effects of dwelling attractiveness on route choice comes from research into participants' responses to urban scenes in four different empirical studies. To be considered aesthetically pleasing, architecture and building frontages should evoke feelings of pleasantness, excitement and calmness [156], with colour, novelty, complexity, shape, type and style all playing a part. In a study of reasons for enjoyment of scenic routes [120], single family housing was considered more attractive than blocks of flats or buildings containing multiple family units, and stone walls and historic sites increased how pleasant an area was thought to be. In addition, architectural designs can encourage walking [180], and the presence of older or more traditional buildings is preferred in photographs taken from scenic routes [121].

The use or meaning of a structure, be it a building, a monument or even a fountain, may also be of significance for human preference. This characteristic is covered by the term points of interest, indicating that more than just a building frontage is being considered. Previous studies have shown that these 'sights' make driving more memorable and pleasurable [120], and even encourage walking [180]. Scenes with churches had the highest preference in an earlier study on scenic quality [121], and points of interest are of particular importance for tourists [206].

2.1.3 The Relationship Between Attributes and Journey Types

Research into the cognitive components of wayfinding [6] indicates that when planning a route, the decision process is dependent on the type of journey to be completed. An example of this could be the differences between the daily commute to work and the route taken when going for a stroll on a summer morning. Four main types of journeys can be identified [62, 221]:

1. Tourism - A typical example of this would be an individual visiting an area to see the 'sights'.

2. Business Trips - Short trips to new destinations for work purposes.

3. Leisure Journeys - Usually recreational, these can be in a familiar or unknown environment, and are mostly aimed at finding aesthetically pleasing areas to travel through.

4. Everyday Navigation - Trips performed regularly such as the daily commute to work, or visiting the local shops.

Of these, business trips may be considered to have many of the same characteristics as everyday journeys, and were therefore excluded from the present research.

If selecting a route between two fixed points, either on a map or in the field, is assumed to portray everyday travel, then previous research has suggested that simplicity and attractiveness both affect this everyday journey type to different extents [83]. Distance has the most influence on everyday journeys, with the preference for shorter route lengths for commuting and other regular trips confirmed by the work of several other researchers [134, 197, amongst others]. In addition as the number of turns and decision points encountered increase the perceived distance between two points [210], so it would seem likely that these too would have a role. With the longest leg first attribute also increasing complexity [12, 83], then this too should be added to the list of attributes which affect route choice for everyday routes - this time with the longer the initial leg being preferred. Finally, although aesthetic quality is composed of many different characteristics, a further review of the environmental attributes which affect walking [171] indicates that land use is an important characteristic of aesthetics for walking for all purposes.

In contrast, recreational journeys usually involve much longer routes [134, 153, 233], indicating that attributes associated with length will be far less influential and possibly have a positive relationship with preference. Land use is considered highly influential in walking for pleasure or exercise [134] and, as walkability [193] is thought to influence this type of journey, vegetation, dwellings and points of interest are also considerations [78, 106, 134, 151, 207]. Unfortunately, no previous research has considered the influence of the remaining attributes associated with simplicity.

Of the three journey types, the previous research on which attributes affect route choice for tourist journeys is the most limited. What id well known is that for tourist trips, points of interest are expected to be the most important factor when selecting routes [205]. Aesthetics can be added to the characteristics affecting this type of journey [158], with dwellings and land use being suggested as influential factors [225]. Several studies have indicated that vegetation may also sway the directions taken during this type of travel [66, 145, 225], but very little is known about the remaining route and environment attributes.

2.1.4 Summary

Many route and environment attributes may affect route choice for different types of journey. This section has identified three common journey types (everyday, leisure and tourist), and nine different attributes which may influence route choices (Landmarks, turns, decision points, vegetation, land use, dwellings, points of interest and length). All nine attributes can all be taken directly from a map, fall into the category of route simplicity or attractiveness (although length will be treated separately), and have already suggested by previous research to affect route choice for differing journey types as shown in Table 2.3.

Category	Attribute	Everyday	Leisure	Tourist
Simplicity	DPs	✳	?	?
	Turns	✓	?	?
	Landmarks	✓	?	?
	Initial leg Length	✓	?	?
	Length	✓	✳	?
Attractiveness	Land use	✓	✓	✳
	Dwellings	?	✓	✳
	Vegetation	?	✓	✳
	POIs	?	✓	✳

Table 2.3: Attribute influence on route choice for each journey type. Previously shown effects (✓), effects which are inferred from previous studies (✳) or unknown effects (**?**). (POIs - points of interest, DPs - decision points)

The present research will use known influences to establish the validity of a test methodology, establish the influences which are unknown or only implied, and then determine the order of importance of each attribute in selecting routes for the three chosen journey types.

2.2 Existing Navigation Aids in Research

In addition to those available commercially, previous research has also suggested many wayfinding aids which can be applied to pedestrian travel, with some of the most successful specialist wayfinding support being those created for individuals with visual or cognitive impairment. Each of them focuses on presenting relevant information, or the use of different route criteria, for either single types of journey, specific user groups or to develop adaptable or adaptive systems. Table 2.4 lists some of the approaches already available, and shows the different attributes used by each (where available). Although

the details of some of these systems are sparse, none definitively use all of the attributes to be considered by the current research. This section will briefly review these existing systems, and discuss their advantages and drawbacks.

2.2.1 Single Journey Type Systems

Research has offered alternative solutions for specific types of journey, such as tourist trips and everyday journeys, as well as for small geographic areas. Tourist aids such as GUIDE [42], CAERUS [155], HIPS [19], CyberGuide [2], TellMaris [195], and Deep Map [145] give routes between attractions usually within cities or towns. Their main purpose is to provide multimodal transport suggestions, or extra information on these destinations through the use of multimedia data displays. The majority, therefore, concentrate on data which is not readily available (or needs to be annotated) [2, 145] and visualisation techniques and human interfaces [155, for example], rather than on the algorithms to choose routes between points of interest. In fact, all but two of these systems give little or no detail about how the routes are chosen, with one [42] mentioning only that length and aesthetics are involved, and the other [145] being based on a complex user database system. Furthermore, although the HIPS system [19] implies that its approach is extendable to other journey types and locations, and a number of the other approaches imply that they can be applied to other geographical areas [42, 145, 195], little or no detail on how this would be achieved is given. Also, where evaluation of the aids are discussed, all rely on field usability tests focused on the data provided and its delivery method, rather than the routes selected.

Other systems relate to everyday types of journey, covering a variety of specific environments such as a train station [152], an airport [13, 126], a town in Norway [125] and even a large area of urban Japan [8]. With specialised databases being made available for only these locations, these systems cannot easily be applied to other geographical areas ([152] for example does not use map data, and would therefore require all of the routes and landmarks through multiple areas being manually added to the system). One larger example of this type of route recommendation system [8] involves the integration of data from many areas such as train/bus timetables and ticket costs, as well as maps, to produce multiple route alternatives to the user. Although this does consider multiple factors in selecting routes and suitable itineraries, the approach is complicated and requires specialised interfacing between different data sources. In many geographical these resources are simply not available, and where they do exist no suitable interface has been developed.

Previous research has also attempted to produce alternative approaches to the shortest

System	Length	Land marks	Turns	DPs	Veget- ation	Land Use	Dwell- ings	POIs
Single Journey Type Systems								
GUIDE [42]	✓							✓
CAERUS [155]	✓							✓
HIPS [19]	✓							✓
Cyberguide [2]								✓
TellMaris [195]	✓							✓
Deep Map [145]	✳	✳	✳	✳	✳	✳	✳	✓
[152]	✓	✓		✓				
REAL [126]	✓	✓	✓	✓				
Navitime [8]	✓							
"Simplest" Paths [57]		✓	✓	✓				
[93]	✓		✓	✓				
[31]		✓		✓				
Specific User Type Systems								
Drishti [95, 184]	✓							
MOBIC [179]	✓	✓	✳	✳				
RouteCheckr [228]	✓	✳	✳	✳				✓
InMyOneWay [64]	✓	✳						
Adaptable/Adaptive Route Planning								
[4]	✓				✳	✳	✳	✳
Turas [147]	✓			✓				
[176]	✓			✓				
[40]	✓			✓	✳			
[101]	✓	✓	✓	✓	✓	✓		✓
[33]	✓				✓	✓		✓
[187]	✓			✓				

Table 2.4: Summary of the attributes used by existing navigation aids. Previous research indicates that each characteristic is present (✓) or implied (✳).

path algorithm which combine environment and route attributes, with one group of solutions [31, 57, 93] attempting to minimise complexity by examining the directions required to describe them, or the junctions encountered along the way. Although not discussed directly, these systems could easily be applied to other geographical areas (assuming that the issues associated with landmarks could be resolved for [31]), and use simple minimum cost routing solutions. However, all of these approaches are designed specifically to solve only the the task of finding the 'simplest' route, and are not applicable to any other route types. In addition, evaluation of these systems was carried out by simulation, comparing routes to those produced by a shortest path algorithm, with no human assessment of route suitability.

2.2.2 Specific User Type Systems for Everyday Travel

Of the more specialist wayfinding support tools, some of the most successful are those for individuals with visual or cognitive impairment. Just as these two conditions are very different, the navigational aids provided for them adopt contrasting techniques when providing routes.

Visual impairment aids, such as Drishti [95, 184], MOBIC [179] and RouteCheckr [228], require very specialised environmental representations, involving non-visual landmarks and the incorporation of known hazards. In this way they are distinct from the other tools, necessitating very different environment models including data which is not readily available for many geographic areas, and particular types of delivery systems. Despite this need for specialised data, most of the systems (including two [95, 228] of the three discussed here) are based on simple minimum cost algorithms, which combine attributes such as accessibility and obstacles into weighted route cost functions, with the majority of research focusing on the method for delivering route instructions to the user. In addition, both MOBIC and Drishti have been applied to multiple locations, and field tests suggested that they were both successful. Furthermore, the Drishti system [95, 184] implies that it could be adapted to work for multiple route types and user groups, but again no details are given on how this could be achieved.

In contrast, tools for wayfinding with cognitive impairments may require only minor specialisations to non-specific user systems, such as an alternative method of route delivery [139, 204, 226], or finding a simple solution to guiding individuals along a single route [35–38, 217]. Other approaches detect wandering [140], or allow users to indicate that they are lost and either contact caregivers or suggest routes [53, 64]. These cognitive impairment systems seek to overcome the issues of 'getting lost behaviour' but offer very

different solutions, the majority of which are applicable only to this group of users. All but one [64] of these aids rely on predefined routes or additional technologies, and no details are given on how the remaining approach chooses suitable routes.

2.2.3   Adaptable/Adaptive Route Planning Models

Other research has sought to produce approaches which are adaptable or adaptive to the individual using them, rather than being tied to specific areas or user groups. The most common of these is to form user preference databases in advance and then employ these to predict the most suitable route [4, 147], in one case [176] by collecting data automatically from the users' everyday wayfinding decisions. Another approach requires users to interact with the system, either by ranking initial route criteria [40, 101] or selecting from a set of suitable routes suggested by the algorithm [33, 187]. In many cases these models have been designed for alternative modes of transport such as cycling [40, 101, 176, 187], but are also applicable to pedestrian travel.

Of all the approaches discussed in Section 2.2, these adaptable or adaptive approaches are the most applicable to different journey types, user groups or geographical locations, but are also based on some of the most complex algorithms. Fuzzy logic is a frequently used solution [4, 33, 40] for adapting route suggestions, along with genetic algorithms [33] and user databases [147, 176, 187].

Although these systems produce routes which are tailored to the user, and therefore more likely to closely fit their individual needs, they are usually much harder to use. The requirement to supply initial data or select from multiple route suggestions implies complex interfaces, which may be too confusing for older users or those with some level of cognitive impairment. Also, recording and reusing previous routes or preferences requires large amounts of data storage to be supplied by the system.

Finally, all but two of these adaptable systems [4, 101] are evaluated only by performing efficiency tests or simulating user profiles. Although these may compare the suggested routes against benchmarks produced by other algorithms [33, 147, 176], few (only [40] and [187] of those discussed here) use real human data, and even fewer compare to human suggested routes [187]. Although these forms of evaluation are valid, they do little to assess the suitability of the system produced routes.

2.2.4   Summary

Very few of the existing systems offer solutions for multiple journey types, geographical areas or user groups, and many rely on data which is not readily available from maps

or similar sources. Also, only 15 of the 24 reviewed here give details of algorithms being used to suggest routes, only seven of which require no additional human input or complex selection algorithms, and of those only one [187] uses human participants to evaluate the suitability of these suggestions. Also, none of these existing systems combine readily available data with a simple algorithm to suggest routes for multiple journey types - principles which form the basis for this research. Some of these points are outside of the scope of this thesis (such as multiple geographic areas and user groups), although the system produced will be conducive to extension to include them at a later time.

In addition, Table 2.4 shows a summary of the attributes mentioned by each of the existing navigation aids that were discussed. It should be noted that there is limited information given on specific attributes in many of the earlier systems, so it is possible that more are used than could be determined here. An obvious omission from this table is the attribute 'Initial Leg Length', which does not appear in any of the previous systems. Table 2.4 also shows that many of the existing approaches rely on just two or three of the attributes to be used in the proposed system, with only two [101, 145] considering more than six, and only one of these [101] refers to them specifically rather than just discussing simplicity or aesthetics. None of the described systems consider all of the attributes detailed in Section 2.1. One reason for this may be that not all attributes are required for all journey types, but this restriction will be examined empirically to determine its validity before introducing it into the system in this thesis.

2.3   Wayfinding Problems

In order to successfully choose appropriate routes through an environment for different types of journey, one or more environment models and route selection algorithms must be constructed. However, the relationship between the model and the algorithm is cyclic - the basic format of the environment model influences the route selection algorithm choice, whilst the algorithm requirements determine the content (and to some extent the format) of the model to be used. Overriding both of these factors are the data which is available, and the wayfinding problems that they are designed to solve. Section 2.1 has considered some of the attributes which are important for different types of journeys and should therefore be made available, but this section will examine the different wayfinding problems that the journeys could pose, and relate each of them to known computing problems which could suggest possible solutions. By defining these problems, appropriate environment models and route selection algorithms can then be determined.

2.3.1 Point to Single Destination Pathfinding

The most common problem in route selection is that of finding a route from a start point to a single specified destination, usually covering the least distance possible. The Shortest Path problem has known and accepted solutions (see [173] for an overview of some of these), dependent on the environment to be searched, but it is not the only problem which can be considered when selecting a route between a point and a single destination.

2.3.1.1 Minimum Cost Problems

The Shortest Path problem is a specific case of a larger group of Minimum Cost problems [94]. Even 'shortest path' could cover multiple problems, such as the physical distance travelled by the route, the time taken to travel it or the effort required to complete the route. Any attribute indicated to have a negative relationship with route choice, could be considered to be a Minimum Cost problem for the purpose of route selection [94]. For the attributes discussed in Section 2.1, minimum cost approaches could be used to find routes with either the minimum number of turns or decision points.

Minimum Cost problems are not directly applicable to attributes with positive relationships to route choice (landmarks, points of interest, vegetation, dwellings, land use, and initial leg length in Section 2.1), although converting these attributes into a form which is appropriate is possible. One approach to this would be to consider the length or proportion of the route which doesn't encounter any of the required attributes, or alternatively to define a threshold indicating low attribute density. The problem then becomes to seek routes which minimise the length of the route which has low levels of the attribute, rather than maximising the length of the route which contains high levels.

One of the biggest advantages of Minimum Cost problems is the ability to combine attributes in a straightforward way. Weighted cost functions allow more than one attribute to contribute to the cost of a route, and therefore influence route selection. In contrast, the conversion of positively related attributes can also lead to problems and drawbacks. Keeping the attributes independent can be complicated, lead to excessively long routes, and require careful consideration of how the attributes should be represented within the chosen environment model.

2.3.1.2 Alternatives for Maximising Attribute Values

Unlike problems which seek to minimise attribute values (cost), problems which aim to maximise attribute values are harder to define and solve. For example, if the aim was to find the longest route through an environment, this could be considered to be an

infinite problem - cycling through paths in the environment continuously increases the length [118]. To overcome this issue, it could be argued that the problem should be redefined to become that of finding a route which traverses all paths in that environment, known as the Chinese Postman (or Route Inspection) problem [59]. This problem aims to traverse each path only once where possible, or with the minimum repeated distance where this is not achievable, and has polynomial time solutions. However, these solutions rely on one specific property of length, and the fact that for each path it has a positive, non-zero value. For attributes such as vegetation or points of interest, this property can not be guaranteed.

In a real-world environment, it is likely that some paths encounter no instances of these attributes. This alters the problem, as the route containing the maximum value of one of these attributes does not necessarily include every path. By restricting the paths to be included in the route, the task becomes more closely related to the Rural Postman problem [60]. The Rural postman Problem is one commonly encountered in many real-world scenarios, from utility meter reading to snow ploughing, in addition to the postal service suggested by the name [60, 97]. In this scenario, only a subset of the paths within an environment are traversed, defined to be those with a perceived benefit (traditionally selecting only the roads containing houses with post to be delivered). This benefit could easily be derived from from each attribute with a positive relationship with route choice, dependent on journey types (see Section 2.1). An alternative way to simplify the process, is to define areas which have a high density of the attribute, and restrict the route to these areas where possible. These high density areas could be considered as an environment model in their own right, with only the shortest routes between them also included, and the remaining low density areas discarded.

An advantage to these approaches is the simplicity of representing the attributes, which can be applied directly to the environment without conversion. They also offer straightforward, and relatively efficient, solutions to the problem of attribute maximisation.

One disadvantage of maximising attributes such as vegetation or parkland is the likely length of the routes produced. Although a route suitable for leisure, such as going for a stroll, should have high levels of the attributes suggested by Section 2.1, having an unrealistic length (say 10 miles or more) would still make it inappropriate for pedestrian travel in many instances. As an alternative, the same high density areas as described above could be used to form shorter routes for certain attributes. An example of this might be directing the route to the nearest area of parkland, and then using some variant of the route inspection problem to select a restricted length route through this area.

A further disadvantage of maximising attribute values is the necessity to combine attributes for different types of journeys. As with minimising problems, the use of weighted cost functions is also possible here, but in real-world environments the attributes may be of conflicting densities in different areas. This would make partitioning the environment according to high and low density areas difficult.

2.3.2 Multiple Destination Pathfinding

The approach of partitioning the environment into high and low density areas, and then travelling to each of the high density ones (as described above), offers a different perspective on the possible wayfinding problems which could be solved - that of multiple destination pathfinding. Unlike single destination problems, the aim of multiple destination pathfinding is to visit all (or as many as possible) of a set of 'destinations' in an environment, whilst still selecting routes that are optimal in terms of one or more additional attributes.

2.3.2.1 Problems Which Include all Destinations

An obvious example of this type of problem within the context of the present research is that of selecting a route which visits all of the points of interest in an area, but does it in the shortest length possible. Points of interest may be the most obvious attribute to which this approach may be applied, but it is not the only one. Assuming that suitable thresholds and partitioning algorithms can be found to separate high and low density areas, any attribute with a positive route choice relationship could be maximised in this way, simply by labelling each high density area as a destination.

If the start and end points are the same, then multiple destination pathfinding reduces to the Travelling Salesman problem [135]. In this, each destination is labelled as a 'city' and the aim of the problem is to visit all of the cities, with a route which travels the minimum distance possible. Alternatively, if the start and end points are different, then the generalised Minimum Length Hamiltonian Path problem [88] applies. In both cases, a number of both exact and heuristic solutions are available. However, as mentioned previously (Section 2.3.1.2), the routes selected are likely to be longer than appropriate for pedestrian travel.

2.3.2.2 Problems with a Subset of Destinations

It may be possible to reduce the number of required destinations, using methods such as filtering by importance (possibly according to user interests). In many respects, this

problem could be considered similar to the Travelling Purchaser problem [178]. This type of problem seeks to reach a subset of destinations, with both minimum distance and the optimisation of another independent cost. It could easily be adapted to maximise a specific type of point of interest, or minimise repetition of similar areas of parkland. In addition, the travelling purchaser problem could be extended to consider more than one of the attributes discussed in Section 2.1, for instance both the amount of vegetation and the type of dwellings inside and area specified as a 'destination', when selecting the best route.

This approach offers advantages in terms of both route length and adaptability to the user. The distance traversed by the selected route could be varied by adjusting the number of destinations included in the subset, by allowing the user to select which points of interest are essential, or offering the possibility of choosing which areas of vegetation should be included. However, the same problems with combining attributes, as discussed in Section 2.3.2.1, may still be encountered. In addition, it is likely that more information on the nature of the specific attributes would be required, such as the type of points of interest or features of parkland and vegetation. These may be difficult or impossible to gather directly from a map.

2.3.3   Summary

This section indicates that although the shortest path problem is the most commonly solved wayfinding problem, it is by no means the only one. Whilst minimum cost approaches do offer the best solution for minimising negatively related attributes, there are several alternatives which may offer more appropriate answers for maximising positively related attributes.

The decision over which problems, and therefore possible solutions, are to be used by the present research will be discussed later in this chapter. However the wayfinding problems mentioned in this section will form a basis on which to examine suitable environment model and algorithms in the following sections.

2.4   Environment Models

In order to be able to select routes through an environment, the environment must first be modelled in a way that is accessible by a computer algorithm and suitable for route selection. Environments have been represented in many different ways by previous research, from ontologies [87, 127] and image collages [219], to deformed grids [99], graphs [100]

and 3D models [125]. These representations include both spatial and non-spatial models, as well as visual and non-visual approaches. Although the majority of these environment representations have been used for route navigation, not all are easily created from maps or GIS databases or suitable for comparing routes. This section compares three environment models which are applicable to the problems described is Section 2.3, discusses their advantages and drawbacks, and establishes which of the three is most appropriate for the research to be conducted in this thesis.

2.4.1   Graph Models

Of the 24 navigation aids discussed in Section 2.2, 15 give detailed descriptions of their environment representations, 13 of which are based on graphs (the remaining two use grid or 3D representations with no details of route recommendation approaches, and so are not appropriate here). The popularity of this approach can be understood if the perception of urban environments and availability of routing algorithms are considered. A pivotal work on urban navigation [142] defines five elements used to describe urban areas:

1. Paths - Roads, walkways and other channels which can be traversed.

2. Nodes - Junctions or changes in transportation methods.

3. Landmarks - Visual cues used to navigate through urban environments.

4. Edges - Barriers which are either physical such as walls, or perceptual as with lines between districts.

5. Districts - Two dimensional areas considered to have some common characteristic.

 Each of these elements can be related directly to objects within most GIS databases, if barriers such as railway tracks are considered edges, and areas such as woodland are districts. Furthermore, paths and nodes can be considered to map to the edges and vertices of a graph, which can then be used to describe the possible routes through an area. Figure 2.1 shows a small section of the campus map, and two variations on graph representation.

 The most obvious advantage of graph models is their ability to be used in solving all of the problems discussed in Section2.3. Extending this representation to be a weighted graph, by annotating the edges with the distance between the nodes, allows the commonly used shortest path approach to route selection to be easily (and computationally efficiently) solved using graph theory [54], making this form of model very attractive.

(a)

(b)

(c)
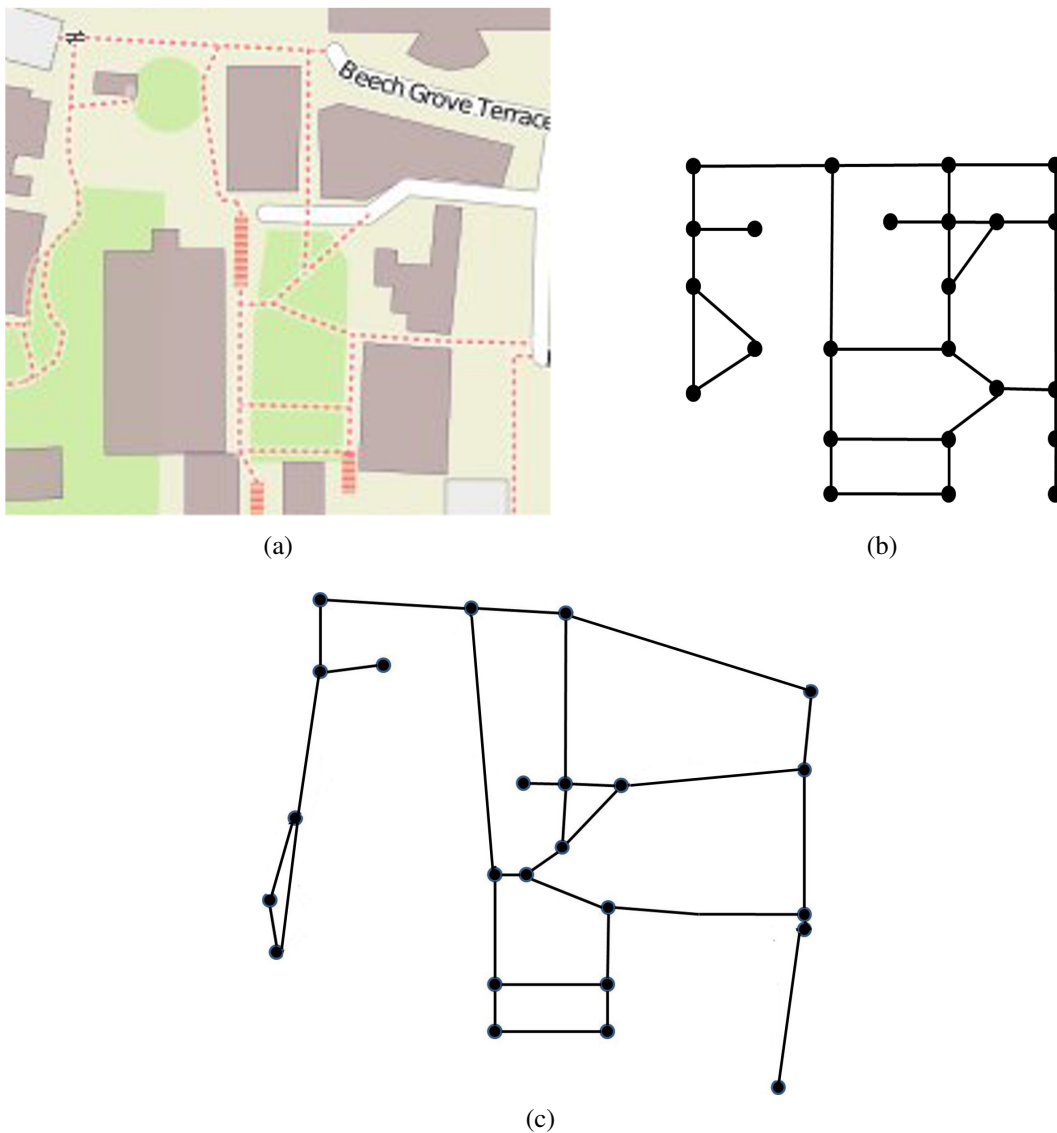
Figure 2.1: Three environment representations of a small section of the test area, (a) the campus map, (b) a graph representation and (c) a deformed grid.

In reality all of the problems discussed in Section2.3 are also well known graph theory research areas, with a large number of solutions available for each.

Another of the major advantages, and possible disadvantages, of a true graph representation (Figure 2.1b) is that the resulting model no longer has any spatial references associated with physical location. Although this makes for a concise environment model, reducing the required storage space, the relationship between the map and the representation is lost. For successful route suggestion this relationship is not necessary, but in the present research the turns attribute will be calculated at runtime so the angle between edges must be represented within the model. In addition, location data is essential for

visualising the model and routes. However, the remaining attributes can be successfully annotated to either the edges or vertices of the graph, so extending the graph representation to incorporate some spatial data is a sensible approach. One example of this retention of spatial information is the deformed grid shown in Figure 2.1c, which is commonly used in space syntax.

2.4.2   Space Syntax

Spatial information is required for a variety of different tasks, and previous research has offered alternatives to the true graph representation which retain this data. Most modern cities can be considered to form deformed grids [99], which can in turn be represented by graphs where nodes retain their geographical location (Figure 2.1c). Space Syntax uses this deformed grid as a start point to examine the relationship between space and human society, and produces representations which express lines-of-sight between points, integration and movement rates across urban environments [11, 99, 100, amongst others].

Space syntax aims to develop descriptions of manmade space, which allow theories to be formed about how spatial configuration affects behaviour [11]. Several different representations can be used in space syntax research, but in all cases they eliminate the concept of cost (especially in terms of distance) when calculating movement through an environment. Of these, the axial map is the most relevant to the present research. In the axial map the links in graph representation are replaced by lines of sight, and the nodes represent the points at which one path becomes visible to another. The map is constructed by drawing the longest lines of sight possible between parts of the environment, and continuing until all visibility between walkways have been represented. This results in a representation which looks very different to the underlying real-world environment, an example of which is shown in Figure 2.2.

A measure of integration for each point is then calculated by finding the mean depth of a node from all others, that is the number of turns that have to be made from one street segment or area to each other, or how many nodes must be passed through to reach this node [100]. This integration value has been found to be an indication of how many people can be found in the area that it represents, and therefore the expected flow of users (pedestrians in this case) [11].

In its traditional form the axial map cannot be automatically generated from a GIS database, and is somewhat subjective due to the need for human input, but this issue can be resolved to some extent by the use of angular analysis [49, 218]. Angular analysis replaces the nodes representing junctions with those representing straight path segments, and the

(a)   (b)

(c)
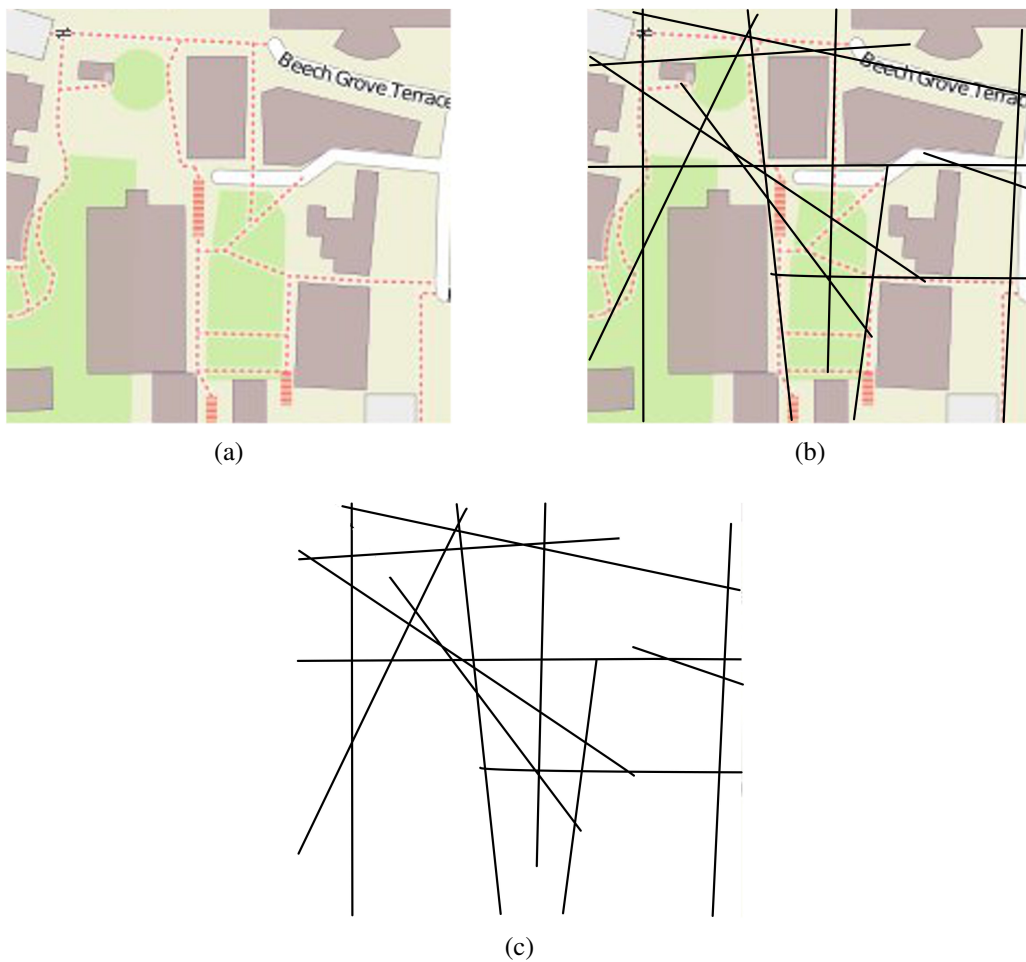
Figure 2.2: Three environment representations of a small section of the test area, (a) the campus map and (b) one possible axial map overlaid on the original, and (c) the axial map alone.

integration measure of a standard axial map with the mean angular change required to travel from one line segment to another [218]. This makes it easy to see that space syntax can offer an approach which statically incorporates the turns within a route, and although this gives a much simplified version of the environment model, in its traditional form it removes all of the attributes on which the present research is based. As with the existing shortest path approaches, it will give only one solution covering all of the possible journey types. In order to incorporate the attributes being considered by this thesis an additional layer of representation and computation would be required, adding to complexity and negating one of the main advantages of using this model.

2.4.3   Voronoi Diagrams

Unlike graphs and axial maps, Voronoi diagrams are not representations which connect locations, but are designed to partition environments into regions surrounding important locations [71]. Each of these regions defines the area of the environment which is nearest to the salient point or location contained within it, delineated by boundaries where the influence of points changes as shown in Figure 2.3. This results in a much simpler representation than the graph from which it is produced, containing only $n-1$ vertices and nodes at most (where n represents the number of important points within the environment).

<table>
<tr><td>(a)</td><td>(b)</td></tr>
</table>
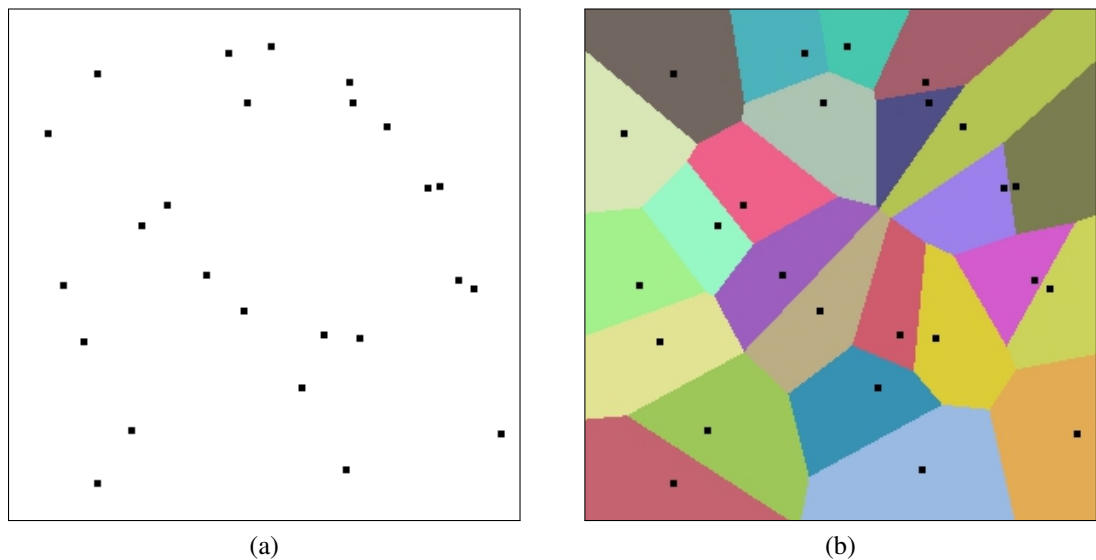
Figure 2.3: Images showing (a) 25 points of interest (b) the Voronoi diagram of the area around these points of interest.

Several algorithms have been suggested that can form Voronoi diagrams in realtime, such as Fortune's algorithm [70], divide-and-conquer algorithms [58, 198, amongst others] and incremental algorithms [162, for example]. Different measures of distance may be used in their construction in addition to the normal Euclidean approach [71], and the resulting structure also lends itself to rapid dynamic insertion and deletion of points through well known approaches [89, is one of the fastest].

Voronoi diagrams have many uses including nearest-neighbour queries (either single or for k neighbours), establishing closest pairs, cluster analysis, collision free path planning and regularly arise within computer games (for examples see [9, 69, 105]). In addition the dual Delaunay triangulation (a further representation which can be calculated from Voronoi diagrams) can be used to calculate minimum spanning trees, giving

reasonable if not the shortest Hamiltonian paths [9]. In this context they are most useful for solving the Hamiltonian path and Travelling Purchaser problems, although Voronoi diagrams could also be very useful for establishing which important locations are nearest the start and end points as well as defining the k nearest points to restrict the route length.

Despite the increased speed with which Voronoi diagrams can be used to access nearest-neighbour information, and insert or remove new points, they have two substantial drawbacks. The first is the need to place new start and end points within the diagram. If the Voronoi diagram is assumed to be based on Euclidean distance, then obviously the nearest important points to the required origin and destination can be calculated in a straightforward manner. However, if the representation is based on actual route length, then these calculations become more complex. In addition, displaying the selected routes to the user is difficult with on the Voronoi diagram structure. Both of these issues can be overcome by retaining the original graph environment on which the diagram is based, but this removes the biggest advantage of this approach - the simplified representation.

2.4.4   Summary

Of the three environment representations discussed each has its own advantages and disadvantages, depending on the wayfinding problem which is to be tackled. However, it is clear that the graph model offers the best solution as it allows straightforward incorporation of all of the route and environment attributes to be used by the present research, and forms the basis for known solutions to all of the problems suggested. Therefore, graph based models will form the basis of the research here, although it will be modified to incorporate the spatial information required to overcome the disadvantages that it poses. In this sense it can be considered an extension of the deformed grid, an example of which is shown in Figure 2.1c.

2.5   Route Selection Algorithms

Selecting a route with specific characteristics has formed the basis for a whole field of research spanning many decades. From early generalised approaches such as breadth- and depth-first searches (for an overview see [190] Chapter 3 or [47] Chapter 6), to specialised algorithms for networking [61] and aviation applications [186]. Section 2.3 indicates that the wayfinding algorithms which may be used to select routes for the present research can be divided into the four problem categories; Minimum Cost algorithms, Hamiltonian Path algorithms, Rural Postman algorithms and Travelling Purchaser algorithms. Actu-

ally, the Hamiltonian Path and Rural Postman problems can be both solved by converting them to the widely researched Travelling Salesman problem (of which they are generalised versions), reducing this to three categories. This section will therefore focus on algorithms used to solve these three well known problems, with details of the conversions required between the Travelling Salesman and Hamiltonian Path or Rural Postman solutions and differing approaches which are also applicable to them given. In all cases previous research has suggested far more solutions than can be reviewed here, but a variety of approaches for each will be briefly explained and performances compared where possible. In all cases, the algorithms in this section are designed to be run over a graph G consisting of N nodes (vertices) and L links (edges).

Table 2.5 shows the different algorithm approaches taken by the systems in Section 2.2 which provide relevant explanations, with the majority using minimum cost algorithms. Not all of the systems give details on their base algorithms, and some are far more complex than others. For example, one system [33] uses a combination of fuzzy logic sets and a genetic algorithm, to generate personalised routes.

This highly complex system [33] first processes a linguistic input from the user to determine a fuzzy set of desired route attributes. It then creates a population of paths by genetic coding, creating sentences which represent the nodes required to travel from the start to the destination. This population is then evolved using elite selection, crossover operations and mutation, to form the second generation solutions. This process then continues, with the stopping criteria being determined by processing time, and some measure of 'goodness' in the solution.

In contrast, although the weight-assessing system in this table [40] uses a complex method to create dynamic weights, this resulting set of weights is applied in a similar way to the more straightforward minimum cost approach. The cased based system [147] is also relatively straightforward; using an algorithm which stitches together routes segments which are known to the user (where possible), with A* search filling the spaces inbetween.

Finally the remaining system [176] relies on a decision tree approach, the C4.5 DTL algorithm, to choose the most appropriate route for the user. In essence it does this by forming a tree which represents a series of tests on the supplied attributes, forming splits on specific attribute values. When a set of user requirements are input into the tree these are tested against the attribute values at each split, progressing down the tree to a leaf node which represents a specific route. If the user chooses an alternative route, then this information is used as feedback to the system and the tests are modified as required.

One thing that the four systems described previously have in common, is the required storage of user data, which is beyond the scope of the present research. Referring back

to Table 2.5, the most commonly used type of algorithm which do not record user data are the minimum cost approaches. With that in mind the minimum cost problem will be examined first.

Route recommendation system	Algorithm	
	Type	**Specific**
GUIDE [42]	Minimum cost	Agent based
Deep Map [145]	?	?
REAL [126]	Minimum cost	Dijkstra
Drishti [95, 184]	?	?
RouteCheckr [228]	Minimum cost	Dijkstra
"Simplest" Paths [57]	Minimum cost	Dijkstra
[93]	Minimum cost	Dijkstra
[31]	Minimum cost	Dijkstra
[147]	Case based	Segment reuse
[176]	Decision tree	C4.5 DTL
[40]	Weight-assessing model	Fuzzy logic
[33]	Genetic algorithm	Multiobjective selection
[187]	Minimum cost	Dijkstra

Table 2.5: Summary of the algorithms used by existing route recommendation systems.

2.5.1 Minimum Cost Algorithms

The most basic form of Minimum Cost algorithm is the Uniform Cost search, which uses the cost of reaching a node to determine when or if it will be expanded. However the addition of supplementary data, heuristics and multiple iterations may be used to improve what can be relatively long processing times of this greedy approach, although this increase in speed often leads to the degrading of route accuracy and increasingly more complex algorithms. This section will describe some of the simplest approaches, along with some which are more complex, and a small number of algorithm improvement techniques that can be used to increase speed and accuracy. In general, solutions to the Minimum Cost problem can be divided into two groups - 'label setting' algorithms in which the cost of travelling to a node is calculated only once, and 'label correcting' algorithms which allow costs to be changed if better ones are found during the search. Of the label setting algorithms Dijkstra's Algorithm [54] is the best known, and by far the most widely used in commercial wayfinding scenarios, and will be the only one considered here. However, a number of previous studies have looked into the performance of several different

approaches to minimum cost route selection [123, amongst others]. Six of the label correcting algorithms tested, along with Dijkstra's Algorithm, are detailed below, followed by an examination of their performance for different types of graph and environment.

2.5.1.1 Dijkstra's Algorithm [54]

Dijkstra's Algorithm is a greedy search strategy based on the principle of optimality. It is also the most widely accepted and commonly used solution for finding the shortest path through a positively weighted graph, and will only be briefly outlined here. The traditional algorithm extends the the node with the least existing cost (distance) during each iteration (beginning with the start point), records the cost and predecessor of each neighbouring node, and continues until the destination is found. It then defines the route by working back along the chain of predecessors, starting with the destination and ending with the start point. Many variations of Dijkstra's algorithm have been developed over the years, with most altering the format in which the nodes are stored as they are discovered, in order to reduce the time taken to find the least cost node for each iteration [41, gives an overview of some of these]. In all cases a node is explored only once, leading to a time complexity of $O\left(|N^2|\right)$ for the original algorithm, reduced to $O(|L| + |N|\log|N|)$ if a Fibonacci heap approach is applied.

2.5.1.2 Label Correcting Algorithms

Unlike Dijkstra's Algorithm which establishes which node to expand based on the cost of travelling to each, the A* Algorithm estimates the cost remaining to reach the destination to make this decision [94]. For example, when attempting to select the shortest route between two points, the remaining cost to the destination could be estimated by calculating the Euclidean distance between the discovered node and the end point. Similarly, the angle between the discovered partial route and the destination point may give an estimation of the number of turns required to reach it (although this is far less certain). As with Dijkstra's Algorithm, there are several variations of the A* algorithm, some of which have been specifically designed for wayfinding [81, 109, for example]. The time complexity of the A* Algorithm is dependent on the heuristic implementation and the graph to be searched [74], but as it becomes Dijkstra's Algorithm if each link has an equal length, then it can be considered to the same as described above.

The Bellman-Ford Algorithm [16] is an early example of a label correcting algorithm, which in essence applies Dijkstra's algorithm N-1 times (where N is the number of nodes). As nodes are discovered or their costs are improved, they are added to the back of a FIFO

queue. The process then continues until the queue is empty, and the route is constructed by traversing the chain of predecessors as with Dijkstra's algorithm. This produces an algorithm with $O(|N||L|)$ time complexity.

Incremental Scan Algorithms are very similar to the Bellman-Ford algorithm in their approach, but they handle improved nodes first, rather than using a simple FIFO queue. In the case of the Pape-Levit algorithm [175], newly discovered nodes are placed at the bottom of processing queue, whereas nodes with improved cost are placed at the top (and therefore processed first). Pallottino's Graph Growth algorithm [173] extends this approach further by using two queues, the 'scanned' and 'unscanned' arrays. The unscanned array is processed first (initially containing only the start position), with previously undiscovered nodes placed at the bottom of the scanned array, and nodes with improved cost placed at the bottom of the unscanned array. Once the unscanned array is empty, the contents of the scanned array is copied over, and the process continues.

The Threshold Algorithm [80] combines the advantages Dijkstra and Bellman-Ford algorithms with those of incremental scans. The algorithm uses a two queue approach, with all nodes below a threshold cost being held in a processing queue (labelled NOW) and all others placed in a holding queue (labelled NEXT). Any new or changed nodes are added to the holding queue until the processing queue is empty, at which time any held nodes below the threshold are passed back to be processed again. This ensures that the k best nodes are processed first, giving a time complexity of $O(|N|^2|L|)$

As with the Threshold Algorithm, the Goldberg-Radzik Topological Ordering Algorithm [41] again uses a two queue approach, array A which is linearly ordered and is the processing queue, and array B. Nodes in A are extended in order, and new or improved nodes are stored in B. Once A is empty, then the nodes in B are ordered topologically, and passed back to A. This continues until both A and B are empty, with the algorithm running in at worst $O(|N||L|)$ time.

The Bucket of Queues Algorithm [123] employs an array of buckets, with each bucket containing a single queue. As nodes are expanded, new or improved nodes are placed into the queue in the bucket appropriate to their costs. The nodes are then expanded in order of bucket, starting with the queue in the lowest value bucket and moving to the next bucket once this is empty. The algorithm returns to lower valued buckets if any nodes are placed in them during processing, continuing until all queues and buckets are empty, and giving a time complexity dependent on the number of buckets used.

2.5.1.3 Algorithm Improvement Techniques

A number of different improvements may be added to the majority of the discussed algorithms, from the use of buckets to heuristic selection techniques [41, 123, amongst others]. Bucketing approaches such as Dial's algorithm [52] or heaps such as the *k*-ary heap approach change the storage structures of algorithms to improve search speeds, whereas heuristic approaches reduce the search area itself by estimating remaining cost as in the A* search [123]. Additional improvements can be made by considering the order in which the nodes are actually expanded [41, 123]. Examples of these include Parent Checking [41], where nodes are only extended if their parent is not in the queue (to be in the queue, then the parent must have improved since its child node was added); the Largest Label Last heuristic [23], where nodes with a cost value greater than the average in the queue are automatically moved to the bottom of the pile; and the the Small Label First Algorithm [22], where discovered nodes with costs smaller than those at the top of the queue are moved to the top, and all others to the bottom of the queue.

2.5.1.4 Algorithm Performance

For each of the algorithms discussed, the time complexity given indicates the worst case performance, but not necessarily the actual performance of the algorithm on different types of graph environments. Several studies [41, 123, 235, for example] have investigated these differing performances, and the more applicable results are summarised here.

Initial testing on simple, small, grid based graphs indicate that the shape of the grid can provide contradictory results [41]. For small square grid and long narrow grid graphs, Incremental Scan Algorithms performed best with the Bellman-Ford and the Topological Ordering algorithm performing the worst, but these performances are reversed for short wide grids. For non-grid graphs Dijkstra's algorithm and the Topological Ordering algorithm outperform all other approaches, indicating that these algorithms are more applicable to complex graph based environments [41] .

Studies which examined the performance of algorithms on real road networks [123, 235] suggest that the Pape-Levit algorithm and the Threshold algorithm perform well as does the Bucket of Queues approach. In addition, although Dijkstra's algorithm performs poorly in its most basic form, it is surprisingly competitive if the algorithm employs improvement techniques such as Dial's Bucketing approach [52]. Derivatives of the A* algorithm have also been shown to perform well on real road networks [81, 109], in many cases outperforming all other methods. Although some of the road networks used in these studies are likely to be much larger than those required by the present research, these re-

sults are the most comparable and convincing. To decide between the five best suggested solutions, factors other than accuracy and speed must be considered. In all cases these results were produced for shortest path problems, and as discussed previously, converting positively related attributes may be complicated. This may decrease the efficiency of these algorithms, but this decrease in efficiency is likely to be similar across all five options.

Dijkstra's algorithm has the advantage of being the only exact algorithm proposed and is straightforward to implement, but is restricted to solving graphs containing only positive costs. Although not immediately obvious, the inclusion of positively related attributes such as vegetation may lead to seemingly negative weights, depending on the method of representation used (see Chapter 5 for a full explanation of this). As such, the algorithm will need to be modified before it can be applied. The A* algorithm is again easy to implement, but is not applicable to all of the attributes required. The two examples of length and angle given in Section 2.5.1.2 are by no means the only cost estimations available; however, it is hard to see how a value for the number of junctions or amount of vegetation likely to be encountered along the route could be calculated without providing much more information to the algorithm. The three remaining better performing algorithms also have advantages and disadvantages. All have been shown to provide good route accuracy (if not optimality) and speed, and all can cope with both positive and negative weights within graphs. However, the Threshold algorithm poses the additional challenge of requiring a suitable threshold to be set, and the Pape-Levit and Bucket of Queues algorithms are more complex to implement. Overall the simplicity, flexibility and accuracy of Dijkstra's algorithm mean that it was chosen to be the most appropriate for the wayfinding problems posed by the present research.

2.5.2   Travelling Salesman Algorithms

The Travelling Salesman Problem is one of the most widely researched of all wayfinding problems [7], and the algorithms used to solve it can also easily be converted to offer solutions for Hamiltonian Paths and the Rural Postman Problem. Whether traversing a directed or undirected graph the Travelling Salesman problem is NP-Hard, meaning that there is no polynomial time algorithm available which can solve all instances of the problem. However, there are both exact and heuristic approaches which have been shown to be appropriate for different problem scenarios. As all routes in pedestrian travel can be considered to be undirected, that will be the focus of this discussion. With so much research available, there are numerous algorithms which can be applied to this problem

- far more than can be discussed here. This section will concentrate on some of the best known and widely used solutions, followed by examining how the algorithms can be converted to suit the individual problems and their performance in providing solutions.

2.5.2.1   Exact Solutions

The Travelling Salesman Problem is a minimisation problem subject to a series of linear inequality constraints and as such, the exact solutions to the problem take the form of Linear Programming. In an ideal world, this set of inequalities would be small and fully understood, but this is rarely the case. Despite this, several differing Integer Linear Programming (ILP) Formulations have been suggested [170, contains a survey of these], with the most widely accepted ILP approach to solving the Travelling Salesman Problem being the Graph Cutting Algorithm.

The Graph Cutting Algorithm begins by using a small subset of inequalities to compute an initial optimal solution, and then attempts to identify an inequality which is valid for the problem but violated by the current solution. This new inequality is known as a 'cutting plane', as it seeks to cut the existing graph into valid and invalid solutions. The discovered cutting plane is then added to the original linear programming problem, then the process is repeated until no more inequalities can be found.

Where the Graph Cutting Algorithm produces a non-integral solution, or fails to provide an accurate enough 'optimal' solution, the Branch and Cut Algorithm offers an extension to help overcome these issues [77, 97, 172]. It continues the cutting process by breaking down the search space into smaller subspaces (branching), then finding optimal solutions for these subspaces. If these solutions are found to be globally optimal (applicable to the entire search space), they are used as candidates to be added back into the original linear programming problem. The possible candidate solutions are pruned by defined bounds, which eliminate those which cannot be optimal when compared to others. There are several possible branching strategies including Pseudo Cost Branching, Strong Branching, Enhanced Branching and Most Infeasible Branching [3, 138]. In addition, where non-integral solutions exist these are used as higher bounds, and integral solutions as lower bounds, although this is not the only possible approach. The remaining subspace solutions are then added back into the linear programming and the process is repeated.

As an alternative to exact algorithms, many heuristic solutions to the Travelling Salesman Problem have been proposed to improve empirical performance [146]. Among these are approaches based on local search optimisation [114], spanning trees (such as [72]) and ant colony simulation [56]. All of these algorithms can be divided into two groups -

tour construction and tour improvement approaches.

2.5.2.2   Tour Construction Heuristics

As the name suggests, Tour Construction Heuristics aim to create a good route from scratch. They begin with the start node, then continue to add new nodes (according to the heuristic) until a complete route has been constructed [160]. Four of the many possible options [114, 160, for more examples of these] are discussed below.

Christofides Heuristic [43] consists of four parts - construct a minimum spanning tree, find the minimum cost matching for each odd degree node, combine the tree with these matchings and apply shortcuts to remove repeated nodes. The heuristic was one of the best early heuristics developed for solving the general problem, and is relatively simple to implement. It generally produces a good route, but not necessarily the best, and has a time complexity of $O(|N|^3)$. Despite this, the routes can have up to 3/2 times the length of the optimum for a Hamiltonian cycle, and 5/3 for a Hamiltonian path where both start and end are specified [104].

Nearest Neighbour Heuristic [17, 188] forms a solution by moving from the start to its nearest neighbouring 'city', and so on until all required nodes (cities) have been visited. This heuristic usually performs worse than Christofides Heuristic in terms of route length, but has a lesser time complexity of $O(|N|^2)$.

The Multiple Fragment (Greedy) Heuristic [20, 114] aims to construct partial routes that can be connected to provide a whole. It does this by sorting all links according to their size, and then adding the shortest link to the route if it doesn't create a cycle with fewer edges than the number of cities, increase any node to have a degree of more than two, or already exist in the route. This process is then repeated until a complete route is established. The Greedy Heuristic performs better in terms of route length than the Nearest Neighbour Heuristic, but worse than Christofides Heuristic. Similarly, it has a time complexity of $O(|N|^2 log(|N|))$, which sits between the two previous approaches.

The Clarke-Wright Savings Heuristic [45] begins by calculating individual routes for a central hub to each of the required nodes, and can be used to solve several variations of the Travelling Salesman Problem. It relies on some initial preprocessing of the search space, and then progresses by creating and connecting partial routes until a whole solution has been found. The preprocessing stage calculates the cost of travelling from the hub to each node individually, and the savings that can be gained by travelling between nodes without returning to the hub in between. With these savings placed in descending order it then starts at the top of the list and either begins a new partial route (where neither node occurs in an existing one), adds the pair to an existing partial route (where only one node

already occurs) or merges two partial routes (where the two already contain one node of the pair). This heuristic will give a single route in the majority of cases, but may need to pass through the start point multiple times. Due to this it has problems with inaccuracy, but again has a time complexity of $O(|N|^2 log|N|)$.

2.5.2.3   Tour Improvement Heuristics

Unlike tour construction heuristics, tour improvement heuristics begin with a complete route (such as one created by tour construction methods) and then seek to make it more optimal [160]. As with the construction approaches there are many heuristics available for this improvement, but only three of the most powerful algorithms will be discussed here.

The two-opt method [48] is a well known local search algorithm, commonly used to solve the travelling salesman problem. In its simplest form it begins with a solution (calculated using the nearest unvisited neighbour method for example), then deletes two edges, and reconnects the vertices in an attempt to find a better solution. This continues until all combinations have been computed, and the best found [114]. Similarly, the three-opt method performs the same procedure, but deletes three edges in one move rather than two. The performance of both the two- and three-opt methods can vary in terms of length according to the method used to produce the initial route which is to be improved. For example, with an initial route produced by Christofides Heuristic, they have been shown to produce routes which are no more than 3/2 times optimal in length [114]. One of the main problems with two- and three-opt methods is their time complexity. It has been established that naive implementations of these methods can encounter problem instances which take $O(2^{N/2})$ moves to find a local optimal solution [34]. However, permitting the algorithm to remove only crossing links may reduce this to only $O(N^3)$ (due to triangle inequality) [224].

The Lin-Kernighan Heuristic [137] is one best known and most widely regarded solutions to the TSP, noted for producing good end routes. It is a generalisation of the two- and three-opt methods, but is not restricted to replacing only two or three links. As with the two-opt method the algorithm begins with a pre-generated route and then starting from a fixed point, attempts to form a sequential series of 2-opt moves which reduce the length of the route. Each 2-opt flip must result in a complete route (ie no parts of the route must become detached), the links produced by the flip must not have previously existed in the route (disjoint) and the candidate nodes for each flip are restricted to the k nearest neighbours. Unlike the methods described above, the Lin-Kernighan Heuristic requires only that the cumulative gain from the sequence of flips is positive (ie the resulting route

is shorter than the original). This allows each flip to use links which are longer than the original as long as the total is shorter, therefore avoiding local minima. The complexity of the Lin-Kernighan Heuristic is generally considered to be $O(N^{2.2})$, although studies have shown that the actual worse case scenario can be much greater than this [174]. Several improvements to this algorithm have been suggested (see [114, 185] for examples of these), many based on varying the value of k to increase the neighbours which are possible candidates or the termination criteria.

The Iterated Lin-Kernighan Heuristic [113] is a very powerful extension of the original algorithm, which repeatedly applies the Lin-Kernighan method with the aim of producing better and better results [114]. Beginning with a route generated using the Lin-Kernighan Heuristic, a single random 4 opt flip (double-bridge move) is applied to produce a slight variation on the original solution. This flip produces an effect similar to that of mutation in genetic algorithms, but has the stipulation that the offspring produced must be shorter than its parent. The resultant child route is then used as a new start point for the next iteration of the Lin-Kernighan algorithm, with the best solution so far being stored and only updated if the result of a subsequent iteration is better. The Iterated Lin-Kernighan Heuristic has many varying implementations, but is generally considered to be the most cost-effective way to improve on the Lin-Kernighan approach [114],

2.5.3 Minimum Length Hamiltonian Path Algorithms

The conversion between the Travelling Salesman Problem and the Hamiltonian Path Problem is the most straightforward of those to be considered here. It simply involves the addition of a zero cost link between the start and end points, ensuring the inclusion of this link in the final route. By removing this link after route selection, the cycle is then converted to a path, thus solving the problem. For the graph representing the environment to be searched, only the start and end points and points representing high value areas need to be included, thus greatly reducing the complexity of the graph. To perform this reduction, it is required that the shortest paths between each point to be generated, although for all but the start and end point these links can be precomputed. The resulting simplified representations should produce small, relatively dense graphs. Using a real world example 20 nodes, each representing a single point of interest, would be considered a large number of locations to visit in one day but would produce quite a simple graph on which to perform route selection.

An important consideration when considering the performance of the Travelling Salesman algorithms discussed on the Minimum Length Hamiltonian Path problem, is the size

of the graph to be searched. In the context of the present research, the graphs required for the Minimum Length Hamiltonian Path problem are likely to be so small that it may be feasible to use the Graph Cutting or Branch-and-Cut exact methods, which would give the most accurate results possible. However, these approaches may require complex code, have large memory requirements and possibly long processing times [131]. In addition, there are no studies available which directly compare ILP and heuristic approaches to finding Hamiltonian paths, making it difficult to judge whether the increases in accuracy are worth the increased processing time. Heuristic solutions are susceptible to losses in accuracy, but are usually more straightforward to implement and run faster. With this in mind, only the performance of the heuristic approaches from Section 2.5.2.2 and Section2.5.2.3 will be discussed in any detail here.

For tour construction heuristics, the performance of each of the algorithms varies according to the size and type of graphs on which they are tested. Given the small size of the graphs that can be generated in this instance, standard test cases may be considered sufficient to assess performance. In a summary of experimental results [114], suggests that Christofides Heuristic and the Clarke-Wright Savings Heuristic produce slightly better routes in terms of optimality than the Nearest Neighbour Heuristic or the Multiple Fragment Heuristic, although all perform well for small graphs. In terms of running times the Nearest Neighbour Heuristic outperforms the remaining three construction algorithms, with Christofides Heuristic taking the longest to select suitable routes, but again all perform well.

For tour improvement heuristics, the tour construction heuristic which was used to construct the original solution also plays a part in the algorithm performance [114]. The two- and three-opt methods produce slightly better results than the tour construction heuristics in terms of route optimality, with the Multiple Fragment Heuristic providing the best initial solution for producing good results from both approaches. However, the three-opt method increases the time taken to produce a solution is compared to the two-opt method, and both will obviously take a substantially longer time than any of the tour construction heuristics required to produce a start route for improvement. The Lin-Kernighan Heuristic outperforms the three-opt method substantially in terms of route length on the test instances used in one study [114], and also increases the running time by only a small amount for small graphs, suggesting that it is a better solution for these cases. Assessing the performance of the Iterated Lin-Kernighan Heuristic is slightly more complicated, with the ability to vary the number of iteration allowed creating a trade-off between running time and route optimality. On a small number of nodes, the Iterated Lin-Kernighan Heuristic can be shown to reduce the route length increase to half that of the standard

Lin-Kernighan Heuristic, although due to its iterative nature it will clearly take much longer.

In addition to route accuracy and running time, algorithm complexity should also be considered when determining which approach is likely to be best for the wayfinding problem posed. For the present research, both versions of the Lin-Kernighan Heuristic would be much more complex than required, suggesting that the two- or three-opt methods may be the most appropriate. However, even Christofides Heuristic is likely to produce acceptable results on such small graphs.

2.5.4   Rural Postman Algorithms

The Rural Postman Problem is a variant of the Travelling Salesman Problem, where each required link represents a city. Creating a graph for which the algorithms discussed in Section 2.5.2 can solve the Rural Postman Problem is slightly more involved than for Hamiltonian Paths, but is still relatively straightforward. The nodes within the graph are reduced to those which are attached the ends of the required links, the required links and then an arc which represents the shortest path is added between each pair of nodes to connect them through the remaining (so far unrepresented) graph. Figure 2.4 shows an example of this process.

As with the Travelling Salesman Problem, exact solutions exist for the Rural Postman Problem. Successful Integer Linear Programming formulations have been suggested by three previous investigations [44, 46, 136], and although they will not necessarily solve all instances of the problem, they have been shown to be reliable in the majority of cases.

Alternatively, heuristics may be based on spanning trees (such as [60, 72]), local search [90, 96, give examples of these] or Monte Carlo principles [63] amongst others. In addition to the heuristic approaches given in Section 2.5.2, probably the most notable tour construction solution to the Rural Postman Problem is Frederickson's Heuristic [72]. This heuristic first constructs a minimum spanning tree, then uses a matching procedure to replace series of edges connecting odd-degree vertices, with arcs representing the minimum distance routes between them. This reduced graph can then be used to create a suitable route.

Several variations of the two and three-opt methods have been proposed [63, 96, for example], although the majority offer only high complexity solutions (up to $O|N|^5$ in the case of [96]). Both the traditional two and three-opt methods adjust only the order in which the vertices but one approach suggests that it may be more optimal for the Rural Postman Problem to also adjust the direction in which the edges are travelled [90]. It has
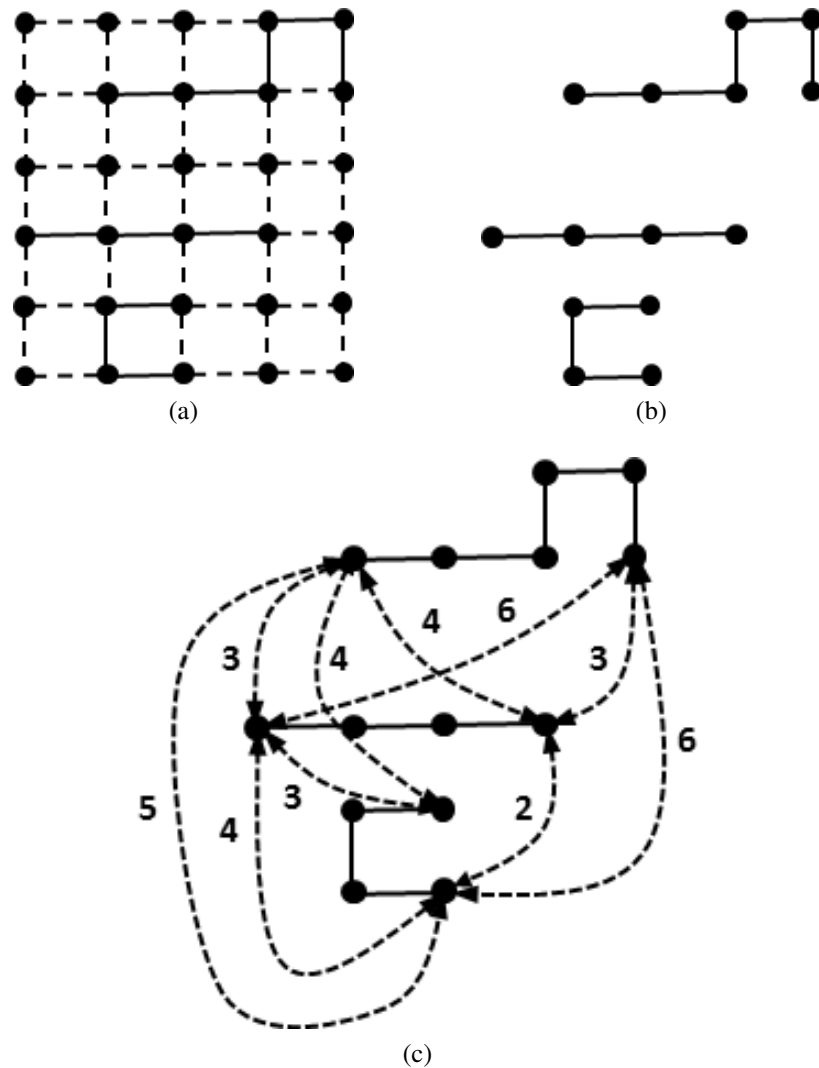
Figure 2.4: Three graph representations of the Rural Postman Problem conversion process, (a) the original graph (solid lines are required links, dotted lines the links which aren't required and each link has a length of one), (b) the required links and associated nodes only (each link has a length of one), and (c) the new graph complete with added links (dotted lines show the arcs and each unlabelled link has a length of one).

also suggested a way to reduce the complexity of the algorithm, and therefore the time required to run it, by applying a minimum cost approach similar to Dijkstra's Algorithm [90]. In this, the costs for the previous solution are used to determine if the new solution will indeed be better, by ceasing when it reaches a cost which is higher.

Unlike the Minimum Length Hamiltonian Path Problem where only a small number of cities are likely to be required, the graphs produced to represent the required areas in this case are likely to be much larger and more complex. Other than the study discussed in Section 2.5.3 and two additional experimental analyses of the Travelling Salesman

algorithms [115], there is very little indication on how these algorithms would perform on graphs representing real world environments when solving the Rural Postman Problem.

For the exact solutions, a Branch-and-Cut approach has been shown to be successful [77]. However, the maximum test graph used is only 350 vertices, which is much smaller than would be expected in many real-world situations. This would suggest that the use of heuristic methods is likely to be more appropriate for the wayfinding problem here.

Examining the performance of the Travelling Salesman tour construction approaches on larger graphs [115], the Clarke-Wright Savings Heuristic again performs well in terms of route quality, but the Nearest Neighbour Heuristic outperforms this in terms of speed. In terms of tour improvement approaches, the two-opt local search technique can generate routes relatively efficiently and accurately in large graphs [90, 115], as can the Lin-Kernighan and Iterated Lin-Kernighan Heuristic [115]. In this case however, the use of the more complex Iterated Lin-Kernighan Heuristic is more likely to be justified, with the variable iterations which create a trade-off between speed and accuracy bringing flexibility.

2.5.5 Travelling Purchaser Algorithms

As with the Rural Postman Problem, the Travelling Purchaser Problem is a generalisation of the Travelling Salesman Problem, and can easily back to it by reducing market to selling only a single product. Unfortunately this property is less useful for this type of problem [27], as none of the algorithms discussed in Section 2.5.2 can be applied directly to generate a solution. However there are several possible approaches, some of which are modifications to those suggested for the Travelling Salesman problems, and others which incorporate similar principles. As with all of the problems discussed previously, these can be broken into two groups - exact algorithms and heuristics - some of which will be briefly described in this section.

2.5.5.1 Exact Solutions

Three exact solutions have been found for the Travelling Purchaser Problem - two of which are different to those discussed in Section 2.5.2, and one that is a slight modification of an approach already described. One of the earliest exact solutions to be proposed was a lexicographic algorithm [183], which uses an alphabet table of words which represent each possible combination of 'markets' (nodes). Each of these possible solutions is then check for feasibility, and the least cost route is chosen. Unfortunately this approach is in no way efficient, and is rarely used.

The Branch-and-Bound Algorithm [202] proposed for the solution of the Travelling Purchaser problem is very similar to a standard solution for the Simple Plant Location Problem, which is then used to solve subsets of the original problem. These smaller problems are then used to calculate lower bounds (see [202] for a full explanation of this), and these bounds are in turn used to calculate a single optimal solution. The branching techniques which can be used are the same as those in the Branch-and-Cut solution to the Travelling Salesman Problem (see Section 2.5.2.1), and the process continues until only one solution remains.

In addition to the two new exact solutions described above, a Branch-and-Cut Algorithm [132] very similar to that in Section 2.5.2.1 has also been formulated. This uses the same techniques as discussed previously, with a new ILP formulation being used. As with the Travelling Salesman Problem, these exact solutions are generally inefficient, and can only be used on small simple graphs at best. Fortunately, several specialised heuristics have been developed to overcome these difficulties.

2.5.5.2 Heuristic Solutions

There have been many heuristic solutions proposed for the Travelling Purchaser Problem from ant colony optimization [28], to tabu [79] search strategies [227]. This section will concentrate mainly on the most commonly used approaches, to restrict these suggestions to a sensible number.

The Generalized Savings Heuristic [82] begins with an initial route which contains only the start point, and the 'market' (node) 'selling the most products' (containing the most variables) at the least cost. In each iteration, the unvisited market with the maximum saving (travel and purchase costs) is added to the route. This continues until all products have been included, and no more savings can be made.

As with the tour improvement algorithms suggested for the Travelling Salesman Problem, the Tour Reduction Heuristic [165] starts with a route containing a subset of markets (typically those selling at least one product at its lowest price) which covers all of the required products. It then iteratively attempts to drop a market from the route if it produces a cost saving, and continues until no savings can be made.

The Commodity Adding Heuristic [178] takes a list of all of the required products, and constructs an initial route which includes only the start point and the market which has the minimum cost for the first product in the list. The algorithm then adds the least cost market for each product in turn, assuming that they do not already exist within the route. The algorithm terminates when all of the products have been included in the solution, and the route is complete. Variations of the algorithm are based mainly on the order of

addition of products from the list, including random-order and sequence-order additions.

The Market Adding Heuristic [132] is very similar to the Commodity Adding Heuristic, but here the insertion criteria are determined by the maximum savings approach. In addition, the product can be any which has not already been included, rather than the addition order determined by the generated list. The approach was originally designed to be applied to fractional solution of linear programming [132], but when additional algorithm improvement techniques are applied such as market drop and market exchange (see Algorithm Improvement Techniques below), it was found to be a feasible standalone solution.

Perturbation heuristics [27] are post-optimisation techniques which are performed on existing solutions, with random changes applied in much the same way as mutation in genetic algorithms. The proposed algorithms remove x markets, and randomly insert new markets according to the cheapest insertion criteria. During each iteration improvement techniques are applied to the resulting route, and the process repeats until a specific number of iterations have shown no reduction in the solution.

2.5.5.3 Algorithm Improvement Techniques

In addition to preprocessing approaches [213, for example], four techniques are commonly used to optimise the routes generated by the majority of heuristic algorithms:

1. Market drop - the removal of any market in the route which can generate a total cost reduction.

2. Market add - the addition of any market to the route which reduces the travel cost, reallocation of product purchase to minimise the buying costs, and removal of any markets in which no products are purchased.

3. Market exchange - the iterative removal of a market from the route and performing market add steps until a feasible, lower costing, solution is found.

4. TSP heuristic - the application of a Travelling Salesman heuristic to the existing route in an attempt to shorten it.

2.5.5.4 Algorithm performance

Although there are many approaches to solving the Travelling Purchaser Problem there are few studies which compare more than two approaches, and none which directly compare exact and heuristic algorithms. Exact solutions to the problem however, tend

to be inefficient and may become unwieldy for all but small simple graphs [178]. As such, the results of three computational studies which compare only heuristic approaches [27, 178, 213] will be used to examine algorithm performance here.

A performance comparison study using randomly generated test examples [178] suggested that an implementation of the Commodity Adding Heuristic (Random-Order Commodity-Adding) outperforms all of the tested Generalized Savings and Tour Reduction heuristics, along with additional Search Algorithm implementations (not detailed here), in terms of accuracy across a range of graphs (10-50 markets and 5-60 products). It also reaches these solutions substantially faster than any implementation of the remaining three approaches.

With larger numbers of markets and products (up to 200 of each) [213] the Generalized Savings Heuristic performed better than the others in terms of route accuracy, but again the Commodity Adding Heuristic performed best in terms of processing time. Adding preprocessing techniques has been shown to improve these values, but not to affect the relative ordering. Many of the graphs examined in this study are far greater in size and complexity than would be required by the present research, and therefore the preprocessing would be unlikely to be required.

A further study [27] compared the Market Adding Heuristic, Commodity Adding Heuristic and two variations of Perturbation Heuristic. This suggested that the Perturbation Heuristic outperforms all those tested for 50 to 350 markets and 50 to 200 products. Unfortunately the precise details of the algorithms used are not included in the paper concerned, but from the overview, are complex.

In conclusion, for the size and type of graph which would be produced by the environment and variables in the present research, the above research suggests that the Random-Order Commodity-Adding Heuristic would probably prove to be the most appropriate Travelling Purchaser solution in this context.

2.5.6 Summary

This section has briefly described some of the possible solutions available for the wayfinding problems outlined in Section 2.3, their performance, and identified the most suitable for the specific issues posed by this thesis. It has suggested the four following solutions, stemming from the three algorithm categories examined (where the Minimum length Hamiltonian path and Rural postman solutions are differing Travelling salesman algorithms):

- Minimum cost problem - Dijkstra's algorithm.

- Minimum length Hamiltonian path problem - Two-opt method.

- Rural postman problem - Iterated Lin-Kernighan heuristic

- Travelling purchaser problem - Random-order commodity-adding heuristic.

Each of these approaches offers an appropriate trade-off between route accuracy, processing time and algorithm complexity for the wayfinding problems described. In addition they all offer a measure of flexibility, allowing the present research to consider a number of differing attributes. Despite these advantages, only one approach can form the basis of this thesis. In addition each wayfinding problem has its own benefits and drawbacks, the most important of which must also be considered.

The Hamiltonian path, rural postman and travelling purchaser approaches allow all attributes to be considered in a relatively straightforward way, with positive attributes being represented as required locations, and negative ones being the cost of reaching these. However, this also introduces its own drawback, with the two types of attribute being treated in very different ways. Each of the three problems also offers the ability to reduce the complexity of the graph being examined, either by reducing the number of nodes to only those representing the required locations, or by reducing the total environment to only the required links and those required to travel between them. However this also introduces the need to define these locations (with the addition of finding routes within them in some cases), determining appropriate thresholds for designating areas to be of 'high attribute value' and the need for environment preprocessing with associated time increases.

The minimum cost approach does allow all attributes to be treated in the same way, although converting positively related attributes brings additional complexities. It does not allow for graph simplification, but does remove the need for preprocessing and the definition of which areas should be included or omitted. Overall the simplicity and flexibility of this wayfinding problem and its solution were the main influencing factors when choosing it for the basis of this thesis.

2.6 General Summary

This chapter has briefly discussed the commercially available route suggestion systems, and determined that the majority rely on only route length to choose between available options. Although this is appropriate for generating the shortest routes required to travel from one place to another, it is unlikely to produce routes which are suitable for different types of journey. The present research aims to use environment and route attributes to produce more suitable routes for varying journey types, by combining these attributes

into an individual algorithm for each types of journey. In order to do this, four areas of research are required

1. Attribute influence on route choice - The previous research discussed in Section 2.1 has been used to select nine attributes relating to route simplicity and attractiveness (length, turns, decision points, initial leg length, vegetation, parkland, points of interest and dwellings), chosen to be underlying criteria for route selection. Each of these attributes are expected to influence the choice of routes for three journey types (everyday, leisure and tourist), and although the order of importance for some of these attributes and journey types is know, for others it has yet to be investigated. Chapter 3 gives details of a series of experiments which were designed to either confirm, extend or establish the known importance ranks in order to successfully use these to select routes for each journey type.

2. Construction of an environment model - In Section 2.4 three possible environment representations have been discussed. The graph model was chosen as the most appropriate for the present research, being simple enough to be applicable to many geographical areas, but detailed enough to allow route selection. Chapter 4 will give details of how a suitable environment model was constructed from available data sources, and how each of the chosen attributes was incorporated into this model.

3. Development of suitable route selection algorithms - A number of the possible wayfinding problems that the present research could address have been discussed in Section 2.3, and Section 2.5 has considered some of the many possible algorithms which are available to solve them. Of these a minimum cost approach, based on Dijkstra's Algorithm, has been chosen due to its relative performance, flexibility and ease of implementation. Chapter 5 and Chapter 6 describe how algorithms were developed firstly for each of the individual route and environment attributes, and then how these attributes were combined to form algorithms which can select routes for each of the proposed journey types.

4. Suggested route evaluation - Finally, the aim of this thesis is to produce suitable routes for each of the journey types, and therefore offer alternative solutions to those available in existing commercial systems considered in Section 2.2. Chapter 7 discusses a user study which allowed comparisons to be made between the routes suggested by human participants, and those selected by the developed algorithms.

Chapter 3

Influence of Environment and Route Attributes on Route Preference

Chapter 2 (Section 2.1) has explained that a number of route and environment attributes may influence route choice for different types of journey, but in order to successfully construct a route selection engine for multiple journey types, the influence of environment and route attributes must first be ascertained. This chapter begins by hypothesising about the influence of those attributes on route preference. It then discusses the methods which may be used to establish influences, and compares these in the context of the present research. The remaining sections describe a series of experiments designed to examine the influence of each of the chosen attributes (turns, decision points, landmarks, initial leg length, vegetation, land use, points of interest, dwellings and length), firstly as part of the simplicity or attractiveness of a route, and then for everyday, leisure and tourist journeys. Experiment 1 and Experiment 2 were devised to establish the validity of the experiment approach, by reproducing the results seen in earlier research, and to provide a rank for the attributes within the attractiveness category (vegetation, land use, points of interest and dwellings). Experiment 3 and Experiment 4 then examine the influence of attributes on each journey type (everyday, leisure and tourist), creating ranks in each case. Finally, Experiment 5 and Experiment 6 examine the influence of length on everyday, leisure and tourist routes, and integrate these findings into the previously established ranks.

3.1 Experiment Hypotheses

Table 3.1 shows the list of hypotheses which are developed from the findings previous research. The rationale behind each hypothesis is given below, grouped according to the attribute category or journey type to which they refer.The experiments within this chapter are arranged in pairs, with the first experiment establishing whether or not a set of attributes have an effect on route choice for either an attribute category or journey type. The second experiment then determines the rank of influence of the attributes found to be influential by the first. The numbering of the hypotheses reflects their placing within these experiments, and therefore the order in which they were tested.

Hypothesis		Description	Exp.
H1	✓	Simplicity will increase as the number of decision points or turns required at decision points decreases, and as the number of landmarks or the length of the first leg increases.	**Exp 1**
H2	✓	Attractiveness will increase as the amount of vegetation increases, as land use shifts from urban to farm and from farm to park, as the type of dwellings changes from multiple occupancy dwellings to single occupancy dwellings and then to historic or large dwellings or as the number of points of interest increases.	**Exp 1**
H3	✓	Simplicity will be most influenced by landmarks, followed by decision points, then turns at decision points and finally the length of the first leg.	**Exp 2**
H4	✓	Attractiveness will be most influenced by land use, followed by vegetation, then points of interest and finally dwellings.	**Exp 2**
H5	✓	The preference for routes for everyday journeys will be increased by a lower number of decision points and turns at junctions, by the presence of farm or park land and by routes with longer initial leg lengths.	**Exp 3**
H6	✓	The preference for routes for leisure journeys will be increased by the presence of park or farm land, single occupancy or large dwellings, vegetation and points of interest.	**Exp 3**
H7	✴	The preference for routes for tourist journeys will be increased by the presence of park or farm land, single occupancy or large dwellings, vegetation and points of interest.	**Exp 3**

Hypothesis		Description	Exp.
H8	✓	Everyday journeys will be most influenced by the number of turns, followed equally by land use and the number of decision points and finally initial leg length.	**Exp 4**
H9	✳	Leisure journeys will be most influenced by land use, followed by vegetation and then dwellings.	**Exp 4**
H10	✓	Tourist journeys will be most influenced by points of interest.	**Exp 4**
H11	✓	The preference for routes for everyday journeys will be increased by shorter route lengths.	**Exp 5**
H12	✳	The preference for routes for leisure journeys will be increased by longer route lengths.	**Exp 5**
H13	✓	Everyday journeys will be influenced by route length more than any other attribute.	**Exp 6**

Table 3.1: Hypotheses for perceived route simplicity, attractiveness and preference according to journey type. Previously shown effects are indicated by (✓), and effects which are inferred from previous studies are marked (✳).

The relationships between levels of the same attribute, or between different attributes, are well tested for some of the hypotheses such as those associated with simplicity (H1 and H3) and which attributes are important for attractiveness (H2). However in other cases existing data for relationships are more scarce, for example those associated with leisure and tourist journeys (H9 and H7).

3.1.1 Hypothesis H1 and Hypothesis H3 - Attribute Influence on Route Simplicity

The attributes which affect route simplicity have been studied by a wide range of previous research [12, 83, 102, 142, 163, 164, 191, 199, and many more]. This research indicates that an increase in number of landmarks, or decrease in the number of decision points, number of turns and initial leg length all improve route simplicity as shown in Hypothesis H1.

Despite this wealth of research indicating that these attributes all affect route simplicity, there is very little evidence about the relative importance of these factors required for Hypothesis H3. The strongest argument is for the importance of landmarks indicating that, although these visual cues are most influential at decision points [110], they have been shown to decrease wayfinding errors [110, 152, 189] on even labyrinthine routes with many decision points. This reduction in errors implies an increase in simplicity, and leads

to the assumption that landmarks will be of more importance than either decision points or turns. Empirical evidence [210] also suggests that the number of decision points in a route should outrank the number of turns in terms of influence on simplicity. Even though the difference in perceived length discovered by the previous study is small, it does indicate an increase in mental load and therefore complexity. If it is assumed that participants will select the simplest route when choosing from alternatives between two points on a map, then the number of turns along this route will be of more importance than the length of the initial leg [83]. Although this assumption may seem like a leap, it is the basis for including the longest-leg-first attribute in the present research, and will be maintained here. This leaves only the order between the number of intersections and the length of the first leg to be predicted. Here there is no firm indication of importance, but it will be assumed that intersections will have more influence due to the differences in perceived length being only comparatively minor. These findings all contribute to Hypothesis H3.

3.1.2 Hypothesis H2 and Hypothesis H4 - Attribute Influence on Route Attractiveness

In contrast to simplicity, several studies [120, 121, 156, 180, 203, 222, for example] have suggested that a very different range of attributes affect the attractiveness of a route as shown by Hypothesis H2. Increases in the amount of vegetation, parkland, large or single family dwellings or the number of points of interest have all been shown to increase preference for photographic scenes, and would therefore be expected to play a part route attractiveness.

Information on how the influences of these attributes compare used to form Hypothesis H4, comes from three main studies [120, 180, 222]. One, an investigation into responses to visual landscapes [222], suggests that land use has a greater influence than vegetation as country views are considered more attractive than urban views, even those with vegetation. According to a study looking at scenic quality [120], vegetation is more important than points of interest, and these in turn are more influential than dwelling types, which is confirmed by research into environmental factors that affect walking and cycling [180].

3.1.3   Hypothesis H5 and Hypothesis H8 - Attribute Influence on Everyday Route Choice

Route choice for this type of journey has been shown to be affected by the number of turns and decision points encountered, aesthetics and the length of the route's initial leg, and addition to the route length (see Hypothesis H11 and Hypothesis H13) [12, 83, 102, 199, for example] as indicated in Hypothesis H5. One complication is the definition of aesthetics as discussed in Section 2.1, however landuse has been indicated to have a positive influence on everyday routes [171].

The evidence for the order of influence of the above attributes on everyday journeys required to establish Hypothesis H8, is somewhat mixed. To some extent the order of importance of these attributes for everyday journeys is clearly defined, with number of turns outranking all others, followed by aesthetics and finally the longest leg first criteria. However, as turns and decision points were equivalent in much previous research, and aesthetic quality is composed of many different characteristics, in some cases the influence is not so easily determined. A review of the environmental attributes which affect walking [171] indicates that the most important of characteristic of aesthetics for walking is land use. The preference for shorter route lengths for commuting and other regular trips is confirmed by the work of several other researchers [134, 197, amongst others], and as the number of decision points encountered and turns made increase the perceived distance between two points [210], it would seem likely that these too would have a role leading to Hypothesis H8.

3.1.4   Hypothesis H6 and Hypothesis H9 - Attribute Influence on Leisure Route Choice

The attributes of attractiveness are all expected to be important for leisuure journeys [78, 106, 134, 134, 151, 193, 207], suggesting Hypothesis H6. Land use is considered the most influential in walking for pleasure or exercise [134] and, as walkability [193] is thought to influence this type of journey, vegetation, dwellings and points of interest are also considerations [78, 106, 134, 151, 207].

Unfortunately, these studies fail to show significance for the order of the attributes. What evidence there is suggests that vegetation and dwellings are less important than land use [134], and that vegetation is preferred to dwellings [151] (Hypothesis H9). As points of interest are not mentioned in either of these studies, their influence cannot be established.

3.1.5 Hypothesis H7 and Hypothesis H10 - Attribute Influence on Tourist Route Choice

For tourist trips, points of interest are expected not only to be influential (Hypothesis H7), but also to be the most important factor when selecting routes [205]. Aesthetics can be added to the characteristics affecting this type of journey [158], with dwellings and land use being suggested as influential factors [225]. Several studies have indicated that vegetation may also sway the directions taken during this type of travel [66, 145, 225], and combined these findings suggest Hypothesis H7. Despite agreement that points of interest will have the most influence on tourist routes (Hypothesis H10), only a very vague indication of the rank of the others is given in any related literature.

3.1.6 Hypotheses H11, H13 and H12 - Influence of Length on Routes for Different Journey Types

One attribute missing from the discussion of leisure and tourist journeys so far is route length. Many studies have suggested that route length not only affects choice for everyday routes, but that it is the most influential attribute involved in this decision [83, 134, 197, amongst others] (Hypothesis H11 and Hypothesis H13). However for leisure journeys, the influence of route length is much less clear. Although a number of studies indicate that length may play a part in route choice for walking for exercise or leisure [120, 171, for example] (Hypothesis H12), the relationship between this attribute and the route chosen is not investigated in any detail. In addition, no previous research has suggested how influential route length is on choice for leisure journeys. Furthermore, for tourist journeys, the relationship between length and route preference is even less well defined.

3.2 Overview of Methods for Evaluating Preferences

Many different fields require the determination or evaluation of people's preference, from marketing and economics to travel and health care. Despite the widely varying fields in which preference is tested, only two approaches to this testing are commonly used - stated preference and revealed preference [177, 216].

Revealed preference uses participants' real-world actions, such as travel information or shopping habits, to determine preference [216]. One major advantage to revealed preference is its accuracy in reflecting the actual choices of participants, but the real-world nature of this approach limits the attributes and attribute combinations which can be tested,

and a lot of data must be collected in order to determine preference [216]. These issues can lead to revealed approaches suffering from insufficient variation between choices, and difficulty in understanding the importance of each of the tested attributes.

In contrast, the stated preference approach determines preference by directly questioning the participant [216], such as choosing from a variety of possible packing options or environment scenes. Stated preference overcomes the drawbacks of revealed preference alternatives, by allowing the researcher to design trials which have as many attributes and variation as required. It also allows for preference estimation without real-world data; however, one substantial drawback is that what people say is not necessarily the same as what they do [216].

For route and environment attribute evaluation, both revealed and stated preference could be used to elicit participant choices, but the practicalities of not needing large amounts of real-world data overcome the disadvantages of the stated preference approach, which is why it will be used in this thesis. Within the stated preference approach, there are several techniques which could be used, dependent on the choices to be made, or available information. Four of the most appropriate - contingency ranking, preference choice, contingency rating and paired comparisons - are briefly outlined below.

Contingency ranking requires respondents to rank a set of alternative options [92], such as ranking photographs of environment scenes. It can be considered a sequential choice process in which the participant chooses their most preferred option first, remove it from the pool and repeat until no more options are available. Contingency ranking can produce more statistical data than approaches such as choice experiments, but choices are not independent and can be unreliable and inconsistent across ranks due to complexity and cognitive effort required. This type of preference testing has been used previously in areas such as evaluating forest landscape attributes [76], assessing environmental determinants [180] and choosing between road options [144].

Preference choice offers participants two or more options via images or descriptions, and asks them to state which of these would be preferred [92]. It has been used to elicit responses to alternative scenarios in many different fields, including travel choice [141], bicycle route selection [107] and to evaluate walkability [119]. Choice experiments rely less on the accuracy of available information and displaying an entire set of options at once [29] than other approaches, and offer a more flexible approach which relates directly to many tasks experienced in real-world situations.

Contingency rating is very similar to contingency ranking in that respondents are presented with a set of alternative options, but here they rate each of these alternatives on a standard scale [92]. It is one of the most attractive approaches in many fields, and has

been widely used in previous research (examples include determining the scenic quality of roads [120,121], assessing environmental determinants [106,180] and finding perceptions of the neighborhood environment [78]). Contingency rating has the same advantages and disadvantages as contingency ranking, although using a rating approach can give much more statistical information. However, ratings may not be comparable across individuals, and an assumption of the cardinality of scales is required.

Paired Comparisons combine the rating technique of contingency rating, with the small comparison requirement of choice experiments. Participants in paired comparisons are asked to give a numerical indication of rating between two alternatives [92], removing the need to display all of the options at once, but retaining many of the same advantages and disadvantages as contingency rating. Paired comparisons are not as commonly used as alternatives described above, but there are examples where it has successfully determined user preference [201, for example]

Although any of the described approaches could be used for the evaluation of route preference, within this user study constraints such as the lack of complete information (not all attributes which could affect the decision are being used) can affect which is the most appropriate. With a flexible and simple approach, no requirement for complete data, and the availability of advanced statistical analysis, choice experiments represent the best option in this case.

In addition to the varying techniques available to determine user preference, there are a number of ways of depicting and therefore presenting alternative routes, from route descriptions and sketch maps to complex virtual environments. Not all convey a suitable amount of information to make an informed choice between two or more routes. Route descriptions and sketch maps are methods for transmitting spatial knowledge that is used every day (for example the response given to asking a stranger for directions in the street, or the written depictions given by pressing the 'directions' button on Google Maps [85]) but neither contains enough detail to enable route choice. However, maps, photos, videos and virtual environments can all be used in a way that enables route preferences to be found.

Maps provide a simple approach to environment representation, providing a static scene which can be made to show the entire area at a single glance, communicating knowledge to the participant without the need for it to be discovered [24]. Selecting a route from a two- or three-dimensional map is one of the most common techniques that people use to plan travel in an unknown area, and has been used many times previously for testing user preference, [12,83,143, amongst many others]. Two-dimensional maps have been shown to be better for understanding relative position [208], whereas three-dimensional maps

give more understanding to the shape and layout of terrain [208], and symbolic representations may be used successfully on maps for wayfinding tasks [73, 122]. Compared to other approaches, maps take a relatively short time to create and require little participant training [50], but map reading ability varies across individuals [108] (with some participants struggling to locate landmarks, or determine routes through three-dimensional spaces). In addition two-dimensional (and to some extent three-dimensional) maps may lead to biases against certain attributes such as land use, when compared to others such as length due to visibility [220].

Although photographs may be considered real-world map representations, they can be used to display more detail than traditional maps, and can remove some of the disadvantages associated with map reading [73]. For example aerial photos retain spatial relationships, but may remove some of the biases present in maps by making attributes such as vegetation more visible. In addition, photographs of salient landmarks may also aid orientation and situatedness [15, 51], improve perception of location, and assist in wayfinding [142, 199, and many more]. Combining this with two-dimensional maps can also make visual cues more identifiable [51]. However aerial photographs can be difficult to comprehend and pose problems for determining location in some cases [204], and many photographs may be required for even small environments.

Video or other moving visual recordings have some of the advantages of photographs, such as improving situatedness and perception of location [15, 51] by giving access to salient landmarks, but also disadvantages such as requiring long recordings to cover even small areas. In addition, unlike aerial photographs, videos are normally taken at ground level, removing much of the spatial information required for route choice. However, one major advantage of this type of representation is the ability to visually show a walkthrough of the users' real-world experience.

Interactive Virtual Environments allow traversal along a route by creating a three-dimensional model in which the participant moves at human eye-level through the environment. They are an increasingly common approach to empirical testing in fields such as psychology [103] and geographical visualization [124] as they provide a quasi-realistic test platform and a sense of presence for the user. Movement through an environment allows for discovery of information [24] in much the same way as if it had been encountered in a real-world situation, and research between interactive virtual and real environments have shown that valid results can be obtained for the main aspects of environmental perception [25] and path choices [234]. They have a higher level of presence than traditional maps [192], if factors such as involvement and immersion [196] are considered. However, the quality of models and textures used within virtual environments may have a

direct impact on realism and route choice [26, 130]. In addition initial creation of the environments can be time intensive with existing models such as Google Streetview [85] being limited in some areas. They may also require extended participant training periods, require participants to use working memory for comparisons, and some may experience 'cybersickness' [133].

Static Virtual Environments represent a snapshot of a virtual environment, typically viewed from an elevated angle. They provide a static scene which can be made to show the entire route, including start and destination points, in a single glance. Static virtual environments have many of the advantages of maps and photographs as mentioned previously, and textured models have been found to elicit similar responses to real images in perceptual evaluation [161]. Compared to the virtual environments required for walkthroughs, static rendered images take a relatively short time to create, and again require little participant training as most people can easily understand the contents of high quality images [50]. However, there is no indication that the preference decisions made are comparable to those made in the real-world [26], and there may be less presence experienced than in an interactive virtual environment [192].

Each of the approaches to displaying possible route choices described, has its own advantages and disadvantages. For the present research, the use of static virtual environments was chosen, in order to exert the greatest possible control over the attributes being tested, and to reduce the amount of participant training required. The static computer generated scenes used in the experiments here are based on three dimensional maps, also known as Worlds in Miniature [211], shown as pairs of routes in static scenes. These contain buildings and objects that are either mapped with photorealistic textures, or are modeled to appear similar to real-world examples. Ground cover materials have been chosen to be unambiguous and distinguishable, to avoid issues with recognition, and the possible routes are shown with little or no occlusion.

3.3   Experiment Method

A total of six experiments were devised, each based on a similar set of materials and procedures. In each experiment participants were presented with a scenario, shown a pair of routes, and then asked to choose their preferred route for the specified scenario.

The first two experiments asked the questions:

Simplicity: 'Which route is the simplest?'

Attractiveness: 'Which route is the most attractive?'

and were designed to establish the validity of the images, static 3D environments and overall experimental design. In order to achieve this, the experiments aimed to reproduce the findings of earlier research, establishing that each of the tested attributes were of the expected importance and rank position within their respective attribute category.

Experiment 3 and Experiment 4 aimed to establish if the same attributes are important for three different types of journey, and determine previously unknown ranks using the following questions:

Everyday Travel: 'Which route would you choose for everyday journeys? - This could be walking to work or uni, or if you were just popping out to the shops.'

Leisure Travel: 'Which route would you choose for leisure journeys? - This could be walking for pleasure or exercise, so say you were going for a stroll.'

Tourist Travel: 'Which route would you choose for tourist journeys? - Say you were visiting campus for a short time and wanted to explore the area, or you were taking a visitor on a tour of Leeds.'

Finally, Experiment 5 and Experiment 6 used previous three questions to test if length was also important for each of the journey types, and insert it into the correct rank position.

3.3.1 Materials

Pairs of routes were shown side-by-side in single images connected to common start (bottom) and end (top) points as shown in Fig. 3.1. Each route varied by a single attribute relating to either its simplicity or attractiveness, and the images were counterbalanced by alternating which side attributes were shown on. The routes were constructed using Autodesk^R 3ds Max^R 2012 (14.0 student stand-alone version) [10], combined to create environments, and rendered to a 640x480 jpeg image file.

Within each experiment, the routes were based on identical but mirrored basic layouts, with features added according to the attribute being illustrated (see Appendix A for a full set of the route images used). For attributes which required an increase in the number of elements of this type (points of interest, turns, decision points and vegetation), a single feature was added for level one, three features were added for level two and five features were added for level three. These features were selected to be typical examples of structures or elements commonly encountered in urban areas, with churches, water features, statues and public buildings chosen as points of interest, and trees, hedges and flower beds
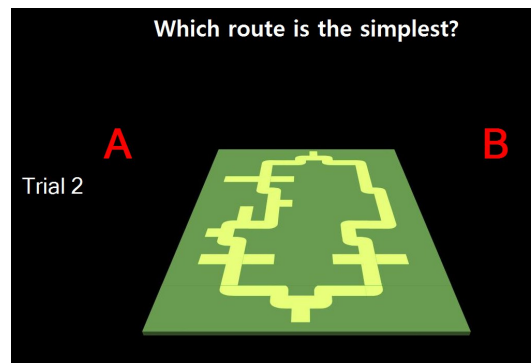
Figure 3.1: Example experiment 1 environment. Category - simplicity, left route - decision points level 5, right route - decision points level 1.

representing vegetation. For example, Fig. 3.1 shows two levels of decision points with the number of decision points being increased to lower the simplicity of the route.

The type rather than amount of land use or dwellings have been shown to affect attractiveness, which is reflected in the levels of these attributes. Multiple occupancy housing is significantly less preferred than any other form of dwelling, whereas historic homes are more preferred [120], and these each form a level for the dwellings attribute. Land use is harder to portray in a single image of these dimensions, especially without using images of housing. To prevent confusion or misunderstanding, ground coverings showing paving (urban), a ploughed field (farmland) and grass (parkland) were selected. Wherever possible, overlaps between feature types were avoided. However, buildings being shown as both dwellings and points of interest, and grass (which could be considered vegetation) used for parkland were considered acceptable, as similar overlaps would exist in real world environments.

For Experiment 1, Experiment 3 and Experiment 5 different levels of the same attribute were compared to establish their importance for the attribute category or journey type presented. For the remaining experiments, the most preferred levels of different attributes were compared to establish their rank positions in this importance.

3.3.2 Procedure

Each experiment used a different group of University staff or students as participants, mostly in large cohorts within a teaching situation. Recruitment initially involved contact with a member of teaching staff to acquire their permission to run an experiment either before or after their lecture. All potential participants were then contacted by email to explain the nature of the experiment, provide an experiment information sheet, and give details on how they could opt out of participation. Where these large groups gave biases in

terms of gender or age, these were addressed (where possible) by secondary recruitment through poster advertising and smaller group sessions. It was felt that a large number of participants would give confidence in the results (for example 43 participants would give a confidence interval of approximately 15%), and that the experiment was appropriate for large scale simultaneous testing, which were the main reasons for choosing this approach to recruitment.

Each recruited individual was provided with a second copy of a participant information sheet and a multiple choice form, to provide further information and collect their preferences (see Appendix B for examples of these). The form gave spaces for participants to record their gender and age, but they were instructed to not write their name anywhere on the sheet. In addition, two boxes were provided for each screen, including those in the training phase, one marked 'A' and one marked 'B'. During the experiment participants were asked to mark the letter corresponding to their preferred route in each trial on this form. The experiments were all checked for validity by a psychologist, approved by the Faculty Ethics Committee, and informed consent was provided by participants returning the completed multiple choice forms.

The experiment was divided into two phases; a training phase and a test phase that together took a total of approximately 10 minutes. During the training phase, instructions were provided to the participants, and up to eight screens were displayed in succession, one illustrating each of the pairs of attributes or levels to be tested. Questions were displayed above the routes relating to different attribute categories or journey types (Figure 3.1), and they were asked the appropriate questions. Participants were then asked to indicate that they had expressed a preference for a route before moving on to the next pair. A small number of images were then shown at the same interval as those in the test, to enable participants to acclimatise to the speed with which they would have to make a selection.

Once this phase was complete the test algorithm was run, with the question order and sequence of up to 36 images (3 per attribute, with up to two layouts, for each of the questions) being randomly selected. Between each image a simple sound was played indicating that the next was being shown and a black screen indicated that the trial was over. The completed sheets were collected at the end and the participants were free to leave.

The total experiment time, number of images rerun at test intervals during training, time for which image was displayed and the number of images used was dependent on the experiment being run. These variations were established to be appropriate during a short pilot study, and through informal discussions with participants after their session.

For example, participants indicated that sessions of more than ten minutes may mean a loss of concentration (due to boredom or confusion over many repetitions of the same attribute combinations), that reducing the display time for each image would be appropriate (information provided by participants of both the first and second experiments) and that comparing different attributes was considered to be easier than comparing routes of differing lengths.

3.4 Experiment 1: Which attributes affect attractiveness and simplicity?

This experiment investigated whether the level of a single attribute has an effect on how simple or attractive a route is judged to be. The experiment compared four attributes for each of simplicity (landmarks, decision points, turns and initial leg length) and attractiveness (vegetation, points of interest, land use and dwellings), at three different levels. A within participants design was used, with each participant shown 24 pairs of routes for ten seconds each, and asked to state which route in each pair they preferred.

3.4.1 Participants

A total of 63 individuals (8 females; 55 males) aged 18 to 48 (mean 23.2 years, SD 6.6 years, 1 withheld) participated in this experiment, and all were either University students or members of staff.

3.4.2 Results and Discussion

Any blank forms submitted were rejected as were those where participants had given their name, but the decision over the use of partially completed sheets was determined by the percentage of usable answers. In this case, one submission contained only 74% of the expected responses and was therefore removed, but all others were considered sufficiently complete. Submissions with no answers for the training questions are not included, as no conclusion could be reached as to whether or not they had completed the appropriate preparation for the task.

The votes for each screen were gathered, and combined to give a single value for each route per participant. The Friedman test for k related samples was chosen as the most appropriate non-parametric test to analyse the data. This test ranks k ordinal samples from a population according to their overall differences [200], but does not require the data to

have a normal distribution. This was used to establish rank with a significance level of $p < .05$. A Wilcoxon's Signed Rank post-hoc test with a value of $p < .017$ (Bonferroni corrected), which gives more weight to attributes having a large difference between their conditions, was also performed on the outcomes of the pairwise comparisons.

The results of this analysis are shown in Fig. 3.2, indicating that there is a statistical significance between all levels of decision points, turns, dwellings and points of interest, and statistical significance between the majority of levels of land use and vegetation (as predicted in Hypothesis H1 and Hypothesis H2). However, landmarks and longest leg first attributes show little or no statistical significance between levels. These results are compared to the predicted effects in Table 3.2.
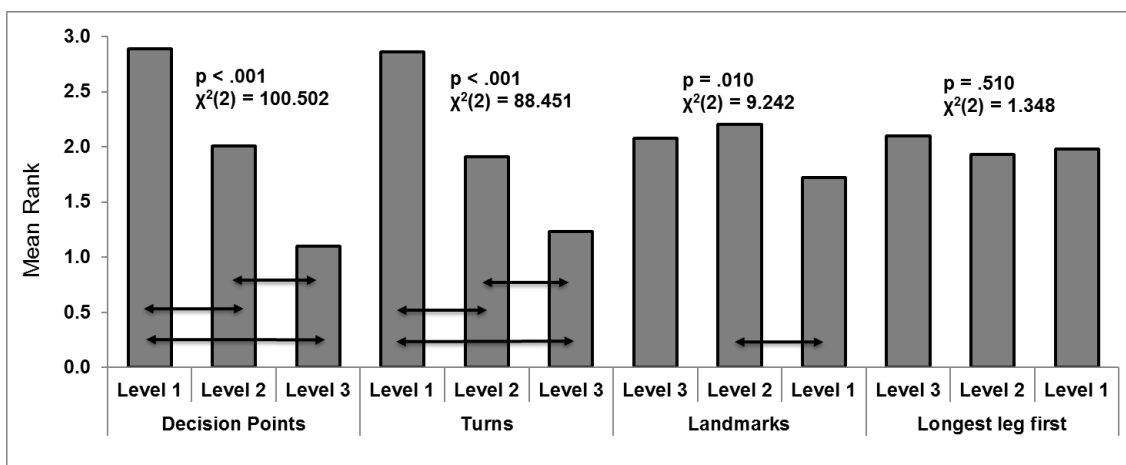
(a) Simplicity Attribute Results
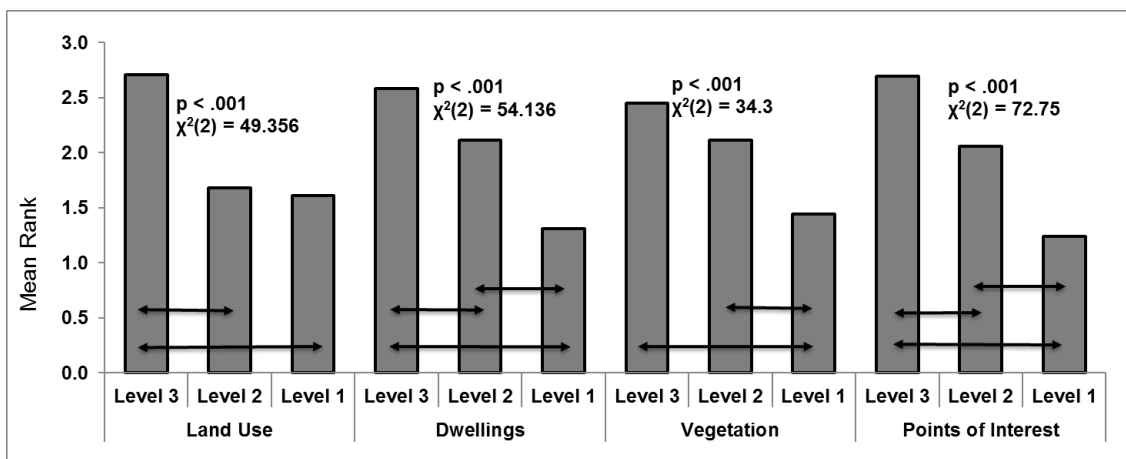
(b) Attractiveness Attribute Results

Figure 3.2: Experiment 1 Friedman Test Results (rank and p values) - pairwise (Wilcoxon's) statistical significance is indicated by the arrows overlaid on the plot.

From Figure 3.2a, routes are considered simpler if they have fewer decision points or

Attribute	Simplicity		Attribute	Attractiveness	
	H1	**Measured**		**H3**	**Measured**
DPs	✓	✓	Land Use	✓	✓
Turns	✓	✓	Dwellings	✓	✓
Landmarks	✓	✓	Vegetation	✓	✓
LLF	✓	✗	POIs	✓	✓

Table 3.2: Predicted (hypotheses from table 3.1) vs measured attribute effect for each attribute category. Previously shown effects (✓) or no reported effect (✗) are compared against the effects found in experiment 1. (POIs - points of interest, DPs - decision points, LLF - longest leg first)

turns as predicted in Hypothesis H1, mirroring the results of previous work [191, 210]. Landmarks initially affect simplicity in line with pre-existing research [142]; however, a plateau or dip is seen as the maximum value tested is reached. This may show that a threshold value exists, and that adding landmarks to a scene above a particular density may be perceived to escalate the visual clutter, and therefore exaggerate the route's complexity.

Figure 3.2b indicates that the attractiveness of a route increases as the amount of vegetation rises, as the number of points of interest increases, and as land use changes from urban or farmland to parkland, all of which are as predicted in Hypothesis H3 and have been indicated by previous work [120, 121, 180]. A deviation in the predicted trend does occur between farm and urban routes; however, this may simply be due to the images that were selected to represent these, and participants' dislike of scenes containing bare earth [203]. Finally, attractiveness increases as dwellings change from multiple occupancy buildings to single occupancy, and then large or historic dwellings as predicted by scene quality tests performed previously [120, 121].

One surprising result from these tests is the lack of effect for the longest-leg-first comparisons, with all images being considered to be of a similar complexity. A likely explanation is associated with the format of the images used for this experiment. In order to vary the length of the initial segment without affecting the other attributes of simplicity, the remainder of the path was moved backwards and therefore compressed. By decreasing the length of subsequent segments, the overall complexity of the route may be considered to be increased. Although an earlier study [12] used two-dimensional maps to test for the existence of the initial segment strategy, the representations used did not affect the length of the final segment, or the number of turns required to complete the route.

3.5 Experiment 2: Order of influence on attractiveness and simplicity

As with the Experiment 1, this experiment investigated the influence of each attribute, but here they were compared against each other rather than varying the levels of the individual factors. It compared fewer attributes than Experiment 1 with three being considered for simplicity (landmarks, decision points, turns), as the results of the previous experiment indicated that initial leg length should be excluded, and the same four attributes for attractiveness (vegetation, points of interest, land use and dwellings). The most preferred levels of each simplicity attribute from Experiment 1 were compared as were the least preferred, and the same process was repeated for each of the attractiveness attributes. Again a within participants design was used, but each participant was now shown 36 pairs of routes (ten seconds for each image) and asked to state which route in each pair they preferred.

3.5.1 Participants

A total of 60 individuals (28 females and 30 males, 2 withheld) participated in this experiment. They were aged 19 to 53 (mean 21.8 years, SD 5.8 years, 1 withheld), and all were either University students or members of staff.

3.5.2 Results and Discussion

As before, any blank forms or those without answers to the training questions were rejected, and as the minimum percentage of completed responses was 94%, all submissions were considered sufficient. Unlike the first experiment, the collected votes were combined to give a single value per attribute for each participant, including those for both tested levels. Again the Friedman Test was used to establish rank with a significance level of $p < .05$ and the Wilcoxon's approach was used for post-hoc testing, but here a value of $p < .01$ was selected for the pairwise comparisons to correct for multiple comparisons and make analysis easier. The results of this analysis are shown in Figure 3.3.

To divide the attributes into ranked lists, we started with the Friedman rank, and then split this into groups where there was Wilcoxon's significance between adjacent attributes. These groups are indicated by the dashed lines on Figure 3.3, and the resulting ranks compared against those predicted in Hypothesis H2 and Hypothesis H4 are shown in Figure 3.4. This confirms some relative rankings found in previous research (the exception being vegetation versus land use for leisure journeys) and also highlights the previously unknown importance of other attributes, the lack of statistical significance between several

attributes, and the marked difference to the predicted influence of others.

(a) Simplicity Attribute Results   (b) Attractiveness Attribute Results

Figure 3.3: Experiment 2 Friedman Test Results (rank and p values) - pairwise (Wilcoxon's) statistical significance is indicated by the arrows overlaid on the plot. Dashed lines show where statistical significance divides the results into a rank order.

For simplicity, landmarks were shown to have the most impact, with their presence indicating the most simplistic routes as predicted in Hypothesis H2. This corresponds to the assumption that landmarks have a marked effect on the simplicity of a route [152], and may explain the reduction in wayfinding errors when these features are present [110,189]. Turns at decision points were found to rank next for simplicity, contradicting the findings of previous work [191, 210], and the reasons for this are uncertain. Decision points rank lowest for simplicity, but the significance of these results indicate they do reduce the simplicity of a route as expected.

From these results, dwellings did have the least influence on attractiveness as expected [120], but the order of the other attributes is different from those predicted in Hypothesis H4. Land use was expected to have the largest effect on how attractive a route is perceived to be [222], but here vegetation was shown to have a greater impact and points of interest were ranked equally. Conversations with participants after this experiment indicated that they may have been considering adverse weather conditions (such as rain) on surfaces when making decisions. This was unexpected as the images were intended to simulate a sunny day, and it was not anticipated that participants would give judgments using alternative scenarios.

Figure 3.4: Experiment 2 Ranks. Predicted rank (hypotheses table 3.1) is compared to actual rank (right), and arrows show movement within the ranks. (POIs - points of interest, DPs - decision points)

3.6 Experiment 3: Which attributes affect route choice for different journey types?

This experiment investigated whether the level of a single attribute has an effect on which route is chosen for different journey types. The experiment compared two levels of six attributes (land use, dwellings, vegetation, points of interest, decision points and turns), for three journey scenarios (everyday, leisure and tourist). Here only the most preferred level of each attribute as found by experiment 1 is included, and landmarks were excluded as they overlap with the points of interest attribute within the attractiveness category. A within participants design was used, with participants being shown 36 pairs of routes each showing different levels of an attribute. They were then asked to state which route in each pair they preferred, and the results analysed to give an overview of how the attributes affect route selection for different journey types.

3.6.1 Participants

A total of 73 individuals (19 females and 53 males, 1 withheld) participated in this experiment. They were aged 18 to 25 (mean 19.3 years, SD 1.5 years, 1 withheld), and all were either University students or members of staff.

3.6.2 Results and Discussion

Of the 73 submissions the minimum percentage of completed responses was 98%, which was considered sufficient to include all participants in the analysis. The votes for each screen were gathered, and combined to give a single value for each route per participant. The Friedman test for k related samples was again chosen as the most appropriate, to establish rank with a significance level of $p < .05$. A Wilcoxon's Signed Rank post-hoc

test with a value of $p < .01$ was also used as previously. The results of this analysis are shown in Figure 3.5, indicating that there is a statistical significance between the different levels of all attributes for everyday and leisure journeys, and all but turns and decision points for tourist trips. These differences are compared to the predicted effects in Table 3.3. As in earlier research, for turns and decision points there is a negative relationship between attribute level and preference, and for all other attributes there is a positive relationship.

(a) Everyday Travel Results

(b) Leisure Travel Results

(c) Tourist Travel Results
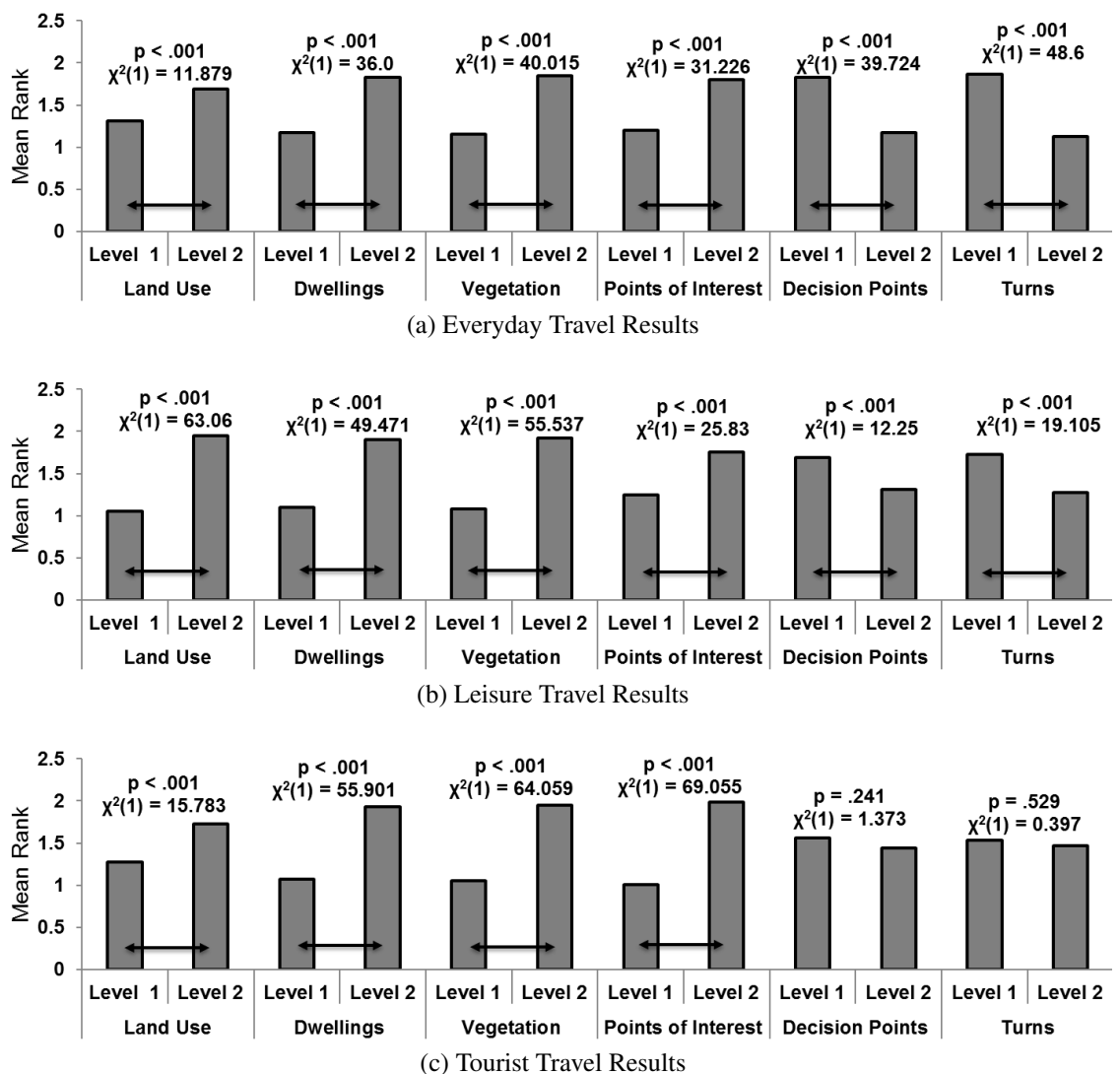
Figure 3.5: Experiment 3 Friedman Test Results (rank and p values) - pairwise (Wilcoxon's) statistical significance is indicated by the arrows overlaid on the plot.

For everyday journeys, the attributes (land use, decision point and turns) predicted by Hypothesis H5 and previous research [83, 171, 210] did have an influence on route choice. However, vegetation, points of interest and dwellings also had an effect. Previous

Attribute	Everyday		Leisure		Tourist	
	H5	Measured	H6	Measured	H7	Measured
Land use	✓	✓	✓	✓	✳	✓
Dwellings	?	✓	✓	✓	✳	✓
Vegetation	?	✓	✓	✓	✳	✓
POIs	?	✓	✓	✓	✳	✓
DPs	✳	✓	?	✓	?	✗
Turns	✓	✓	?	✓	?	✗

Table 3.3: Predicted (hypotheses from table 3.1) vs measured attribute effect for each journey type. Previously shown effects (✓), effects which are inferred from previous studies (✳), unknown effects (?) or no reported effect (✗) are compared against the effects found in experiment 3. (POIs - points of interest, DPs - decision points)

studies [83, for example] discuss a wayfinding criterion termed 'aesthetics', but give little or no indication of the specific attributes being included in this category. The results found by this experiment show that all of the tested attractiveness attributes affect route choice for this type of journey, rather than just land use as predicted.

All of the attractiveness attributes (points of interest, vegetation, land use and dwellings) affect route preference for leisure travel as predicted in Hypothesis H6 and found previously [151]; however, both turns and decision points were also found to influence route preference for leisure journeys, which was unexpected. This may indicate that a more complex route is preferred when walking for pleasure or exercise.

In contrast to the other two journey types, for tourist trips only the attributes predicted in H7 influenced route preference. Points of interest, land use, vegetation and dwellings all affected choice as in previous studies [206, 225], but turn and decision points played no part in the decisions. This indicates that simplicity plays no role in the choice of appropriate routes for tourist travel, with attractiveness being the sole source of influence.

These results indicate that to automatically select routes for these types of pedestrian journeys in the real world, the level of each of the tested attributes should be considered. However, before an automated route selection algorithm can be designed, the relative importance of each of these attributes needs to be determined.

3.7 Experiment 4: Order of attributes' influence for different journey types

This experiment was designed to determine the order of influence of the attributes tested previously in Experiment 3, on routes for the three types of journey. Stated choice experi-

ments were again used, but this time mixed factorial design was employed; with attribute as a within participant factor and journey type as a between participant factor. The experiment again compared six attributes (land use, dwellings, vegetation, points of interest, decision points and turns) for three journey types (everyday, leisure and tourist), but unlike the previous experiment the two routes displayed within each of the virtual environments contained two attributes, allowing comparison between them.

To make the experiment more manageable for participants, but still investigate all of the required attributes simultaneously, this experiment was divided into three separate test conditions. The first test examined the everyday journey type, the second leisure journeys, and the third tourist journeys.

3.7.1 Participants

A total of 169 individuals (90 females and 75 males, 4 withheld) participated in this experiment. They were aged 18 to 53 (mean 23.8 years, SD 7.6 years, 4 withheld), and all were either university students or members of staff. They were divided into three groups with 55 participants for everyday journeys, 54 for leisure journeys, and 60 for tourist journeys.

3.7.2 Results and Discussion

Although the Friedman test suggests a rank for each journey type, the Wilcoxon's results indicate that the order is not as clear cut as it could be. To divide the attributes into ranked lists, we started with the Friedman rank, and then split this into groups where there was Wilcoxon's significance between adjacent attributes. Finally, these groups were subdivided at any other points of statistical significance, as shown by the dashed lines on Fig. 3.6. The resulting ranks compared against those predicted in Hypothesis H8, Hypothesis H9 and Hypothesis H10, are shown in Fig. 3.7, which confirms some relative rankings found in previous research (the exception being vegetation vs. land use for leisure journeys) and also highlights the previously unknown importance of other attributes.

For everyday journeys, turns were found to be more important than decision points and land use, which had equal influence, as predicted by Hypothesis H8. However, all three occurred lower in the ranking than expected, due to the additional attributes found in Experiment 3. Vegetation ranked higher than all other attributes for everyday journeys, which is somewhat surprising, as is the joint second place of points of interest. These results indicate that attractiveness attributes have a larger influence on routes for journeys of this type than expected [83]. Although this may indicate that people look for

different things when choosing routes for everyday travel than previously found, it may also suggest that participants struggle to envision the task being asked of them. Comparisons with real-world routes would be required to investigate these suggestions, although consideration should be given to the differences between testing using real and virtual environments.

Vegetation is preferred to dwellings for leisure travel as predicted in Hypothesis H9
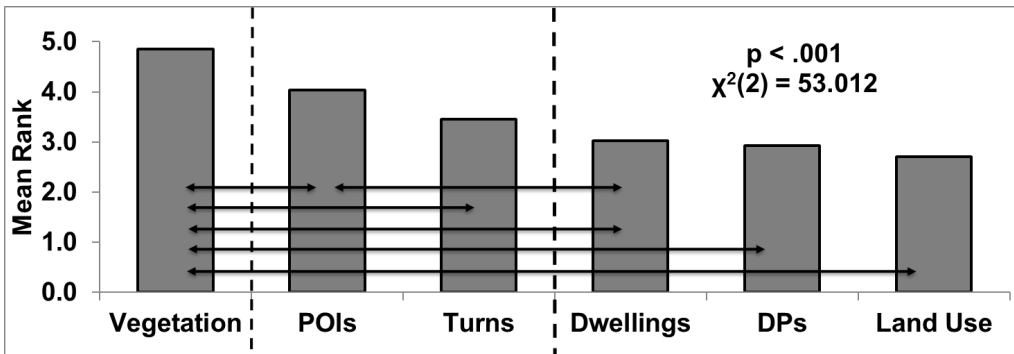
(a) Everyday Travel Results
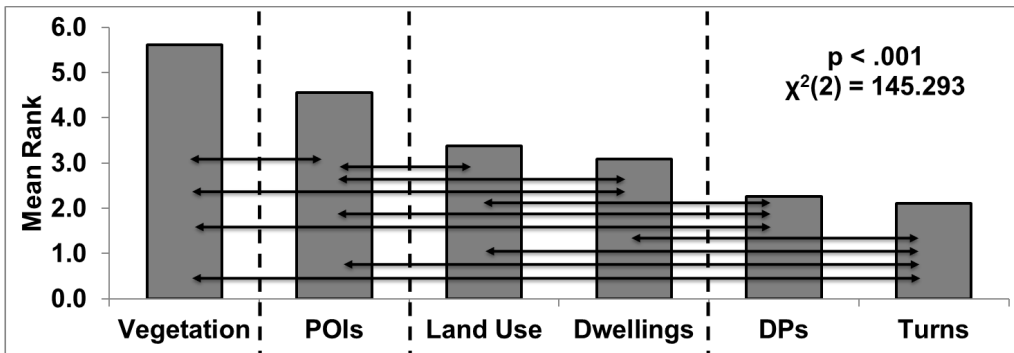
(b) Leisure Travel Results
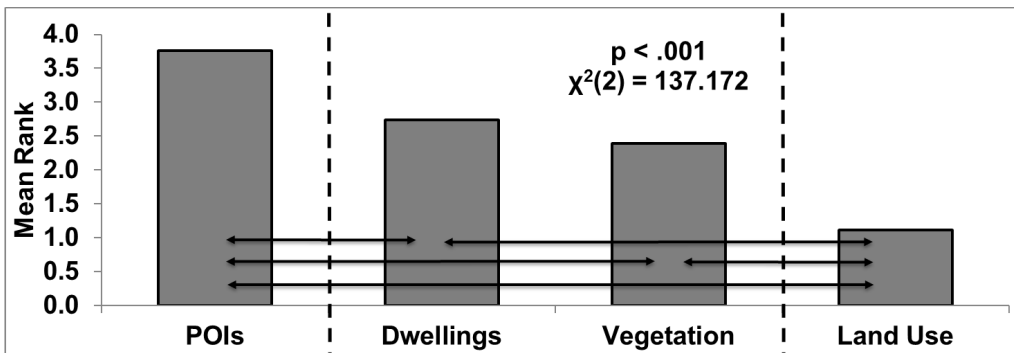
(c) Tourist Travel Results

Figure 3.6: Experiment 4 Friedman Test Results (rank and p values) - pairwise (Wilcoxon's) statistical significance is indicated by the arrows overlaid on the plot. Dashed lines show where statistical significance divides the results into a rank order.
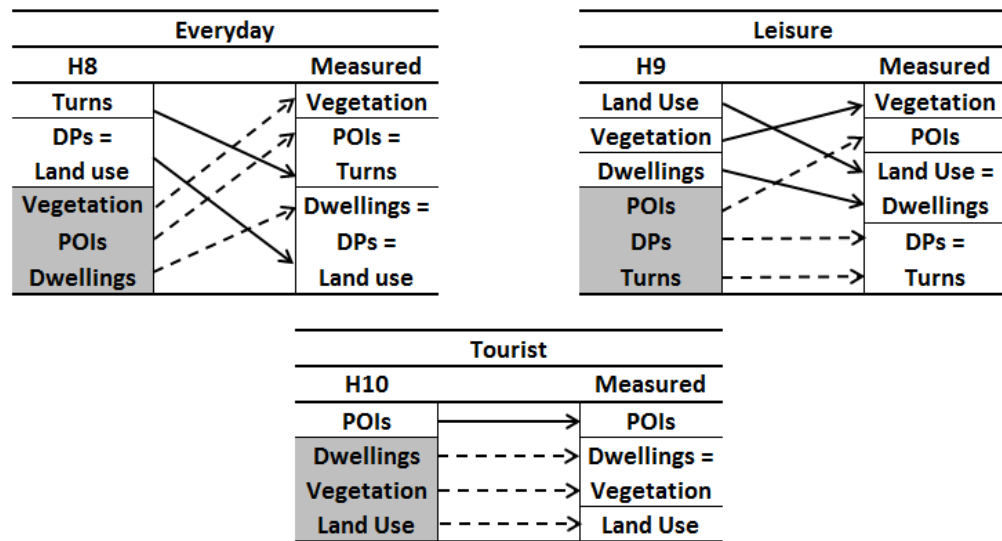
Figure 3.7: Experiment 4 Ranks. Predicted rank (hypotheses table 3.1) is compared to actual rank (right), and arrows show movement within the ranks. Grey predicted boxes and dashed lines indicate that these attributes were initially unranked. (POIs - points of interest, DPs - decision points)

and found previously [151]; however, land use ranks only equal third rather than first as expected [134]. This may be a reflection of the images used for this attribute, and discussions held after the experiment indicated that participants had considered other factors such as weather when selecting a route. The ranks of points of interest, turns and decision points had not been predicted by earlier work as they were not previously tested. Although the relative influence of turns and decision points are not predicted in Hypothesis H9 they may have been assumed to follow those found in everyday routes (Hypothesis H8), but the equality in the rank of these attributes indicates that this is not the case.

As predicted in Hypothesis H10, points of interest play the most important role in the choice of routes for tourist travel [206], but the order of the remaining attributes was not suggested by previous work. Dwellings ranked equally next, as did vegetation. As both points of interest and dwellings may actually be destinations as well as being attributes of the environment for this type of journey, this result is not that surprising. What is more unexpected is that land use ranks last for this type of journey, although this may also be explained by participants considering outside factors as in leisure trips.

3.8 Experiment 5: Does length affect route choice for different journey types?

This experiment investigated whether length has an effect on which route is chosen for three journey types (everyday, leisure and tourist). A within participants design was used, with participants being shown 18 pairs of routes each showing three different lengths. They were then asked to state which route in each pair they preferred, and the results analysed to give an overview of how length affects route selection for different journey types.

3.8.1 Participants

A total of 50 individuals (31 females and 19 males) participated in this experiment. They were aged 19 to 44 (mean 21.4 years, SD 4.0 years), and all were either university students or members of staff.

3.8.2 Results and Discussion

All 50 individuals provided responses to 100% of the required trials, so all participants were included in the analysis. The votes for each screen were gathered, and combined to give a single value for each attribute per participant. The results from the Friedman test ($p < .05$) and Wilcoxon's tests ($p < .01$) for this experiment are shown in Fig. 3.8, which indicates that there is statistical significance between all length levels for everyday journeys, and between the majority of length levels for leisure and tourist journeys.

Figure 3.8a suggests that there is a negative relationship between length and route preference for everyday journeys, meaning that shorter routes would be preferred for this journey type as found in previous research [83, 134, 197, amongst others] and predicted in Hypothesis H11. For leisure journeys, Figure 3.8b suggests that there is a positive relationship between length and route preference as indicated by [120, 171, for example] and Hypothesis H12, but it also suggests that this increase in preference may subside as the length increases. This may imply that there is an optimum length for leisure journeys, although this was not tested. Figure 3.8c shows that length also has a positive relationship with route preference for tourist journeys, which has not been seen previously, but in this case there is no abatement in this preference as the length increases.

3.9 Experiment 6: Extended order of attributes' influence for different journey types

This experiment was designed to determine the rank of length compared to the six attributes tested previously (land use, dwellings, vegetation, points of interest, decision points and turns), on routes for the three types of journey (everyday, leisure and tourist). Stated choice experiments were again used, with a within participant design. For this
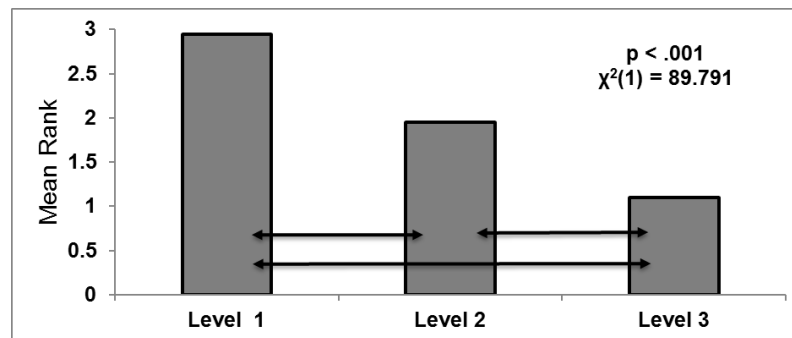
(a) Everyday Travel Results
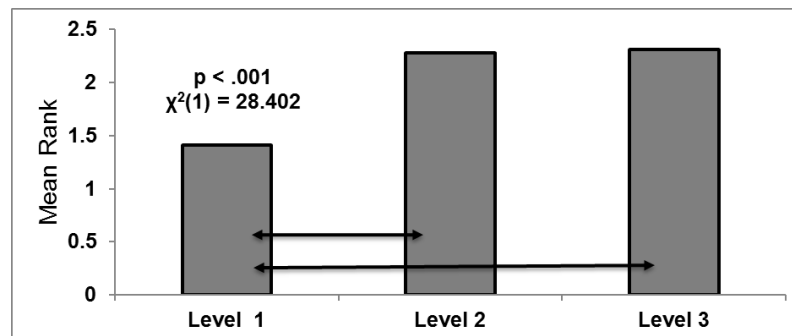
(b) Leisure Travel Results
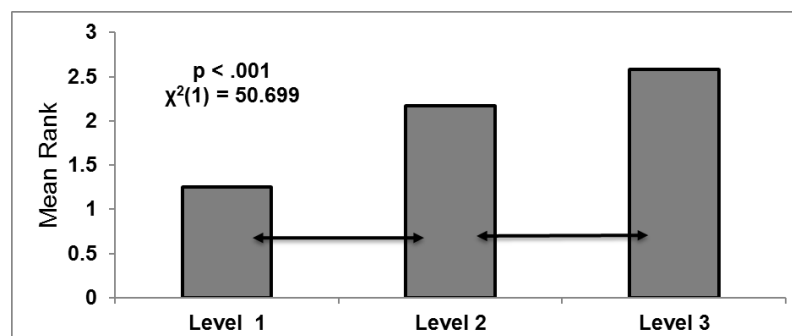
(c) Tourist Travel Results

Figure 3.8: Experiment 5 Friedman Test Results (rank and p values) - pairwise (Wilcoxon's) statistical significance is indicated by the arrows overlaid on the plot.

experiment however, the routes varied by length and one other attribute. The aim of this was to place length in the attributes ranks found by experiment 4, rather than create new ranks, and the design of the experiment was restricted accordingly.

3.9.1 Participants.

A total of 35 individuals (8 females and 27 males) participated in this experiment. They were aged 18 to 52 (mean 20.9 years, SD 5.8 years), and all were either university students or members of staff.

3.9.2 Results and Discussion

The minimum percentage of completed responses was 97%, so all 35 submissions were included in the analysis. The votes for each screen were gathered, and combined to give a single value for each attribute per participant. The results from the Friedman test ($p < .05$) and Wilcoxon's tests ($p < .01$) for each part of this experiment are shown in Fig. 3.9.

A change in the relationship between the length and attribute means (from positive to negative or vice versa) indicates the suggested rank position of length, such as in Figure 3.9b If no change is observed then length will rank highest if the length mean is consistently higher (as in 3.9a), or lowest if the length mean is consistency lower (as in 3.9c).

For everyday journeys, length was found to have the most influence of all of the tested attributes, corresponding to the results of previous research [83] and as predicted by Hypothesis H13, which suggests that the shortest route is the most preferred for this type of journey. For leisure journeys, where no previous research is available on the rank of length compared to other attributes, length is ranked equal to land use and dwellings. Finally, for tourist journeys, where there is again no previous evidence for the rank of length, it is ranked lowest. None of these results are particularly surprising, but together they form full rankings for the attributes on which this research is based.

3.10 Conclusions

The aim of this study was to determine how selected attributes affect route selection for two different attribute categories (simplicity and attractiveness) and three types of pedestrian travel (everyday, leisure and tourist journeys). Unlike previous studies, this research investigated how seven environment and route attributes (length, turns, decision points, vegetation, land use, dwellings and points of interest) affect the preference for a route simultaneously, providing a direct comparison between them for each journey type. These
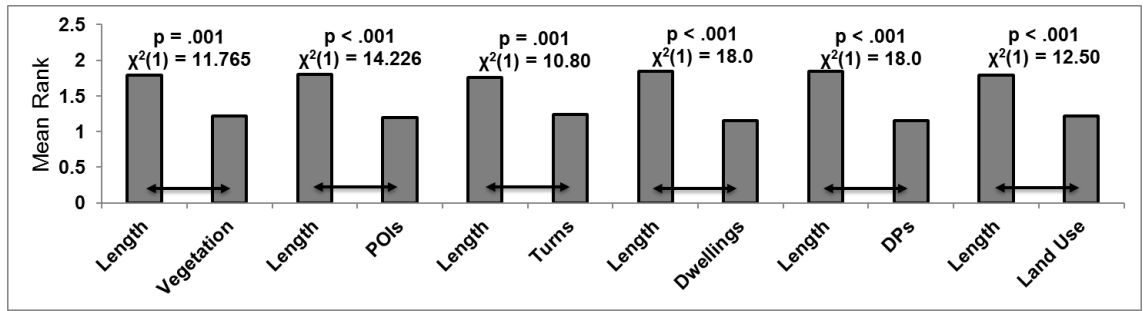
(a) Everyday Travel Results
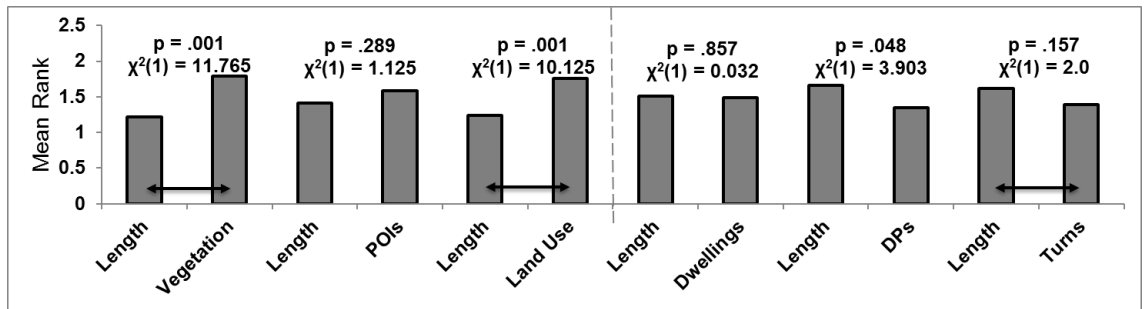
(b) Leisure Travel Results
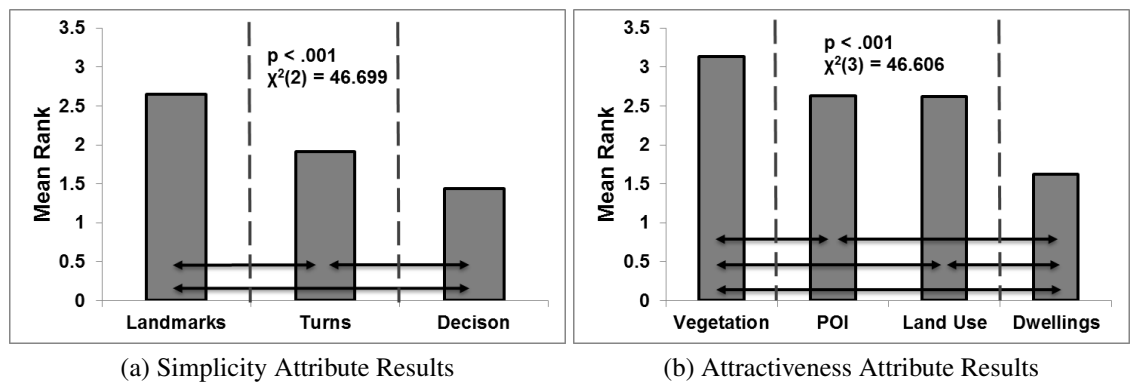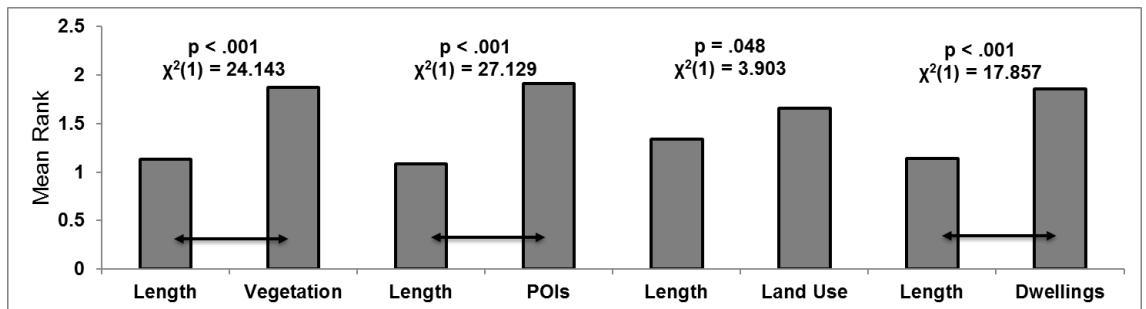

(c) Tourist Travel Results

Figure 3.9: Experiment 6 Friedman Test Results (rank and p values) - pairwise (Wilcoxon's) statistical significance is indicated by the arrows overlaid on the plot.

comparisons were then used to establish the rank of attributes, for both the attribute categories and the different journey types. For the attractiveness category and leisure and tourist journeys these ranks are new, and for everyday journeys the number of attributes known to be influential were increased.

Earlier research has suggested that testing in computer-generated environments may lead to different route choices to the real world [25]. However, the ecological validity of the present study is supported by the results of experiment 1, experiment 3 and experiment 5, which replicated the findings of real-world research for all attributes that have previously been studied with the exception of initial leg length (see Table 3.2, Table 3.3
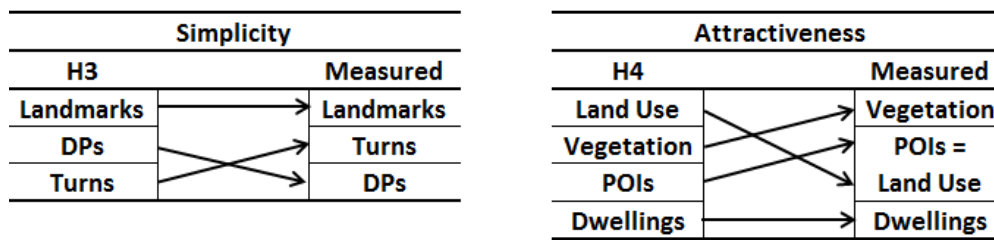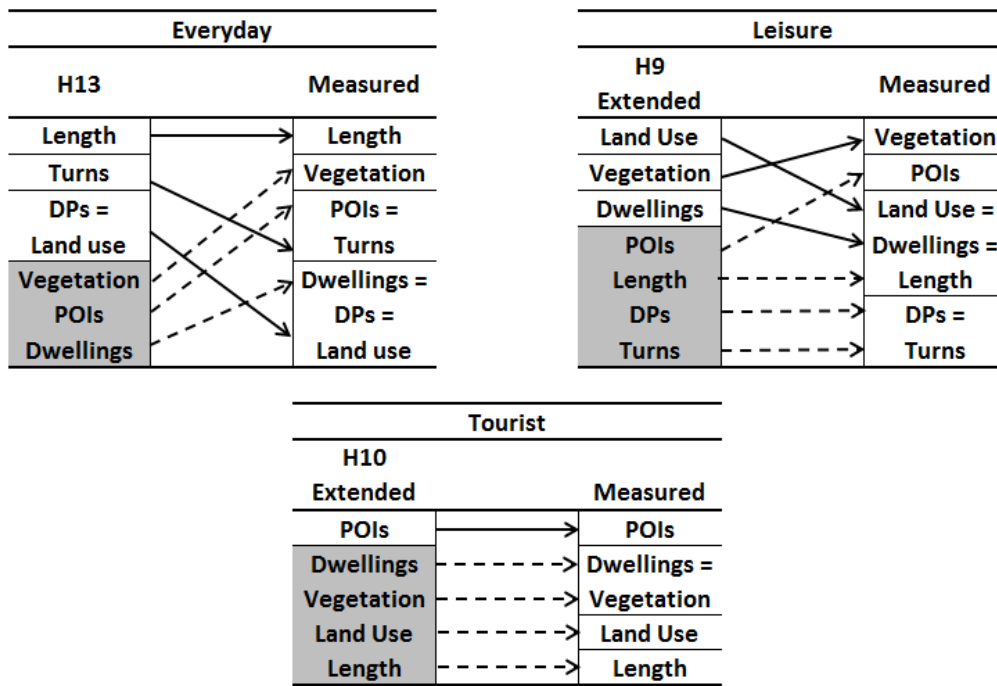
Figure 3.10: Experiment 6 Ranks. Predicted rank (hypotheses table 3.1) is compared to actual rank (right), and arrows show movement within the ranks. Grey predicted boxes and dashed lines indicate that these attributes were initially unranked. (POIs - points of interest, DPs - decision points)

and the discussion in Section 3.8.2).

Three attributes were found to affect simplicity (number of turns, number of decision points and number of landmarks) as seen in previous research and predicted by Hypothesis H1; however, initial leg length showed no significant change in preference between different levels of the attribute. In addition, the results indicate that all tested attractiveness attributes (vegetation, points of interest, architecture and land use) have an influence on preference, mirroring previous research and the prediction of Hypothesis H3. The experiments carried out also successfully produced ranks for the influence of these attributes with respect to attractiveness and simplicity, which despite varying from those predicted in Hypothesis H4, are statistically significant and therefore valid.

For both everyday and leisure journeys, turns, decision points, points of interest, vegetation, land use and dwellings all contribute to the preference of a route. In both cases, this was greater than the number of attributes predicted in both Hypothesis H5 and Hypothesis H6. This may be due to no previous comparisons being made between all of these attributes, or may indeed show trends that are only apparent with large numbers of participants. Attractiveness was shown to affect everyday route selection more than anticipated by Hypothesis H5, and leisure journeys are influenced by simplicity which had not

previously been investigated (see Hypothesis H6). The experiments carried out also successfully produced ranks for the influence of these attributes, which unexpectedly placed vegetation as the most important for both of these journey types. Differences between the placings predicted by both Hypothesis H8 and Hypothesis H9 and actual placings of all remaining attributes were also seen, with land use featuring much lower than anticipated in both ranks.

Despite a lack of previous data on how they were determined, the results of the experiments on tourist journeys confirm that points of interest, vegetation, land use and dwellings all influenced route preference as predicted by Hypothesis H7. They also indicate that simplicity attributes have no effect, as expected. Furthermore, it suggests a rank for the influence of vegetation, dwellings and land use, which had not been established by earlier research (H10).

The influences of length on simplicity and everyday journeys are commonly known and predicted by Hypothesis H11 and Hypothesis H13, and to a lesser extent on leisure routes as predicted by Hypothesis H12, but experiment 5 and experiment 6 indicate that it is also important tourist journeys. For everyday journeys shorter routes are preferred, compared to tourist routes where longer routes are favoured. For leisure journeys, although longer routes are more popular, there is a suggestion that there may be a limit on length for increased preference. Experiment 6 also allowed length to be placed in the attribute ranks for different types of journey.

The results also indicate that the scope of this research should be reduced to consider only the seven distinct attributes (length, land use, dwellings, vegetation, points of interest, decision points and turns), and then only if they have been shown to have an influence over the type of journey required. To this end, decision points and turns will be dropped as attributes that will be included in the tourist route algorithm, and the rank position of each remaining attribute will be used to determine if its inclusion is expected to be critical in individual algorithms.

Although this study is not an exhaustive examination of all of the factors contributing to route preference, it does suggest a basis for how people choose routes. Using these results, a system which selects routes appropriate for everyday, leisure and tourist journeys will be discussed in the following chapters. The relationships determined by this and previous research will now be used to form an environment model, and the ranks will now be converted into algorithms which use weighted equations to generate the cost of a partial route to select the most appropriate.

Chapter 4

Building an Environment Model

In order to successfully use the route and environmental attributes discussed in previous chapters, they must first be incorporated into a model which a computer can use to select routes. This chapter describes how an environment model based on a weighted graph was constructed, and the attributes found to be important in Chapter 3 (turns, decision point, vegetation, points of interest, land use, dwellings and length) were represented within this. Firstly three available sources of map data are discussed, and their advantages and disadvantages within the context of the present research are compared. Then a small test area of the chosen map data was downloaded, and an explanation of the the conversion process used to obtain a graph is given. Next, each of the required attributes are examined, definitions and explanations of representation techniques made as needed, and the values described were annotated onto the test graph. This is followed by details of how the resulting annotated graph model was simplified and converted to a sparse graph. Finally, the visualisation of the graph is described and the issue of displaying multiple routes is addressed.

4.1 Map Data

Digital maps are growing in use, from navigating urban streets with Google Maps[TM] [85], to planning hikes with Ordnance Survey Maps[TM] [167]. Map data is the structure underlying most digital map resources, containing details such as positions and attributes,

which can be separated completely from the method used to display them. There are several different sources of this information which could be used to develop an environment model, and this section will contrast and compare three commonly used digital map data resources - Google Maps, Ordnance Survey Maps and OpenStreetMap.

4.1.1 Google Maps[TM] [85]

Google Maps offers a free (almost) worldwide map resource, used by many researchers (see [55] for some examples). It provides a simple method for displaying maps, and annotating these with additional information, by embedding the map itself into a webpage or application. The main advantage to using Google Maps is its accuracy, with constant updates to the environment which reflect the dynamic nature of urban areas.

However, a lack of available information on the data underlying these displays makes the use of Google Maps more complicated in this research. Very little documentation detailing the format and tags of the data is accessible, with most guides covering only displaying and annotating maps rather than detailing the data already stored within them. Another disadvantage of Google Maps is the severe limitations placed on data manipulation and extension, with almost a complete ban being described within the terms of use [86]. In addition, how the information can be displayed is also closely controlled.

4.1.2 Ordnance Survey Maps[TM] [167]

Ordnance Survey provides two free geographical resources - Ordnance Survey Maps [167] and OS OpenData [168]. For the purposes of this research OS OpenData is the more appropriate resource, consisting of a database of GIS data for the whole of the UK. Unlike Google Maps, this data is provided under a Open Government License, allowing it to be adapted, published and exploited for both commercial and non-commercial purposes.

OS Open Map - Local data (from OS OpenData) provides the majority of the tags required for this research (roads, buildings, woodland, etc), which can be downloaded as a GML 3 file. Specialised tools such as Digimap [223] are available to assist with the download process, allowing easy access to small areas of data. One major disadvantage of OS OpenData is the sheer amount of data made available. The number of tag categories is huge, from coniferous trees to the gauge of railway lines. Although this could be seen as an advantage in many cases, the preprocessing required to combine tags into something usable for the present research would be considerable.

Despite this disadvantage, OS OpenData has several benefits. The data held is accurate and updated regularly, with new downloads being made available every one to two months. The level of detail is high and different 'layers' can be downloaded separately, allowing only required datasets to be pulled from the database.

4.1.3   OpenStreetMap[TM] [1]

OpenStreetMap combines the (almost) worldwide coverage of Google Maps with the database availability of OS OpenData, to provide a free product that can either be used to embed maps into websites or applications, or to download GIS data files of specific locations. The area chosen for download can be user specified, country or worldwide, and this data is provided under an Open Data Commons Open Database License (allowing copying, modifying and redistribution of the supplied data). There are many advantages to using the maps made available by the OpenStreetmap site, including the ease with which data can be extracted for specific geographical locations, and the format of the data delivered.

As with OS OpenData, this OSM data contains extraneous tags such as timestamp and user, which will need to be removed in preprocessing to give a concise format. In addition, the data provided contains many of the tags required by this research and, due to its open source nature, other tags may be added and uploaded at any time. However, the two main disadvantages of this resource are both also related to the open source ethos of its creators.

Firstly not all features have yet been fully tagged, with descriptive tags associated with vegetation and land use faring worst in this deficiency. Despite having the ability to correct and add missing tags, this may take a substantial amount of time for even a small area. Also, the data may contain inaccuracies or errors due to the way that it is collected [67]. Inaccuracy of GPS coordinates is common in urban environments, and non-experts are allowed to map substantial areas without monitoring. Careful checking of the downloaded data will be required if this data source is used.

4.1.3.1   Summary

Table 4.1 shows a summary of the characteristics of three commonly used digital map resources - Google Maps, Ordnance Survey Maps and OpenStreetMap - which are considered most important to this research. It clearly shows that Google Maps is not suitable for use in this context, but that Ordnance Survey Maps and OpenStreetMap are more appropriate bases for an environment model for route suggestion. This leaves two areas for

consideration; accuracy and preprocessing.

OpenStreetMap and Ordnance Survey Maps both produce data which requires preprocessing. Superfluous tags need removing, and in the case of Ordnance Survey Maps combining, to create only the required categories. Examination of both data sources indicates that this process is much more straightforward for OpenStreetMap data. In addition, although some of the required tags were missing from the downloaded OpenStreetMap data, the information that was present was found to be acceptably accurate. This suggests that either source could be used, with very little to choose between them. Here the decision was taken to use OpenStreetMap as the basis for the environment model in this research, mostly due to its humanly legible form.

Map Data	Open License	Offline Use	Required Tags	Accurate	Extended Preprocessing
Google Maps	✗	✗	?	✓	?
Ordnance Survey Maps	✓	✓	✓	✓	✓
OpenStreet Map	✓	✓	✓	?	✗

Table 4.1: Summary of the characteristics of three commonly used digital map resources.

4.1.4 The Test Environment

The environment chosen for investigation was a section of OSM data [1] covering the university campus and its surrounding area measuring 1.74km by 2.26km. It was selected as it has a variety of path densities, building types, land use and vegetation, as well as being convenient for testing suggested routes.

The data was downloaded and compared with Google maps and University-issued maps to check for errors. The roads, footpaths and important buildings were all present and in the correct locations.

The majority of the data within an OSM map is not needed for the environment model, so the downloaded data was preprocessed to remove relations and unnecessary tags, such as authors and change lists, but retained all essential and spatial data such as the node and way IDs and the latitude and longitude coordinates of the nodes. An example of this filtering applied to a node is shown below:

Node before filtering:

```
<node id='7270009' timestamp='2009-03-02T13:59:18Z' uid='2330'
user='LeedsTracker' visible='true' version='2'
changeset='724644' lat='53.8013029' lon='-1.5570631'>
<tag k='created_by' v='JOSM' />
```

Node after filtering:

```
<node id=7270009 lat=53.801303 lon=-1.557063 /node>
```

The pedestrian paths, including all road types, were then used to form the nodes and links (parts of ways [1] between individual nodes) of a graph. To do this, the filtered nodes were copied directly, the ways were divided into sequences of links and any buildings were tagged as geometry. An example of the conversion of ways to links is shown below:

Way:

```
<way id='-7726' action='modify' visible='true'>
    <nd ref='-2560' />
    <nd ref='-2562' />
    <nd ref='-2558' />
    <tag k='highway' v='pedestrian' />
  </way>
```

Links:

```
<link id=-2143392623 node=-2560 node=-2562 way=-7726 /link>
<link id=-2143392624 node=-2562 node=-2558 way=-7726 /link>
```

The graph produced by this process was found to be mostly connected. Any disconnected nodes were either reattached, if ways had not been correctly merged in the data, or removed if all of the links connected to the node travelled outside the bounds of the area being modelled. Figure 4.1 shows the resulting graph displayed as a map by drawing all nodes, links and geometry.

In addition to the nodes and links in the map being converted to a graph, the route and environment attributes need to be represented in the environment. Buildings and areas were retained within the data, as they may be needed to represent attributes such as dwellings, points of interest or land use. However, the data within the original map

---

[1]OSM 'ways' are data structures containing a series of nodes which represent roads, buildings, areas etc

Figure 4.1: Image showing the links, nodes and buildings in the preprocessed map. Buildings are shown in black, and links in grey.

was found to be missing many of the required tags (such as 'historic' for points of interest, 'tree' for vegetation or 'recreation_ground' for land use [2]), and needed to be annotated before being used. Although this process could have been automated, manually inputting the required data allowed more control.

By examining the results of Experiment 1 (Section 3.4) and Experiment 5 (Section 3.8), the relationship between the attribute levels and route preference can be established, and the best method for representing each attribute determined. The following sections describe how the annotation process was completed and analysis of the environment produced. Once these representations have been found and the graph annotated, appropriate routes can be selected by combining the appropriate attribute values to produce costs for the routes themselves.

4.2 Representing Attributes Within the Model

Each attribute discussed in Chapter 3 has its own characteristics and requirements for representation. This section looks at each one, defining terms and detailing how each will be annotated within the model.

---

[2] see `http://wiki.openstreetmap.org/wiki/Map_Features` for the full list of available tags and their uses

4.2.1 Decision Points And Turns

Decision points are the most objective of all the attributes to be used here. They can be easily determined from maps, as they are simply the points where three or more links meet (a junction of two or more paths). Decision points affect the ability to wayfind and perceived length [191, 210], and have a negative effect on route choice - as their number increases, preference decreases. Some previous research indicates that the number of possible paths leaving a decision point and their angle affect its complexity [57, 93, 102, for example], however this research shall concentrate on the number of decision points on a route. Decision points can be annotated directly from the OSM data, by cycling through the list of nodes created from the OSM data, flagging those which join three or more links (navnodes).

Turns are similar to decision points in that they can be determined directly from maps, but only after an angle which represents the threshold of a change in direction has been defined. They are a common wayfinding criteria [83], affect ability to wayfind and perceived route length [191, 210], and have a negative effect on route preference. One important question is - should all changes of direction be considered routes, or only those at junctions? In most previous work [83, amongst others], these two options have been equivalent due to the experimental designs used (although curves and turns are considered as different attributes in [83]), and the data provided by OpenStreetmap would allow for either to be calculated. However, turns at decision points are likely to have the most influence on route simplicity (i.e., by introducing the intersection angle described above) [57, 93, 189, 231], so only those will be considered here.

Turns can be determined automatically from the OSM data, just as decision points can in the context of the present research. Turns were defined as changes of direction over $20°$ at a decision point, where the angle is that between the incoming and outgoing links (found by calculating the dot product of the vectors representing these two links). As they are dependent on which links are used to approach and leave the decision point, and therefore dependent on the route chosen, turns are calculated at runtime rather than being added to the static map. Figure 3.2a (Section 3.4) indicates that route preference decreases linearly as decision points or turns increase. For that reason, both decision points and turns will be modelled as addition of occurrences, with each decision point or turn adding a weight of one to the cost of a route.

4.2.2 Points of Interest

Points of interest can mean many things depending on context; for example, Ordnance Survey has more than 600 categories of suitable features [169]. OpenStreetmap [166] gives examples which range from churches, schools, town halls and distinctive buildings, to postboxes, phone boxes, car parks and speed cameras. The instructions for defining a point of interest state:

> 'A point of interest or POI is a feature on a map (or in a geodataset) that occupies a particular point, as opposed to linear features like roads or areas of land use. A point of interest is not necessarily very interesting, so, for example, post boxes are relatively interesting/uninteresting, depending on context and your subjective opinion.'

This wide range of possibilities indicates that creating a rationale to follow when mapping an area is difficult. One approach to this may be to crowd source opinions, or mine social media for commonly visited locations. An alternative, and the one taken by this research, is to gather points of interest from tourist information maps and other similar sources. Although this approach restricts the points to those considered to be important to the institution or body which produced these sources, they are a typical way from which tourists gather relevant information.

For the purposes of this research, a list of 20 points of interest (POIs), or 'sights', were provided by the University of Leeds Communications and Press Office in the form of a map which is regularly issued to visitors and prospective students. These 'sights' are marked in red on the map shown in Figure 4.2a. Points of interest are somewhat more problematic to represent than decision points or turns, as they are associated with the environment surrounding a route rather than the route itself. Points of interest may occur anywhere along a route, not only at junctions, and therefore could be associated with both the nodes and links within the representation, but by annotating them onto the links an association with the corresponding nodes is implied. The links associated with the points of interest for the chosen environment are shown in blue in Figure 4.2b.

Figure 3.2b (Section 3.4) indicates that route preference increases linearly as the number of points of interest increases. As with decision points and turns, points of interest will be represented by the addition of occurrences, but this time a higher number of POIs should produce a lower route cost. As the number of points of interest remains static within a chosen environment, it is possible to determine the maximum number of points of interest which can occur on a route. Using this maximum, the cost can be calculated
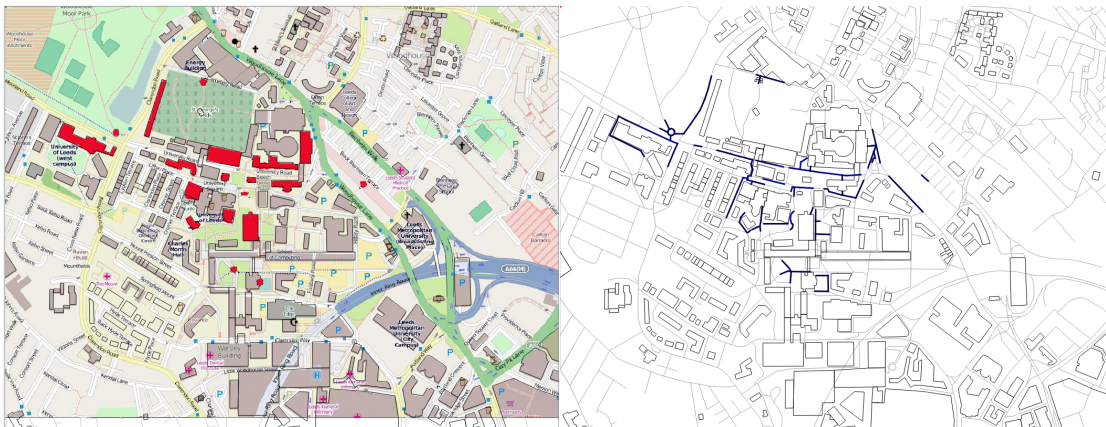
(a) Map showing the points of interest (red).

(b) Screenshot showing the links affected by points of interest on the map.
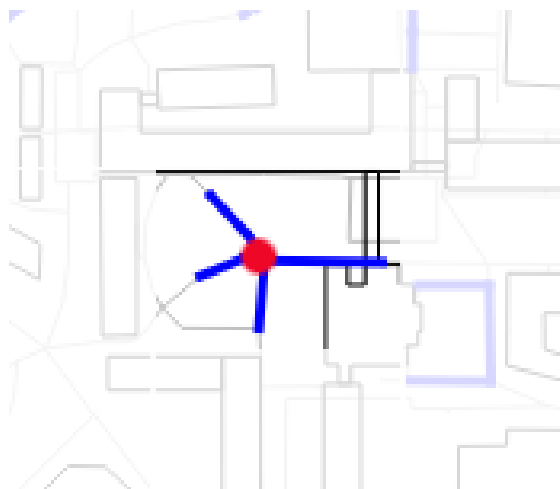
(c) Screenshot showing the links (blue) affected by a single point of interest (red).

Figure 4.2: Images showing how the the points of interest levels were assigned to the final annotated map.

by comparing it against the number of points of interest on the current route. Unfortunately, selecting a route containing the maximum amount of an attribute is not as simple as selecting a route containing its minimum.

The main issue with representing points of interest centres around the fact that a single point of interest may be associated with several links, and one link may be associated with several points of interest. An example of why this is problematic is shown in Figure 4.2c, where four links (shown in blue) are associated with the same point of interest (shown in red) which is located in the centre of the junction between them. If each of these links is annotated with a value representing the number of POIs affecting it, then the POI value of a route using all four of these links will be greater than a route using only two, despite

both routes passing the same number of points of interest. The effect of this method of representation was unclear, so three annotation methods were tested:

Method 1 - Annotate each link with the number of points of interest which affect it. For the example in Figure 4.2c, each of the blue links would have a POI value of 1. This method will encourages repetitions of a single point of interest within a route.

Method 2 - Annotate each link with a proportion of the effect of a point interest according to how many other links are affected by it. For the example in Figure 4.2c, each of the blue links would have a POI value of 0.25. This method spreads the effect of the point of interest over all of the links surrounding it.

Method 3 - Give each point of interest an ID and annotate links affected by this point of interest with its ID. For the example in Figure 4.2c, each of the blue links would have a list of points of interest of length 1 containing the ID of the point of interest at the junction. Counting the number of unique IDs to produce the POI value prevents each point of interest being counted more than once.

For each of the methods described above, an algorithm was created which aimed to produce routes containing the maximum number of points of interest without visiting a link more than once. An example of the results of each of these algorithms are shown in Figure 4.3, indicating that the method of annotation can have a substantial effect on the routes produced.

Both Method 1 (Figure 4.3a) and Method 2 (Figure 4.3b) encourage the algorithm to select all links affected by a point of interest. This leads to several loops in the route, to enable multiple passes of the point of interest. Although this increases the point of interest count, it does not increase the number of points of interest present on the route and may increase the length of the route considerably. In addition, Method 2 makes annotation difficult, as the number of links affected by a point of interest must be counted so that the attribute value can be spread between them.

Method 3 (Figure 4.3c) has the disadvantage that IDs need to be assigned to the points of interest, but it produces routes with fewer loops. Using this approach, the point of interest value reflects the number of unique points of interest on the route, and annotation of links is straightforward. These results indicate that this is the best method for annotating points of interest to the final map and therefore, it was used.

Figure 4.4 shows the distribution of points of interest over the links and the environment. There is a maximum of three points of interest present on a single link, and the majority of the links have none (Figure 4.4b). Figure 4.4c indicates that points of interest
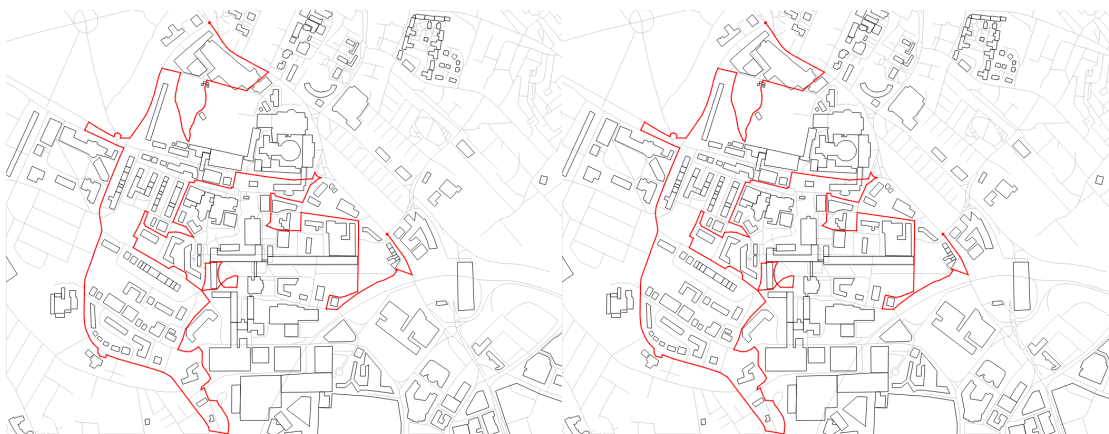
(a) Screenshot showing the route produced using an-notation Method 1.

(b) Screenshot showing the route produced using an-notation Method 2.

(c) Screenshot showing the route produced using an-notation Method 3.

Figure 4.3: Screenshots showing examples of the routes produced by the three different POI annotation methods.

occur mostly in the centre of the map, meaning that in order to contain many points of interest, routes must travel through this area.

4.2.3   Length

As with turns and decision points the length of a link can be found directly from the OSM data, in this case by measuring by the distance between the nodes at each of its ends (given in latitude and longitude), and converting this into kilometers. This is done by using a standard Mercator projection to map the nodes from latitude and longitude to xy coordinates, and then calculating the magnitude of the vector between them. Fortunately, OSM data places a node at each change of direction within a way, so this distance mea-

(a) Screenshot showing the annotated points of interest levels for the final map.
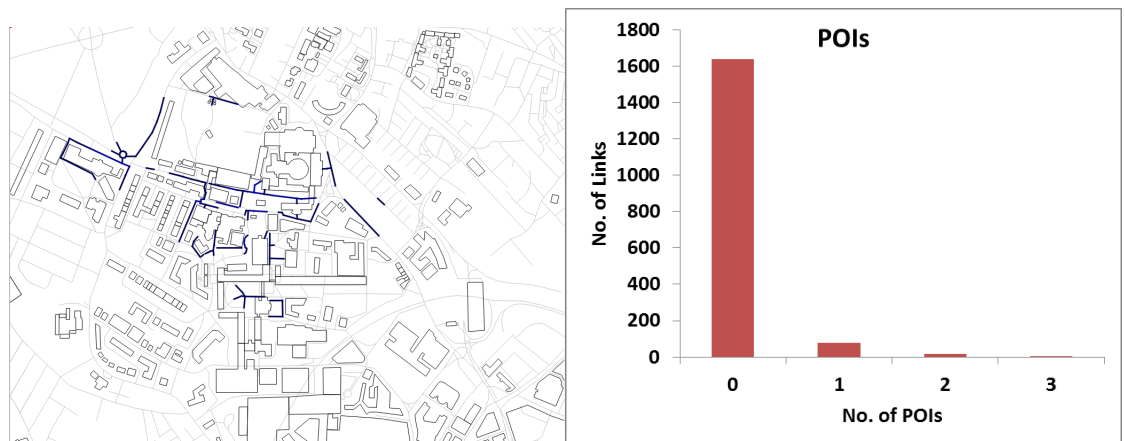
(b) Distribution of the POIs over the links within the representation

(c) Heatmap showing the distribution of POIs over the representation (darker colour represents higher mean POIs per link)

Figure 4.4: Images showing the distribution of POIs

surement can be considered to be sufficiently accurate for the purposes of the algorithm described later.

Figure 3.8 (Section 3.8) suggests that the relationship between length and route preference is dependent on the journey type required. For everyday journeys the relationship is negative, so shorter routes are preferred, but for leisure and tourist journeys the relationship is positive, indicating that longer routes are preferred. This difference in preference affects how the cost of a link will relate to its length, but not how length should be represented. The most appropriate method of representation is to annotate the link with its length in kilometers, which corresponds to the cost for this attribute.

4.2.4   Vegetation and Land Use

In order to define how vegetation should be represented in an environment model, first two important questions must be answered:

1. What constitutes vegetation?

   Any area of planting could be considered to constitute vegetation, including flowerbeds, forests or even small planters. To make this attribute useful however, a minimum size of planted area should be given. There is low preference for weedy, mostly brown, fields and scrubland [116], so small single plants will be discounted, although a single trees will be included as they have been shown to be important in scene preference [203].

2. What types of plants are considered vegetation?

   All types of plants fall under the general classification of vegetation, but low ground-cover types (such as grass) will be discounted from this list and considered instead as part of land use. This omission is justified as they are fairly uniform and smooth, reducing the preference for areas containing only groundcover vegetation [116]. All remaining types of planting will be included in this category, and vegetation types will not be differentiated - although studies have shown that the presence of trees is important, varying results have been found for other vegetation types and density [203].

   Vegetation will be represented in this environment as a continuous attribute along the links of the graph, with both sides of the link taken into consideration.

   Although vegetation is clearly a continuous attribute, land use can be classified as either a continuous or type attribute. Land use is an attribute that is commonly used in the fields of geography, planning and climate studies, although several different methods and classification systems exist for defining it. One commonly used classification system [159] consists of six main categories; (1) cropland, (2) grassland, pasture and range, (3) forest, (4) urban land, (5) special uses (including rural transportation and wildlife areas) and (6) miscellaneous land. These can be generalised to be parkland, farmland and urban land types. Although land use is a type attribute (one in which its type is important rather than just its presence), it can be simplified to be a continuous attribute if only the presence of the most preferred (parkland) type is considered. As the area selected for this research (and most urban settings) contains no farmland, this simplification will be used by the environment modelling approach taken here, with only parkland (the most preferred level) considered for this model. With this in mind, the presence of parkland will be annotated onto the links for this environment to represent the land use attribute.

Figure 3.2b (Section 3.4) indicates that both vegetation and land use are likely to have a positive quadratic relationship. This suggests that vegetation and for this model parkland can be represented as the proportion of the route which they cover. The vegetation value was annotated as the length of either side of the link which it covers, with a similar approach taken for parkland. This gives a maximum value which is twice the length of the link. The cost of a link with high values of either of these attributes needs to be low, so the cost will be given by the proportion of the route which is not covered by the attribute, calculated at runtime as:

$$Vegetation = \left(1 - \frac{Length_{VEG}}{2 * Length_{LINK}} \right)^2$$

and

$$LandUse = \left(1 - \frac{Length_{LAND}}{2 * Length_{LINK}} \right)^2$$

For these attributes, manual annotation was required. To make this process generalisable and scalable, a satellite image taken from Google Earth [85] (Figure 4.5a) was used for the majority of the attribute detection. Using this image, areas containing visible vegetation (such as the trees on St. George's field) were identified. In most cases, these areas are obvious, so this process is straightforward. For parkland, any areas marked as park on the map were identified first, and any areas with visible grass (such as the green areas in the top righthand corner) were added to this category. Where there were difficulties determining whether an attribute was present, or whether or not an attribute covered a link, Google StreetViews [85] was used to show what was visible from the link involved.

To enable annotation, a tool was developed that allowed the image to be placed behind the map; single links were then selected and the appropriate values added graphically. Figure 4.5b shows the data entry section of the tool, and values which have been added for a single link. By using this tool, the vegetation and land use attributes were added to the whole map in just two days. Figure 4.6 shows the distribution of these attribute values across the final map.

Figure 4.6 shows the links affected by vegetation and land use, and further analysis of their distribution is given by Figure 4.7 and Figure 4.8. The histograms in Figure 4.7 show that most links have zero vegetation and parkland. In addition, the heatmaps in Figure 4.8 show that vegetation covers more of the environment than land use, but that both are mostly restricted to the edges of the environment, with only low levels in the centre and little appearing in the lower left quadrant.

(a) Screenshot showing the map overlayed on a satellite image to allow attribute annotation.

(b) Screenshot showing the annotated values for a single link.

Figure 4.5: Screenshots showing the satellite image and data entry panel of the tool.

4.2.5 Dwellings

Many types of buildings may influence route preference, but here we will focus on dwellings. Dwellings are current or former homes (which may have been converted to other uses, but

(a) Screenshot showing the links annotated with veg-(b) Screenshot showing the links annotated with
etation for the final map. parkland for the final map.

Figure 4.6: Screenshots showing the vegetation and land use for the final annotated map.
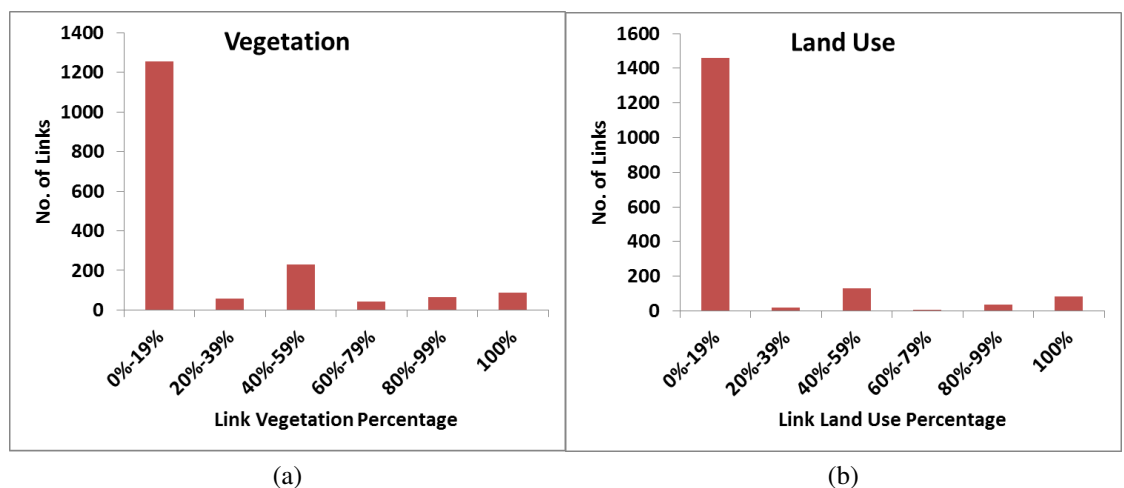
(a)   (b)

Figure 4.7: Distributions of (a) vegetation and (b) parkland over the links of the environment.

retain their external characteristics), and fall into three main categories - multiple occupancy (least preferred), single occupancy and large or historic (most preferred) dwellings [209]. They form type attributes which can be annotated on the links of a graph, although their presence is far more subjective than some of the other attributes discussed. Some buildings are obviously designed to be single occupancy, such as terraced housing, but differentiating between dwellings which were designed to be large single occupancy and those that are multiple occupancy is more tricky. Here multiple occupancy dwellings were defined as those that were specifically built for this purpose (such as blocks of flats), with any which were converted at a later date discounted. Where this distinction was
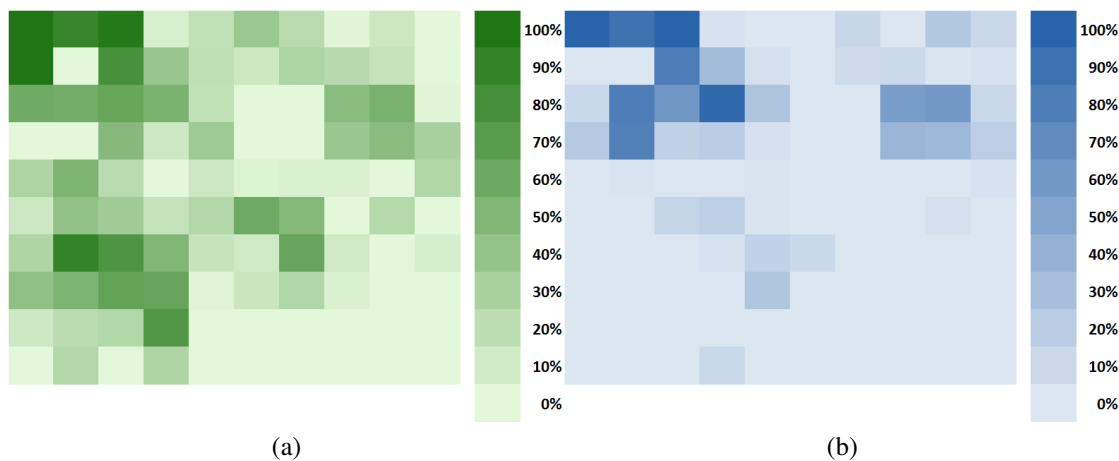
Figure 4.8: Heatmaps showing the distribution of (a) vegetation and (b) parkland over the environment representation (darker colour represents higher mean vegetation proportion).

ambiguous, the final decision was taken by the annotator.

Figure 3.2b (Section 3.4) indicates that dwellings have a positive quadratic relationship with route preference. Dwellings can therefore be added to the graph as the proportion of the link that they cover, but as all levels of this attribute are statistically significant, then both small single occupancy housing and large homes will need to be represented. The most straightforward approach to this is to apply different weights to semi-detached or terraced housing compared to large detached homes. As with the continuous attributes, the cost of dwellings should decrease as its value increases. This suggests that the cost of this attribute will be connected to the proportion of the route not covered by either type of housing, rather than the proportion that is. Dwellings will also be annotated onto links of the graph as the length of either side of the link bordered by dwellings, with the maximum being twice the length of the link.

Identifying dwellings from a satellite image is difficult, although algorithms exist to do this automatically with some success [111, 149, and many more]. To make the identification process easier Google Street Views was used, with each link being walked and the dwellings on both sides recorded. As the dwellings attribute is dependent on the types of dwellings rather than just their presence, terraced and semi-detached housing was given half the value of detached housing. Any buildings which were originally designed to be dwellings but which had been converted to another use (e.g. houses converted to offices) were also included within dwellings, as were all aspects of the buildings (front, back and sides).

Figure 4.9a indicates the links which are annotated with non-zero dwellings values. Like the other continuous attributes, the histogram in Figure 4.9b shows non-linear

and non-normal distributions for the dwellings attribute. However, no links have 100% dwellings coverage. This is due to the the way dwellings are represented on the graph, with a lower value being used for semi-detatched and terraced housing. The heatmap shown in Figure 4.9c indicates that dwellings are again found around the edges of the map, with few appearing in the lower left quadrant.
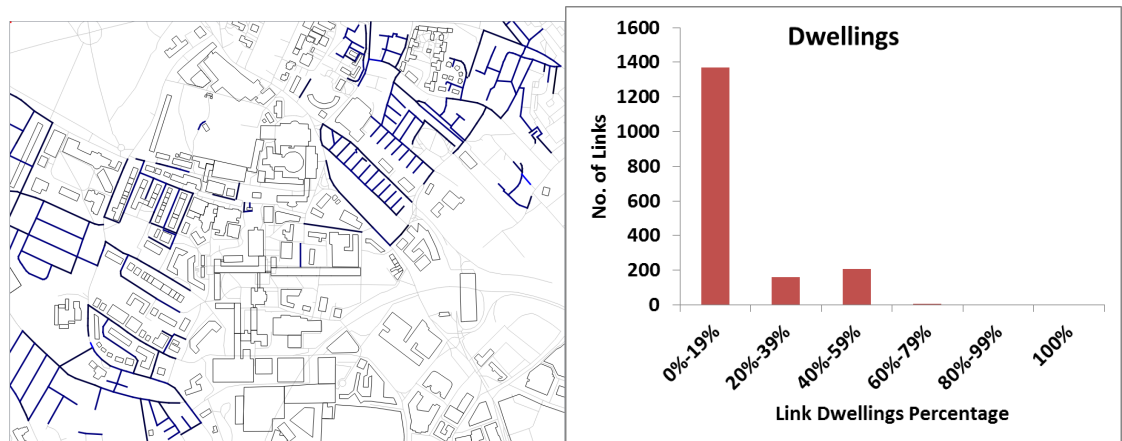
(a) Screenshot showing the links annotated with dwellings (both small and large single occupancy homes) for the final map.

(b) Distribution of dwellings (both small and large single occupancy homes) over the links within the representation
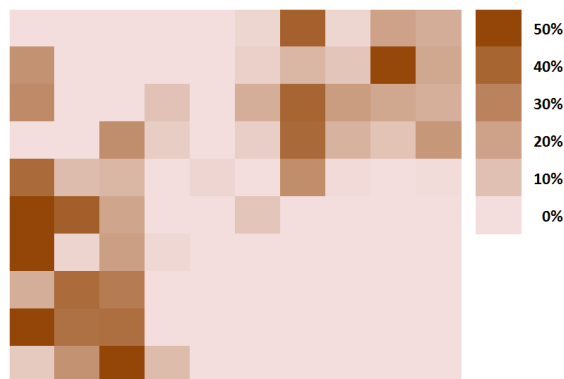
(c) Heatmap showing the distribution of dwellings (both small and large single occupancy homes) over the representation (darker colour represents higher mean dwellings proportion

Figure 4.9: The distribution of Dwellings.

4.3 Converting the Map into a Sparse Graph

With the annotations complete, the final map can be converted into a sparse graph. This is a simple process as the map itself is constructed of nodes and links, as discussed earlier.

The placement of nodes at each object or change of direction along the way means that there are many unnecessary nodes and links within this representation. Simplifying the graph underlying the map reduces the search space and therefore the time taken to find suitable routes.

The simplification process first marks nodes with a single link as navnodes, as long as they occur on pedestrian paths. These nodes represent cul-de-sacs in the environment, and are therefore part of the navigable representation. Secondly, any nodes connecting two or more links are added to this list of navnodes. Series of links between navnodes are then combined to form a single navlink, and any remaining links which are not attached to navnodes (such as those connecting walkways to buildings) are removed. The length of a navlink is found by summing the length of the links within it, and the same process is used to find vegetation, land use and dwellings values. For points of interest, a list of all the unique points of interest IDs found along the navlink is collated. Table 4.2 indicates how the graph simplification reduces the numbers of nodes and links within the graph, with 77% of nodes removed and 34% of links.

Object	Before Simplification	After Simplification
Nodes	4151	937
Links	2615	1735

Table 4.2: Object numbers before and after graph simplification.

The result of this process is a sparse graph, annotated with all of the required attributes. This representation allows the creation of a route suggestion algorithm by forming attribute cost functions using the annotated values described above, and combining them to develop route costs for different attribute categories and journey types.

4.4 Visually Representing Graphs

In order to communicate route information to a user, a visual representation of the model and suggested routes must be created. The format of this display will initially be only for visual inspection although it would also be appropriate for use on mobile environments, but conversion to something appropriate to a more diverse range of users is outside of the scope of this project.

As discussed previously (see Section 3.2), any two or three dimensional visual representation of an environment or route may be termed a map. However, as the display method becomes more complex, the data required to produce the display becomes larger [73]. As the data held within the chosen GIS database (and the environment model)

is restricted to be 2-dimensional, then a 2-dimensional display will be most appropriate in this case. Displaying all of the data held within a small section of a GIS database (in this case OSM map) produces Figure 4.10a; however, this leads to a complex map and the requirement for large amounts of data to be stored within the model. This is contrary to the aims of this thesis, so how can the visual representation be simplified in order to reduce the data needed?

Schematic maps offer a form of simplification in which all but the essential components of the map are removed [73, 122]. Schematics can be very simple visual representations such as in tube maps, where knowing only which stations are connected by which lines are necessary for successfully planning and navigating a route [122, 220]. However, any simplification of a map may be considered a schematisation. Although this type of representation could be used to reduce the amount of data required to display the environment model for this research, an important question must first be answered - what elements of the map are 'essential'? [73]

There are no defined guidelines for schematising a map and, in fact, several papers have suggested that the type of simplification is dependent on the task for which the representation is to be used [73, 122, for example]. One answer to the question of which components are essential may be that only the links and nodes already held within the model need to be retained, as they are the only components required to directly display routes. Although Figure 4.10b indicates that this representation may be sufficient to allow for route comparison, this depiction of the model is not necessarily easily recognisable as the selected area of the map. In contrast Figure 4.10b, showing an alternative solution, indicates that simply replacing the roads and walkways on the map does allow for ease of recognition, but does not remove the majority of the data held within the model. For this research, a compromise between these two representations is desirable.

In both sketch maps and the generalised wayfinding task, buildings are used as landmarks [73, 142, 199, 212]. This is mostly due to their static nature, size and tendency to be identifiable [142]. Retaining building information in a map also retains the spatial relationships between the physical environment and 2-dimensional representation of it, reducing the cognitive load of the user [73], and making it more recognisable. Figure 4.10c shows a map schematisation which removes all map data other than that related to buildings, nodes and paths. This representation requires relatively little additional information to be stored within the environment model, but makes the area more recognisable to the corresponding map. In order to include this additional data, a second layer of representation will also be created (containing the geometry of the buildings) but this will be used for display purposes only.
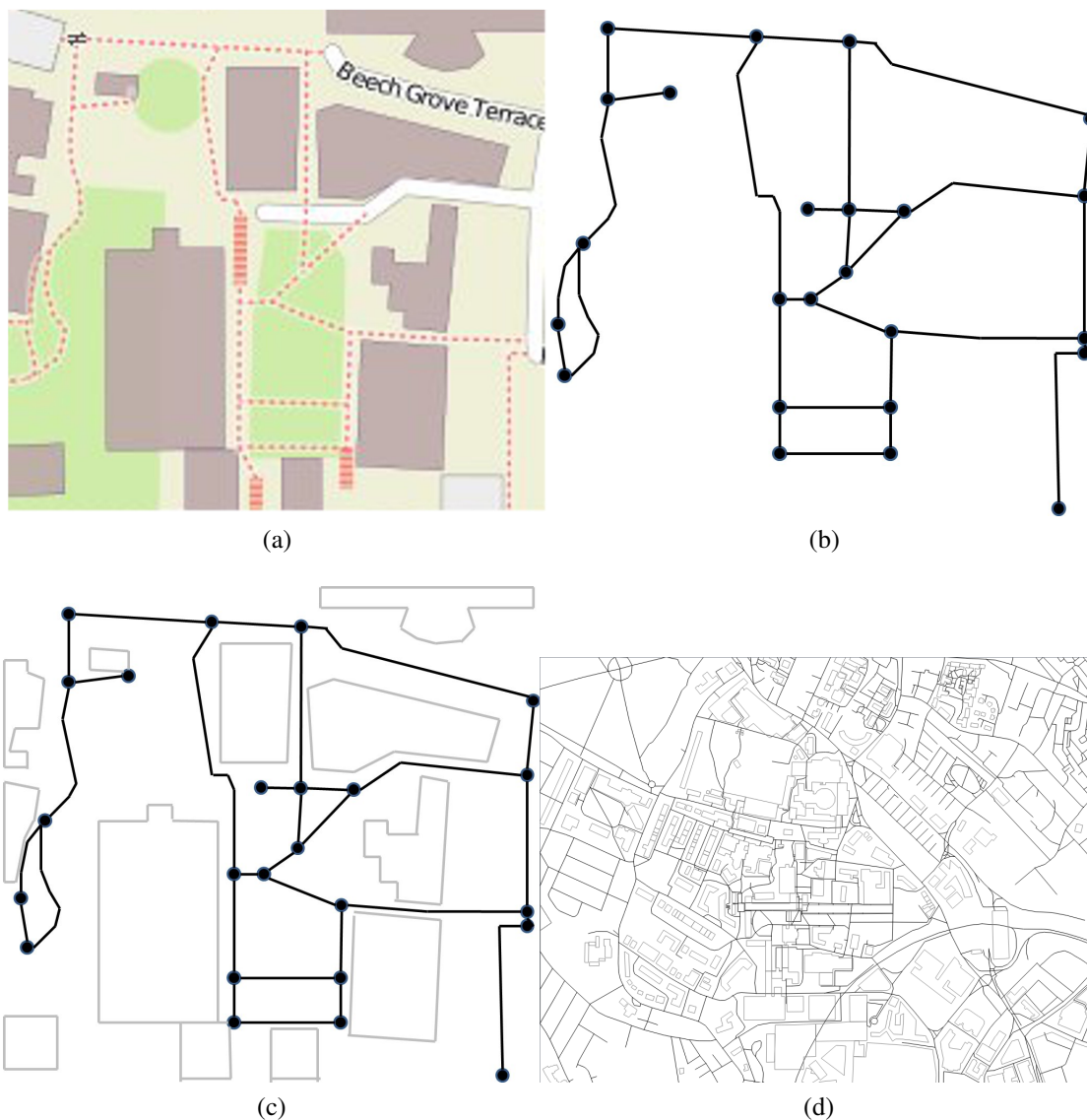
Figure 4.10: Three visual representations of a small section of the test area, (a) the OSM map, (b) roads and walkways replaced with nodes and paths, and (c) all but nodes paths and buildings removed. (d) shows the entire test area in schematic form.

4.5 Displaying Multiple Routes

Displaying multiple routes on a single map is a common problem, especially in fields such as the production of route maps for public transport systems. The issue arises from the contradictory objectives of increasing the number of routes displayed to reduce the number of maps required, displaying the routes so that they are all visible, and not allowing these routes to obscure large portions of the map.

Looking to public transport maps for a solution gives two options shown in Figure

4.11a and Figure 4.11b, which use an approach similar to the London Underground Map [214] in that each route is displayed in a different colour. The representation shown in Figure 4.11a is the simplest approach to draw using a computer, with the width of each line fixed and determined by the order in which the routes are generated. This offers the advantage that a route can be traced by the width of its line as well as its colour, but it could be argued that this approach obscures more of the map than is necessary by using wider lines.

(a)   (b)

(c)
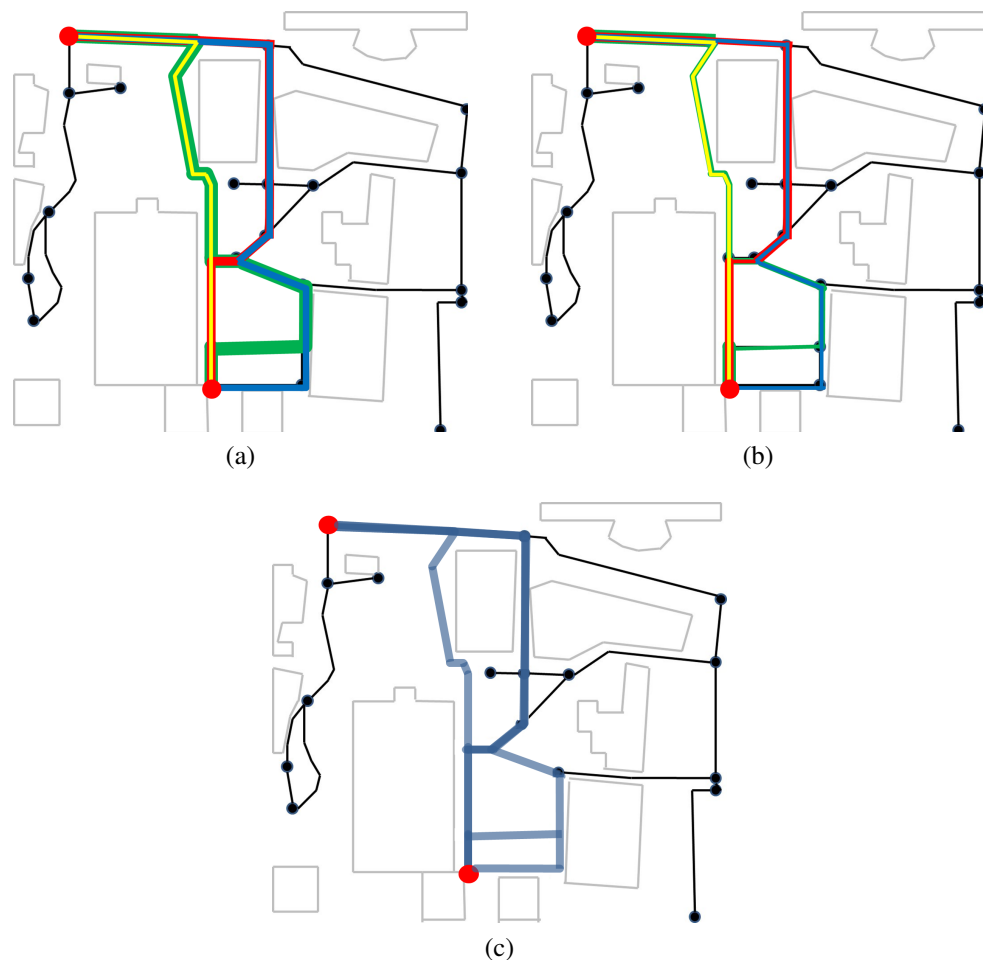
Figure 4.11: Three visual representations of a small section of the test area with multiple routes shown. Shows (a) routes in different colours and widths, (b) routes in different colours and widths varying according to the number of times the section is used, and (c) routes with a fixed level of opacity overlaid to indicate the number of times the section is used.

Figure 4.11b shows an approach that addresses this issue, by varying line widths only where routes overlap. Although this removes the opportunity to follow a route by the width of its line, less of the map is obscured and the width of a route section can be

used to visually determine where route sections are used by multiple routes. However, the approaches shown in Figure 4.11a and Figure 4.11b both suffer from one major restriction - they can only be used successfully to display a small number of routes. With this in mind, the approach used in Figure 4.11b will be used to display and compare only a sample of the routes produced by an algorithm for a wayfinding strategy or attribute.

An alternative approach to representing multiple routes, is to use a visualisation technique which communicates only the popularity of a route or route section [157, 232]. Figure 4.11c shows an example of this approach, where routes with a fixed level of opacity are laid one on top of the other, allowing more popular route segments to be identified by how boldly they stand out from the background. Although the majority information about individual routes is lost, this does allow a large number of routes to be displayed on a single map. It forms a type of geographic histogram which is useful to determine route trends within (possibly) large participant groups.

With the environment model constructed and visualisation techniques defined, this research can now move on to the route selection process itself.

Chapter 5

Constructing a Simple Pedestrian Route Selection Algorithm

The aim of this research is to construct a simple algorithm, which can suggest appropriate routes for a number of different wayfinding needs. The previous chapters discussed which attributes are required for route selection as shown in Table 5.1, and how they were represented in an environment model. This chapter will explain how these attribute values were converted into cost functions and used to suggest routes which maximise or minimise a single given attribute.

Simple	Attractive		Everyday	Leisure	Tourist
Length[1]	Vegetation[2]		Length[1]	Vegetation[2]	POIs[2]
Turns[1]	POIs[2] =		Vegetation[2]	POIs[2]	Dwellings[2] =
DPs[1]	Land Use[2]		POIs[2] =	Land Use[2] =	Vegetation[2]
	Dwellings[2]		Turns[1]	Dwellings[2] =	Land Use[2]
			Dwellings[2] =	Length[2]	Length[2]
			DPs[1] =	DPs[1] =	
			Land use[2]	Turns[1]	

Table 5.1: Importance rankings for route attributes showing simplicity, attractiveness and each of the different journey types. [1] Indicates a negative relationship, [2] indicates a positive relationship. (POIs - points of interest, DPs - decision points)

This chapter starts out by briefly discussing the tool developed to enable route selection and evaluation. Secondly it describes the metrics that were used to determine a

'good' route. Finally, it individually examines the characteristics of each of the attributes in Table 5.1, describes different algorithms which could be applied to find routes which maximise or minimise these and which of the discussed algorithms which can be used to best incorporate the attributes into route selection.

5.1 The Tool

A tool was initially created to graphically display the routes selected on a map. An .osm file is initially loaded and converted into the environment representation as described in Chapter 4, the map is then displayed to the user and options offered for different types of route selection. Two of the most important user interface tabs are shown in Figure 5.1.
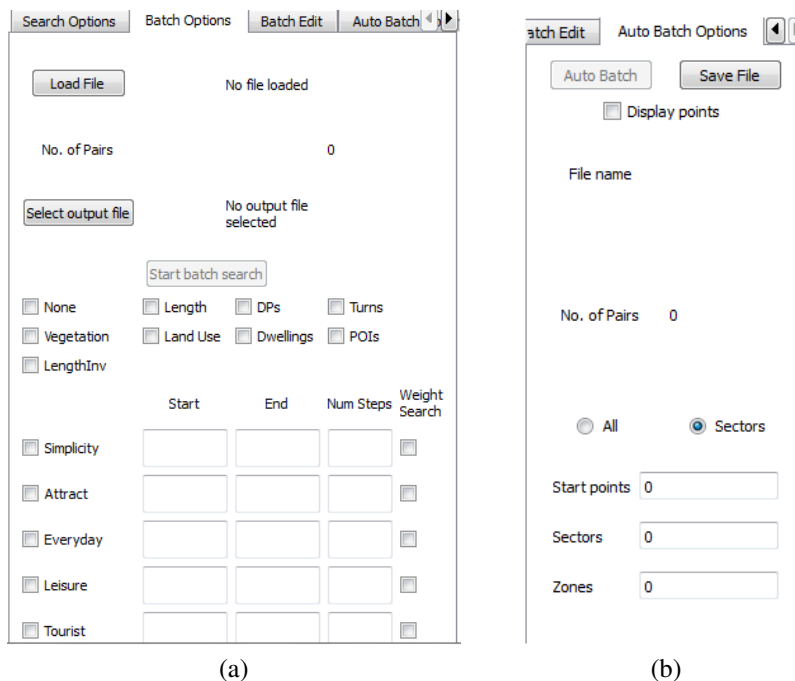
Figure 5.1: Screenshots showing the (a) batch search options tab and (b) autobatch options tab available in the tool.

Figure 5.1a shows the options available on the batch search options screen. Given a number of start-end pairs, each of the single and multiple attribute algorithms can be selected individually, or in combination. Where a multiple attribute algorithm is chosen, a weight range and number of steps can be entered. This allows for the weights of each attribute to be varied by the algorithm in order to find the optimum route and combination of weights.

(a)

(b)
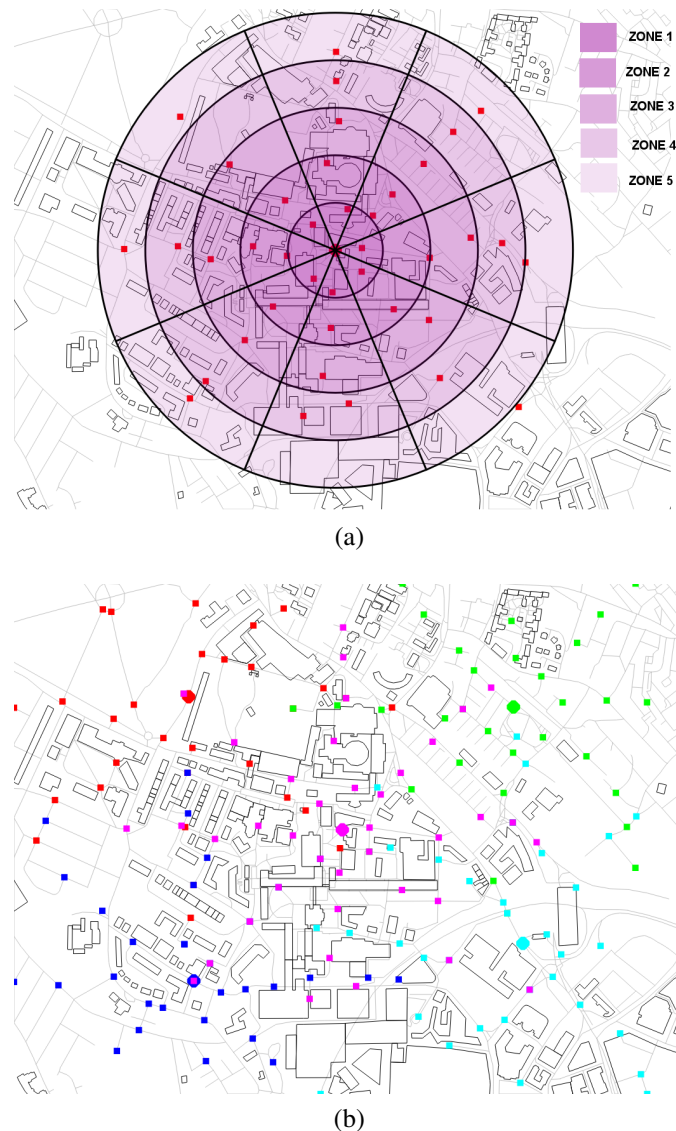
Figure 5.2: Screenshots showing (a) the five zones and eight sectors for a single start point and the generated end points, and (b) the start-end point pairs used for testing. In (b) the five larger circles are the start points, the smaller squares are end points and the colour indicates which end points belong to each start point.

The tool also includes functionality for automatically creating a batch of start-end points (Figure 5.1b). Automatic generation takes three user entered values; number of start points, number of sectors and number of zones. The algorithm selects start nodes spaced evenly across the map, to give the best coverage. For each start node, the map is partitioned into zones (concentric circles around the start point) which are subdivided into areas using the sectors angle (360 °/sectors) as shown in 5.2a. The node closest to the centre of each area is then selected as an end point.

To provide a set of test start-end pairs, the automatic batch algorithm was run with

inputs of five start points, eight sectors and five zones, resulting in 174 point pairs (the remaining 26 were outside of the boundaries of the map used) as shown in 5.2b. These 174 point pairs were used to test the performance of each of the algorithms discussed throughout this Chapter.

5.2   Metrics

Metrics which define how well a route (and therefore set of weights) performs need to be decided, and a sensible first step in determining how to score routes is to consider the tested attributes themselves. For example when examining the simplicity of a route, we should be applying a penalty to excessively long routes or those with many turns or decision points. One possibility is to use the actual length or number of turns or decision points to assess routes directly, but this may not be the best approach. What would happen, for example, if the shortest / only route between two points is relatively long compared to the straight-line distance between them? If the length is used directly, then in this case the route would be unnecessarily penalised. An alternative approach which avoids this issue is to compare the current route attribute cost to that of the minimum route attribute cost (i.e., comparing the current route length with that of the minimum length route) to give the percentage increase ($(att_{current}/att_{min}) - 1$).

In real life, people are unlikely to choose routes which vary substantially between geographically similar points [128], suggesting that some measure of route similarity also should be used to evaluate algorithm suggested routes. There are many ways of measuring route similarity [91], including counting errors at turning points when considering at the ability to follow paths [65]. However, most of these are inappropriate for the present research as they relate to either participant behaviour, or attributes which are being manipulated (such as route length). We need a metric which is independent of the attributes being used to generate route cost, such as route overlap. Route overlap is the total length of the links which occur in both routes, and is regularly used to apply an additional cost to similar routes in systems which seek alternative routes for motor vehicles [32, 181].

5.3   Single Attribute Cost Functions

Before progressing to more complex route types, the behaviour of single attribute cost algorithms was examined. This was partially to confirm the correctness of the algorithm, but also to explore the routes that they produce. The single attribute cost functions can be broken into two groups; those requiring a minimum value (negative relationship in Table

5.1), and those requiring a maximum value (positive relationship in Table 5.1). As they are more straightforward, the attributes requiring a minimum value cost function (length, decision points and turns) were examined first. These are the simplicity attributes from Chapter 4 and have the advantage of being objective, meaning that issues associated with subjectivity were avoided.

5.3.1   Minimum Value Cost Functions

Section 2.5 has established that Dijkstra's least cost search algorithm [54] will form the basis for the algorithms in the present research, due to its simplicity, flexibility and accuracy. If distance is replaced with cost in this algorithm, then it transforms it into a minimum cost algorithm which can be used for any 'cost' values (with the assumption that they are positive). To find a route with the minimum amount of an attribute (i.e., the shortest length), Dijkstra's algorithm requires that the cost of an attribute increases with its amount. For length which is a continuous attribute, the most logical approach to meet this requirement is for the cost of a route (C_{LEN}) to equal the sum of the lengths of the links within that route, as given by Equation 5.1.

$$C_{LEN} = \sum Length_{Link} \tag{5.1}$$

This is the traditional cost function for Dijkstra's algorithm, but other minimum value cost functions are also as straightforward. Modifying the original algorithm to calculate the cost of each partial route, using a weighted function containing the appropriate route attribute, produces the algorithm shown in Algorithm 1.

```
function Dijkstra(pStartNode, pEndNode):
sRoute* pBestRoute=NULL;
m_fMinCost=m_fMinLength=DBL_MAX;
for(int i=0; i<pStartNode->pLinks.size(); i++)
  sPartRoute* pNewRoute=
    new sPartRoute(pStartNode,pStartNode->pLinks[i]);
  CalculateCost(pNewRoute);
  AddToExplore(pNewRoute);
while (m_vToExplore.size()>0)
  sPartRoute* pCurrentRoute=m_vToExplore.front();
  if(pCurrentRoute->fCost<=m_fMinCost)
    cLink* pLink=(cLink*)pRoute->vPrevParts.back();
    if(pLink->m_eSearch =eSearchState_Explored)
      pLink->m_eSearch=eSearchState_Explored;
      cNode* pLastNode=(cNode*)pRoute->LastNode();
      cNode* pNode=pLink->GetOtherEnd(pLastNode);
      if(pNode!=pEndNode)
        for(unsigned int i=0; i <pNode->pLinks.size(); i++)
          cLink* pLink2=pNode->vpLinks[i];
          if(pLink2 && !pRoute->RouteContains(pLink2) &&
              pLink2->m_eSearch!=eSearchState_Explored)
            sPartRoute* newPart=new sPartRoute(pRoute,pNode,pLink2);
            CalculateCost(newPart);
            AddToExplore(newPart);
      else
          if(pCurrentRoute->fCost<m_fMinCost ||
           (pCurrentRoute->fCost==m_fMinCost &&
          pCurrentRoute->fLength<fMinLength))
            m_fMinCost=pCurrentRoute->fCost;
            fMinLength=pCurrentRoute->fLength;
            pBestRoute=ConstructRoute(pCurrentRoute);
    delete(pCurrentRoute);
  else
    delete(pCurrentRoute);
return pBestRoute;
```

Algorithm 1

For the point attributes decision points and turns, the most straightforward approach is to add 1 to the route cost each time one of these features is encountered. This is equivalent to the decision point cost (C_{DP}) being equal to the number of decision points found on that route, as indicated by Equation 5.2, and the same is true for turns (C_{TURN}) as shown in Equation 5.3.

$$C_{DP} = \sum DP_{Route} \tag{5.2}$$

$$C_{TURN} = \sum Turns_{Route} \tag{5.3}$$

(a)

(b)

(c)
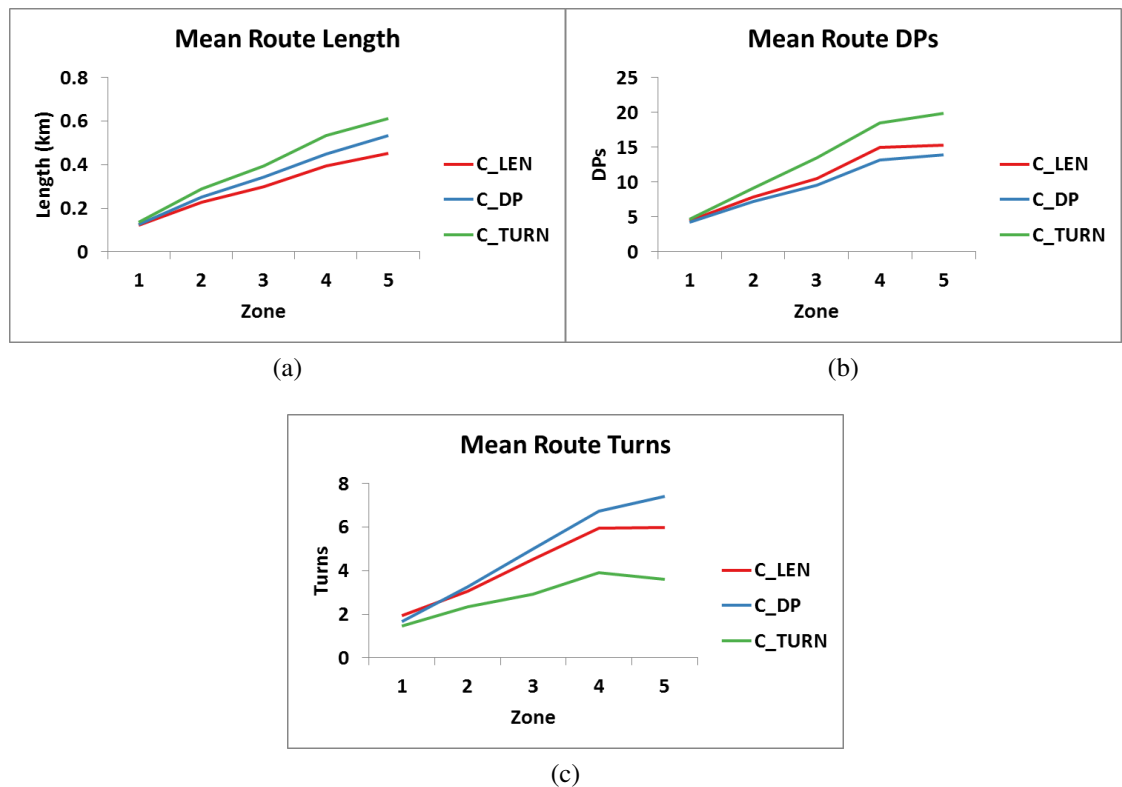
Figure 5.3: Attribute Means - showing the mean route (a) length, (b) DPs and (c) turns for the routes generated by the length (C_{LEN}), DPs (C_{DP}) and turns (C_{TURN}) cost algorithms.

Using the cost functions in Equation 5.1, Equation 5.2 and Equation 5.3 to calculate the route cost in Algorithm 1, routes with either the shortest lengths, minimum decision points or minimum turns were selected. The results of these initial tests are shown in Figure 5.3 and Figure 5.4. Figures 5.3a, 5.3b and 5.3c show how each mean attribute value varies according to the attribute being used to determine cost. All indicate that restricting the value of one attribute produces a route in which the other two are increased. For

example, the minimum length (C_{LEN}) algorithm selects the shortest routes and therefore has the lowest mean length, but these routes contain more decision points than the C_{DP} algorithm and up to double the number of turns compared to the C_{TURN} algorithm.

Figure 5.4 shows examples of the routes selected by the tool for each of the different cost functions. They indicate that not only do different attributes produce very different routes in some cases, but that routes may vary a lot from one zone to the next, despite the points being relatively close. In order to establish just how much variation there is between two routes, the similarity metric discussed in Section 5.2 will be used.

(a)   (b)

(c)
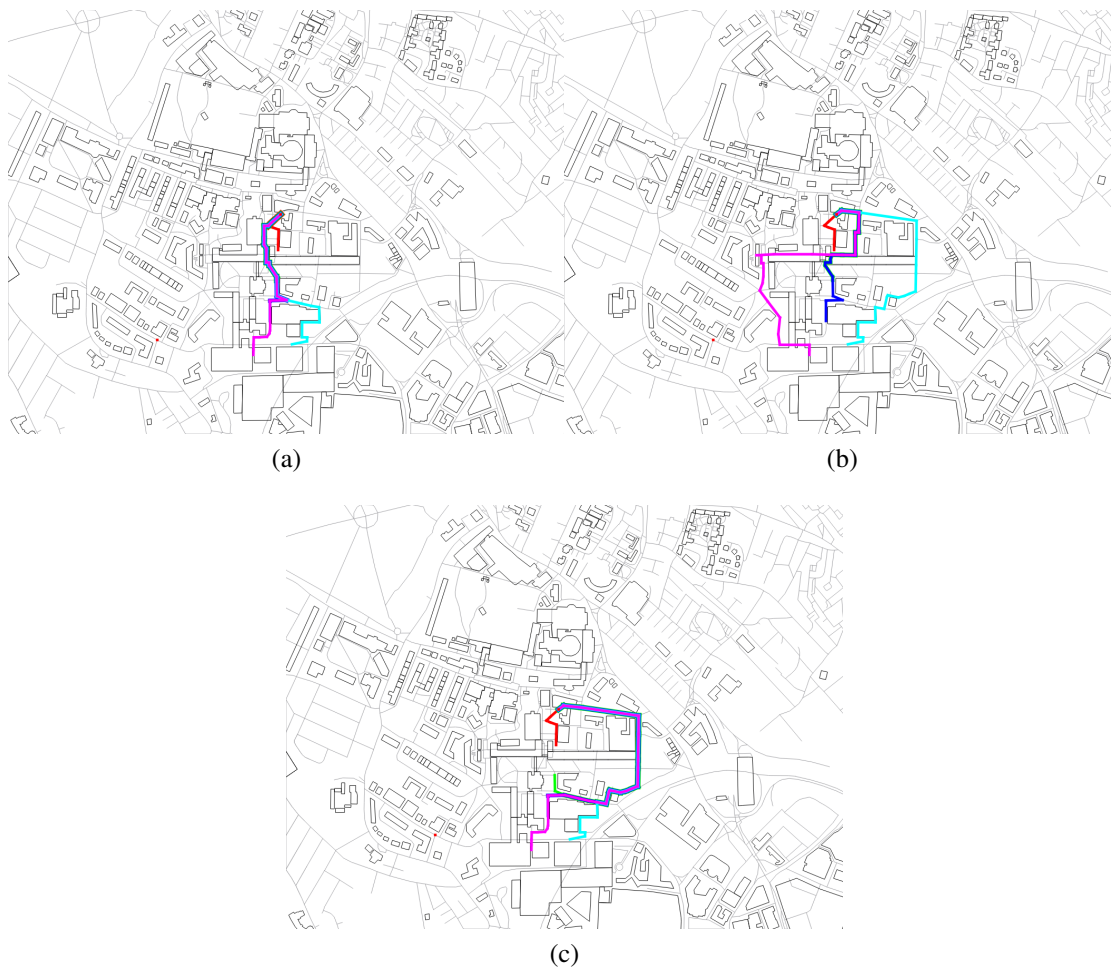
Figure 5.4: Example routes, for a single start point and sector, generated by the (a) length (C_{LEN}), (b) DPs (C_{DP}) and (c) turns (C_{TURN}) cost algorithms. The different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

In the current application similar routes are deemed to be better than non-similar ones, so similarity is calculated to be the percentage length of links which are the same in both routes [18, uses a similar set of criteria to find a set of best routes for vehicular travel].

To do this, routes with the same start point but end points lying in the same sector and in different zones were compared. This is equivalent to routes starting at the same point and travelling in roughly the same directions, but with end points moving gradually further away. Each route is compared against the routes for the previous zones (i.e. Zone 2 routes are compared to Zone 1 routes, Zone 3 routes to those for Zones 1 and 2, e.t.c), and the total length of the common links is converted into a percentage. The maximum route overlap is the length of the shorter of the two routes being compared, which will give a maximum similarity of 100%.
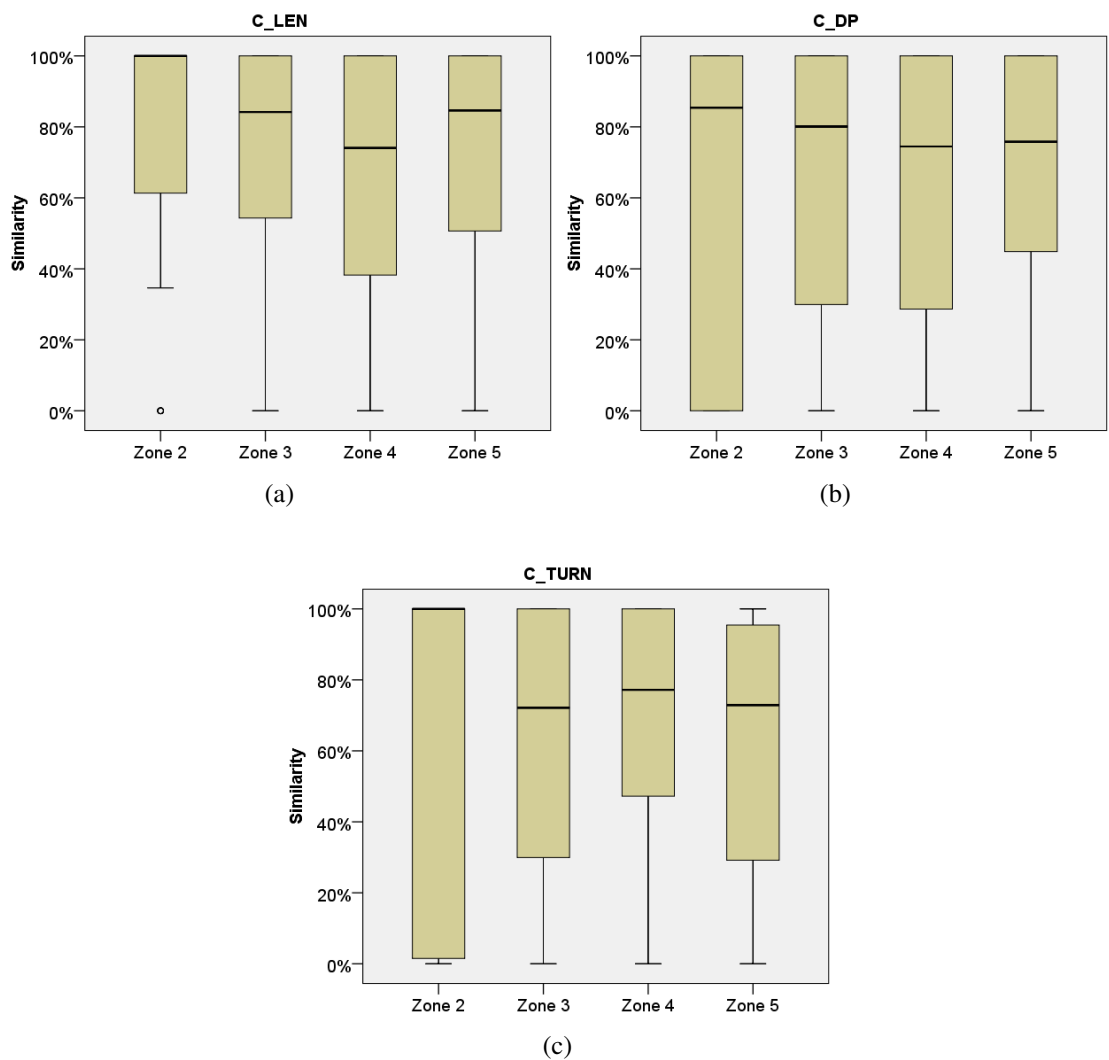
Figure 5.5: Route Similarities - showing the percentage route overlap for routes produced by the (a) length (C_{LEN}), (b) DPs (C_{DP}) and (c) turns (C_{TURN}) cost algorithms. Boxes indicate 25th to 75th percentiles and whiskers indicate these percentages $\pm$ 1.5*interquartile range, 5.5a also shows an outlier outside of this range.

Figure 5.5 shows box plots for the similarity of each of the single attribute cost algorithms. It clearly indicates that the shortest length cost function produces routes with much higher similarity than either decision points or turns, with the exception of the similarity for Zone 4 of the minimum turns cost function. Although these results and those comparing mean route attributes (Figure 5.3) seem to indicate that the shortest length algorithm produces the most appropriate routes of the three tested, Chapter 4 indicates that length, decision points and turns are not the only factors to influence route choice.

5.3.2 Maximum Value Cost Functions

Finding a route containing the highest value of an attribute is a more complicated problem than finding one containing the lowest value of an attribute, requiring large processing power and long computing time. For example, selecting a route with maximum length (the Longest Path Problem) is NP-hard meaning that it cannot be solved in polynomial time [117]. To overcome this issue, creating cost functions that have lower cost as the attribute value rises allows the minimum cost approach to again be used.

From Table 5.1 the maximum value attributes are those with a positive relationship. Points of interest will be considered first here, and used to show some of the problems and solutions involved in selecting routes for these attributes. How these solutions can also be applied to vegetation, land and dwellings will then be discussed.

5.3.2.1 Points of Interest

Although the cost function for points of interest needs to produce a lower value as the number of points of interest increases, this still leaves many possible solutions. The most obvious solution is for the cost to be the reciprocal of the number of points of interest on a link, but the case of zero points will have to be handled. Equation 5.4 and Equation 5.5 show two different approaches to this.

$$C1_{POI} = \sum CLink_{POI} = \sum \frac{1}{POI_{Link} + 1} \tag{5.4}$$

$$
\begin{aligned}
C2_{POI} = \sum CLink_{POI} &= \sum \frac{1}{POI_{Link}} && \text{if } POI_{Link} > 0 \\
&= \sum 10000 && \text{if } POI_{Link} = 0 \tag{5.5}
\end{aligned}
$$

Equation 5.4 simply adds one to the number of points of interest on a link, giving a

weight between zero and one and a maximum cost equal to the number of links in the route. In contrast, Equation 5.5 applies a much larger penalty cost (10000) to any link with no points of interest, which should produce a much lower cost on routes with large numbers of points of interest. It was expected that both of these cost functions would produce routes containing points of interest, with $C2_{POI}$ routes likely to contain more as the higher penalty would create an attraction to these points. However, the histograms in Figure 5.6 indicate that this is not the case.
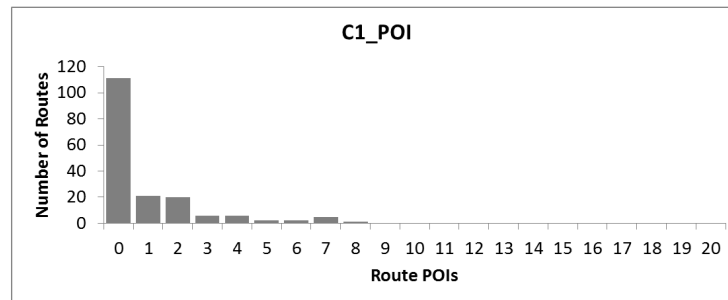
(a)
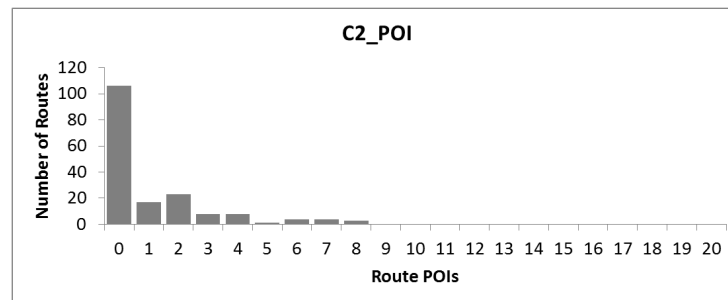
(b)

Figure 5.6: Histograms for (a) $C1_{POI}$ and (b) $C2_{POI}$ showing the distribution of selected routes containing 0 - 20 POIs.

Despite $C2_{POI}$ producing slightly better results than $C1_{POI}$, the majority of routes selected by both cost functions contain no points of interest. Although unexpected, an examination of the environment analysis in Figure 4.2 (Section 4.2.2) provides an explanation for these results. As only 6% of the links in the representation have a non-zero points of interest value, longer routes (with many zero value links) are required to reach links containing any points of interest. In practice, the algorithm is likely to select routes containing fewer links in total rather than detouring to include links with points of interest, as the benefit of including these links is outweighed by the cost of reaching them and then travelling to the end point.

So, what if instead of only reducing the cost of the links containing the points of interest, the cost of any links following the discovery of a point of interest is also reduced?

This removes the cost of travelling between points of interest, and travelling from the point of interest to the end point, which may be enough to attract routes to links with a high point of interest value. Equation 5.6 reflects this as it replaces the number of points of interest on the link with the number on the entire route.

$$C3_{POI} = \sum CLink_{POI} = \sum \frac{1}{POI_{Route} + 1} \tag{5.6}$$

Figure 5.7 indicates that this cost function is better than $C1_{POI}$ and $C2_{POI}$. However, $C3_{POI}$ still produces many routes which contain no points of interest. This is mainly due to the large cost of reaching the points of interest in the first place. Calculating the cost of the route regardless of the number of links within it would remove the penalty of zero value links required to reach points of interest, and would allow a reduction in the entire route cost for any routes containing points of interest.
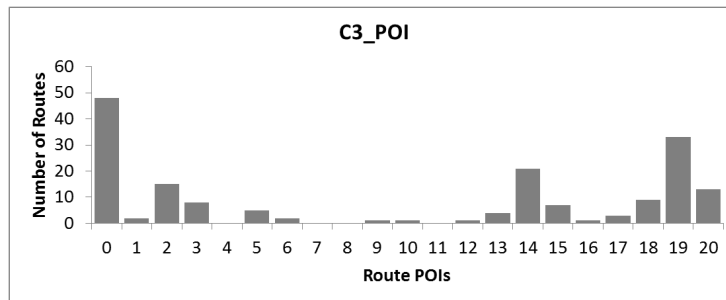
Figure 5.7: Histogram for $C3_{POI}$ showing the distribution of selected routes containing 0 - 20 POIs.

Equation 5.7 shows a cost function which is no longer tied to the number of links in a route, and is determined only by the number of points of interest along it. It also avoids the divide by zero problem by removing the division altogether. The results of running tests using this cost function are shown in Figure 5.8, indicating that it performs much better than the cost functions tested previously. It produces less than half the number of routes containing no points of interest (48) compared to $C1_{POI}$ (111), and 26 routes contain the maximum 20 points of interest. Although lower, there are still many zero value routes, which is because the cost of a route is not reduced until a point of interest is encountered. This effect also reduces Dijkstra's algorithm to a heuristic, which finds routes giving a high points of interest value where possible but not necessarily the highest. An example of why this occurs is shown in Example 5.1.

$$C4_{POI} = POI_{Max} - POI_{Route} \tag{5.7}$$
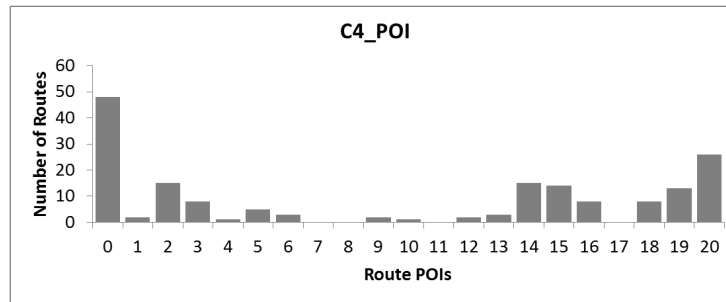
Figure 5.8: Histogram for $C4_{POI}$ showing the distribution of selected routes containing 0 - 20 POIs.

Example 5.1: Environment showing how Algorithm 1 and Equation 5.7 fail to finding the highest POI route due to an early low cost link on another route.

$$C = POI_{Max} - POI_{Current}$$
$$POI_{Max} = 3$$
$$C_{SA} = 3 - 1 = 2$$
$$C_{SB} = 3 - 0 = 3$$
$C_{SA} < C_{SB}$, so expand A
$$C_{SAE} = 3 - 1 = 2$$
$C_{SB} > C_{SAE}$, so stop
But $C_{SBE} = 3 - 2 = 1$
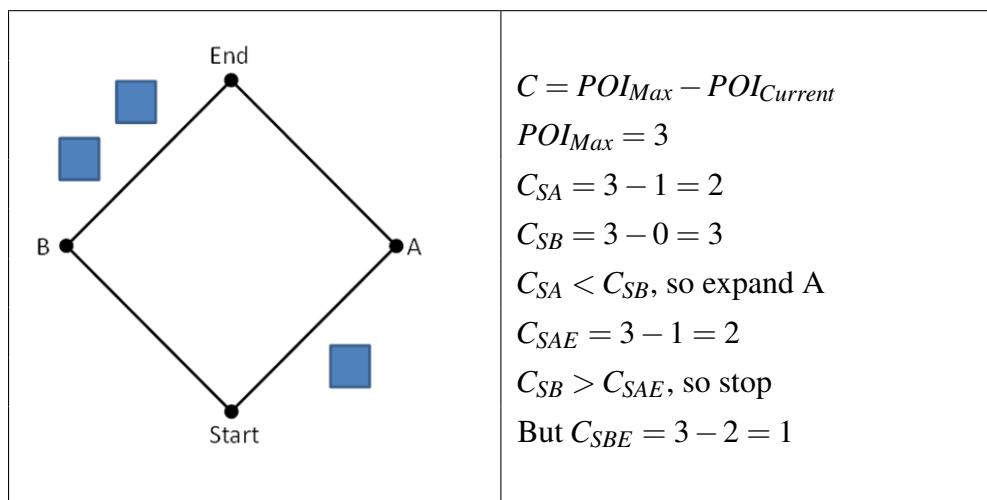
The environment illustrated in Example 5.1 contains three points of interest (shown as blue boxes) and two alternative routes, and is a simplification to show the problems encountered when selecting routes for this attribute. Applying Algorithm 1 to this environment would produce the five steps shown, and incorrectly return the route *Start-A-End* as the best solution. So how can the algorithm be improved to correct this issue?

One approach would be to extend each of the partial routes until it reaches the end point, but Example 5.2 shows why this approach doesn't provide a good solution. Here each of the routes is expanded after the end point has been found, but as each of the nodes is marked as explored when they are expanded, the route containing the most points of interest (*Start-B-D-End*) is never found. In addition, if the POI between the start and node A was removed, the order in which nodes with the same cost are expanded (whether or not a first found-first explored approach is used) may also mean that the best route is ignored. Removing the explored flag would resolve this, but would lead to a brute-force approach

in which all possible non-cycling routes through the environment would need to be found. This is equivalent to the Longest Path Problem, which is NP hard.

Example 5.2: Environment showing how Algorithm 1 and Equation 5.7 fail to find the highest POI route due to the order in which routes are extended.

$POI_{Max} = 3$

$C_{SA} = 3 - 1 = 2$

$C_{SB} = 3 - 0 = 3$

$C_{SA} < C_{SB}$, so expand A

$C_{SAC} = 3 - 1 = 2$

$C_{SAC} < C_{SB}$, so expand C

$C_{SACD} = 3 - 1 = 2$

$C_{SACE} = 3 - 1 = 2$

$C_{SACD} < C_{SB}$, so expand D

$C_{SACDB} = 3 - 2 = 1$

$C_{SACDE} = 3 - 1 = 2$

$C_{SACDB} < C_{SB}$ and $C_{SACDB} < C_{SACDE}$, so expand B

$C_{SACDBS} = 3 - 2 = 1$

All nodes have been explored, so stop
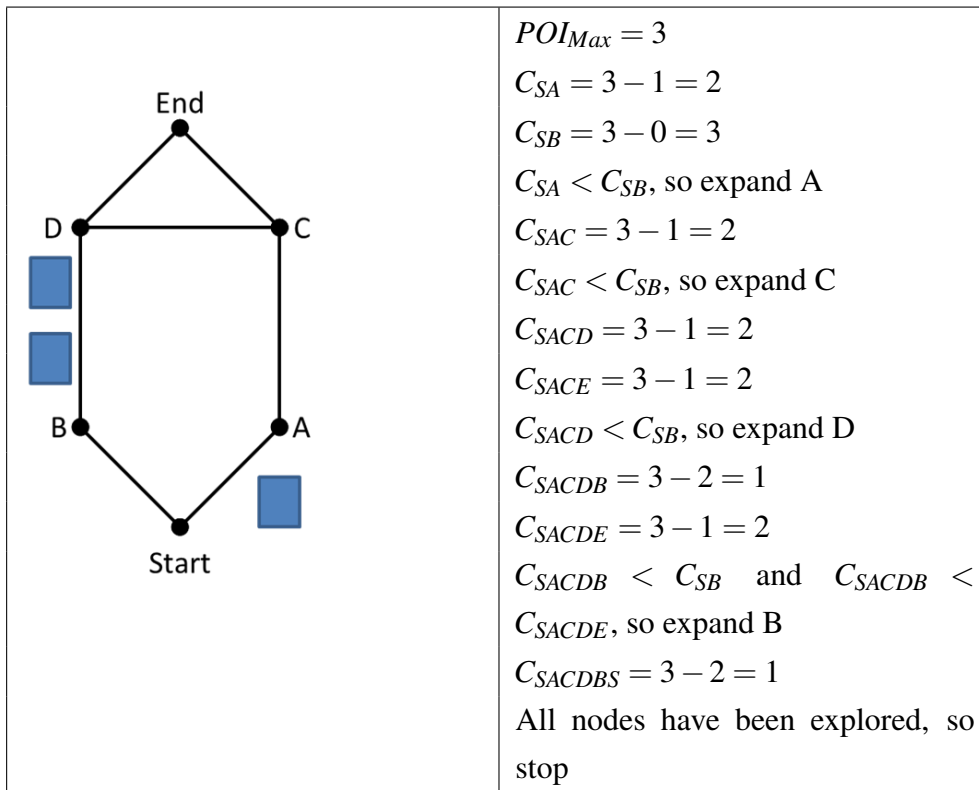
An alternative approach to using brute-force, is to consider the nature of the weights generated themselves. Looking again at Example 5.1, we see that the cost of travelling from S to B is three $C_{SB} = 3 - 0 = 3$, but the cost of travelling from from S to B to E is only one $C_{SBE} = 3 - 2 = 1$ implying that the cost of travelling from B to E is actually -2. Dijkstra's algorithm relies on all costs being positive, and here this is clearly not the case. This negative cost indicates that algorithms suitable for graphs with negative weights must be considered for these maximum value cost functions.

The drawbacks of using common negatively weighted graph algorithms such as Johnson's algorithm [112], the Bellman-Ford algorithm [16] and Floyd's algorithm [68] have already been discussed in Section 2.5. So can we find an efficient heuristic approach that will give us good, but not necessarily the best routes? Looking again at Example 5.1, if the start and end points were switched and Algorithm 1 run, then the correct route would be found (albeit reversed).

```
function Dijkstra_Repeated(StartNode, EndNode):
sRoute* pBestRoute=Dijkstra(StartNode,EndNode);
sRoute* pBestRoute2=Dijkstra(EndNode,StartNode);
if(pBestRoute2->fCost<pBestRoute->fCost || (pBestRoute2->fCost==
pBestRoute->fCost && pBestRoute2->fTotalLength<
pBestRoute->fTotalLength)
   return pBestRoute2;
else
  return pBestRoute;
```

Algorithm 2

Running Algorithm 1 in both directions, as shown in Algorithm 2, would provide the best solution regardless of the start and end positions, and leads to the steps shown in Example 5.3.

Example 5.3: Environment showing how Algorithm 2 and Equation 5.7 overcome the problem of an early low cost link.

$POI_{Max} = 3$

Forwards:

$C_{SA} = 3 - 1 = 2$

$C_{SB} = 3 - 0 = 3$

$C_{SA} < C_{SB}$, so expand A

$C_{SAE} = 3 - 1 = 2$

$C_{SB} > C_{SAE}$, so stop

Backwards:

$C_{EA} = 3 - 0 = 3$

$C_{EB} = 3 - 2 = 1$

$C_{EB} < C_{EA}$, so expand B

$C_{EBS} = 3 - 2 = 1$

$C_{EA} > C_{EBS}$, so stop
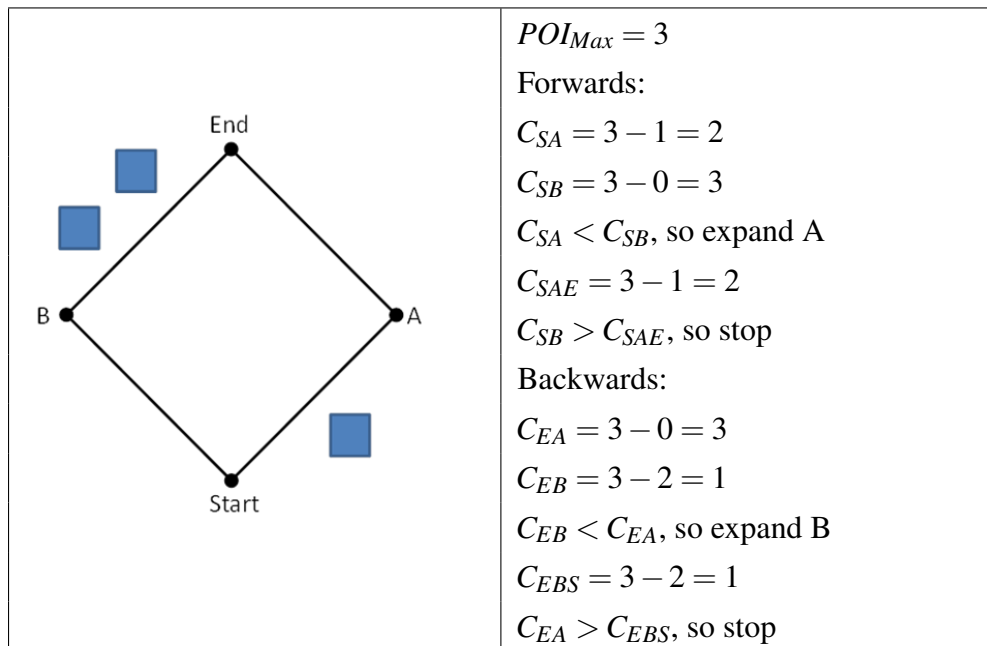
The results of running tests using Algorithm 2 and the cost function shown in Equation 5.7 are shown in Figure 5.9. This indicates that Algorithm 2 produces mixed results compared Algorithm 1. Although the overall number points of interest is slightly lower, the number of routes containing the maximum number is doubled. In addition, of the 174 routes selected, 48% contain less than half of the points of interest in the environment.
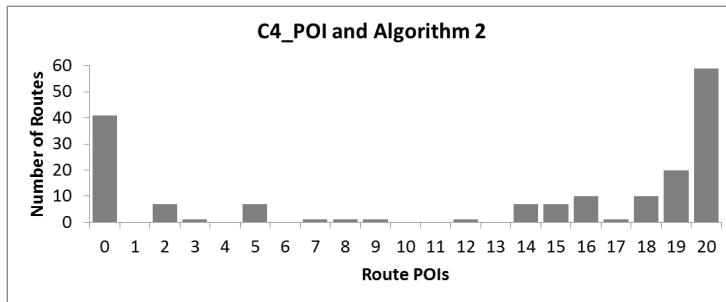
Figure 5.9: Histogram for Algorithm 2 and cost function $C4_{POI}$ showing the distribution of selected routes containing 0 - 20 POIs.

Algorithm 2 is still a heuristic and Example 5.4 shows a scenario in which the best route is still not found.

Example 5.4: Environment showing how Algorithm 2 and Equation 5.7 fail to find the highest POI route due to two early low cost links.

$POI_{Max} = 5$

Forwards:

$C_{SA} = 5 - 1 = 4$

$C_{SB} = 5 - 0 = 5$

$C_{SA} < C_{SB}$, so expand A

$C_{SAC} = 5 - 1 = 4$

$C_{SAC} < C_{SB}$, so expand C

$C_{SACE} = 5 - 2 = 3$

$C_{SB} > C_{SACE}$, so stop

Backwards:

$C_{EC} = 5 - 1 = 4$

$C_{ED} = 5 - 0 = 4$

$C_{EC} < C_{ED}$, so expand C

$C_{ECA} = 5 - 1 = 4$

$C_{ECA} < C_{ED}$, so expand A

$C_{ECAS} = 5 - 2 = 3$

$C_{ED} > C_{ECAS}$, so stop
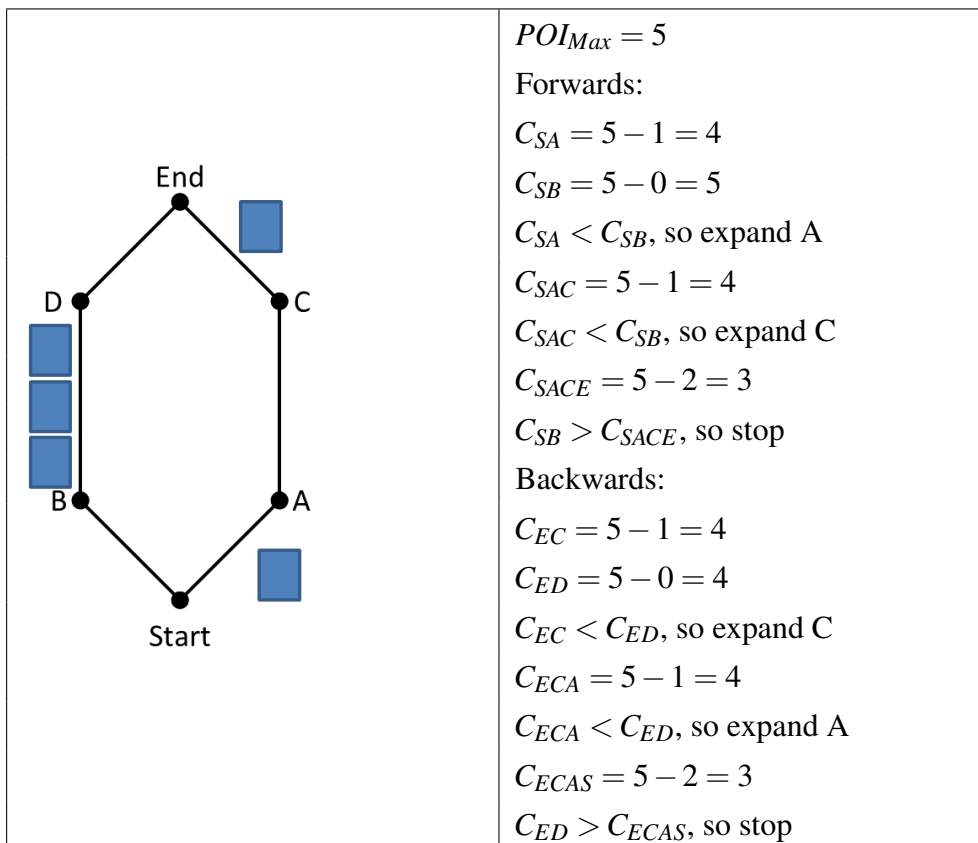
Forty-eight of the routes contain no points of interest at all, suggesting that the scenario illustrated in Example 5.4 is not the only reason for the low results. Two possible additional reasons are available for the presence of zero-value routes. Firstly, the layout of the graph

```
function Dijkstra2(StartNode, EndNode):
sRoute* pBestRoute=NULL;
m_fMinCost=m_fMinLength=DBL_MAX;
for(int i=0; i<m_pStartNode->vpLinks.size(); i++)
    sPartRoute* pNewRoute=
      new sPartRoute(pStartNode,pStartNode->pLinks[i]);
    CalculateCost(pNewRoute);
    AddToExplore(pNewRoute);
while (m_vToExplore.size()>0 && m_fMinCost>0)
    sPartRoute*  pCurrentRoute=m_vToExplore.front();
    cLink* pLink=(cLink*)pRoute->vPrevParts.back();
    cNode* pLastNode=(cNode*)pRoute->GetLastNode();
    cNode* pNode=pLink->GetOtherEnd(pLastNode);
    if(pNode!=m_pEndNode && pRoute->fCost<pNode->m_fMinCost)
        for(unsigned int i=0; i<pNode->pLinks.size(); i++)
            cLink* pLink2=pNode->pLinks[i];
            if(pLink2 && !pRoute->RouteContains(pLink2)
              && !pRoute->RouteContains(pNode))
                sPartRoute* newPart=new sPartRoute(pRoute,pNode,pLink2);
                CalculateCost(newPart);
                pNode->m_fMinCost=newPart->fCost;
                AddToExplore(newPart);
    else
       if(pNode==pEndNode)
          pRoute->vPrevParts.push_back(pNode);
          if(pCurrentRoute->fCost<m_fMinCost || (pCurrentRoute->fCost==
            m_fMinCost && pCurrentRoute->fLength<fMinLength))
            m_fMinCost=pCurrentRoute->fCost;
            fMinLength=pCurrentRoute->fLength;
            pBestRoute=ConstructRoute(pCurrentRoute);
          else
            delete(pCurrentRoute);
delete(pCurrentRoute);
if(m_vToExplore.size()!=0)
   m_vToExplore.clear();
return pBestRoute;
```

Algorithm 3

being used may mean that there are only routes containing no POIs between the two points, and further testing indicates that this is the case for one route. Secondly, the order in which equal cost routes are extended may mean that all nodes are explored before a route containing POIs is found, as explained in Example 5.2.

As marking nodes as explored has such a big influence on the performance of Algorithm 2, then an alternative was investigated where nodes are instead annotated with the cost of the route up to that point. Using this approach, the cost of any partial routes reaching this node can be compared to the annotated cost and any with lower cost expanded. This converts the algorithm to a label correcting approach, with some similarities to the Bellman-Ford Algorithm [16]. Unlike Dijkstra's Algorithm, the first partial route to reach a node does not necessarily determine the final cost. Here any routes encountered with less cost to reach the same node are allowed to replace the original, as long as no cycles exist. However, in contrast to the actual Bellman-Ford Algorithm, the algorithm is not run multiple times. The resulting approach is shown in Algorithm 3.

Although Algorithm 3 removes the problem of no unexplored nodes remaining, the order in which nodes and links are added to the route and the requirement of non-cycling routes will have a similar effect. If a link can only be included in a route once, then a route may reach a point at which it can no longer be extended before any POIs have been discovered, and equally a route containing many POIs may not be able to reach its end point as no links that are not already included are available. Ideally Algorithm 3 would be modified to extend any routes reaching a node which also have an equal cost, but the computing time of this approach is again prohibitive. As a compromise, Algorithm 4 again runs the algorithm in both directions across the environment in a similar way to Algorithm 2.

```
function Dijkstra_Repeated2(StartNode, EndNode):
sRoute* pBestRoute=Dijkstra2(StartNode,EndNode);
sRoute* pBestRoute2=Dijkstra2(EndNode,StartNode);
if(pBestRoute2->fCost<pBestRoute->fCost ||
  (pBestRoute2->fCost==pBestRoute->fCost &&
  pBestRoute2->fTotalLength<pBestRoute->fTotalLength))
    return pBestRoute2;
else
  return pBestRoute;
```

Algorithm 4

The results of tests using Algorithm 3 and Algorithm 4 are shown in Figure 5.10. Figure 5.10a indicates that far more routes containing 20 POIs are found by Algorithm 3 than Algorithm 2, and far fewer routes containing no POIs at all. Although this is a sizeable improvement, Figure 5.10b shows that Algorithm 4 produces substantially better results than any of the previous tests.
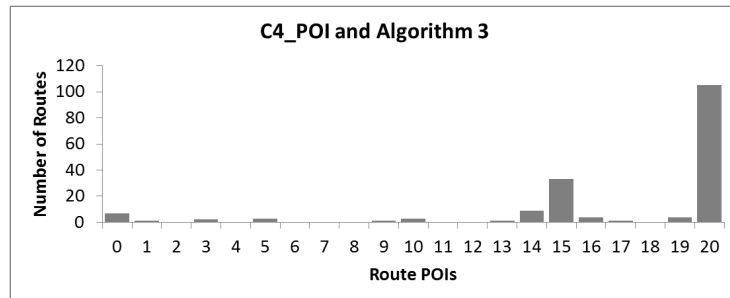
(a)
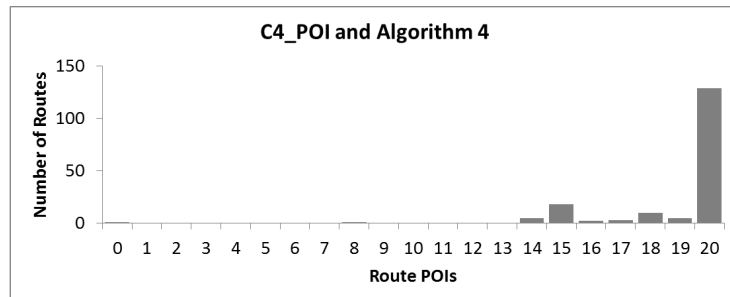
(b)

Figure 5.10: Histograms for (a) Algorithm 3 and (b) Algorithm 4 showing the distribution of selected routes containing 0 - 20 POIs.

Despite the increases in both the number of POIs overall, and the number of routes containing the maximum of 20, Algorithm 4 still selects seven routes with fewer than 15 POIs. The main reason for this is the order in which the nodes are expanded as explained previously, which is best illustrated by the POI frequency shown in Figure 5.11. All of the algorithms in this section up to this point have used a FIFO approach to node expansion - with nodes of the same cost discovered earlier being expanded first. Figure 5.11 shows the results of $C4_{POI}$ with Algorithm 4, but this time with a LIFO ordering approach (nodes of the same cost discovered later expanded first). The vast reduction in performance is a stark example of the problems encountered.
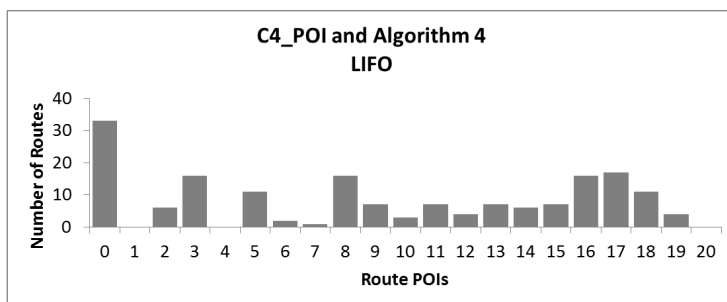
Figure 5.11: Histogram for Algorithm 4 with LIFO ordering showing the distribution of selected routes containing 0 - 20 POIs.

An overview of these results is shown in Table 5.2. From this, $C4_{POI}$ with Algorithm 4 clearly produces better results than the other options, and was therefore selected as the most appropriate.

Cost Function	Algorithm	Total POIs	Routes with zero POIs	Routes with max POIs
$C1_{POI}$	1	168	111	0
$C2_{POI}$	1	200	106	0
$C3_{POI}$	1	1691	48	13
$C4_{POI}$	1	1653	48	26
$C4_{POI}$	2	2208	41	59
$C4_{POI}$	3	2952	7	105
$C4_{POI}$	4	3286	1	129
$C4_{POI}$ LIFO	4	1579	33	0

Table 5.2: Overview of the cost functions and algorithms for POI attribute route selection including evaluation metrics giving the total number of POIs found by all test routes, the number of test routes containing no POIs and the number of test routes containing the maximum 20 POIs.

For ease of understanding, the combination of $C4_{POI}$ and Algorithm 4 will be referred to as C_{POI} from this point forward, and the results of testing this are shown in Figure 5.12. Figure 5.12a indicates that many more points of interest are found by this algorithm than are present on the routes selected by any of the earlier attribute cost functions (C_{LEN}, C_{DP} or C_{TURN}). However, Figure 5.12b shows that this increase corresponds to a substantial increase in the length of the routes selected. In fact, the shortest length for a route containing 20 POIs selected by this algorithm is 2.18km, and for routes containing 10 or more POIs is 1km. Figure 5.12c shows examples of these routes for a single start point and multiple end points, indicating that, for the majority of start-end pairs, the selected routes stick to the central area in which most of the points of interest are located.

(a)   (b)

(c)
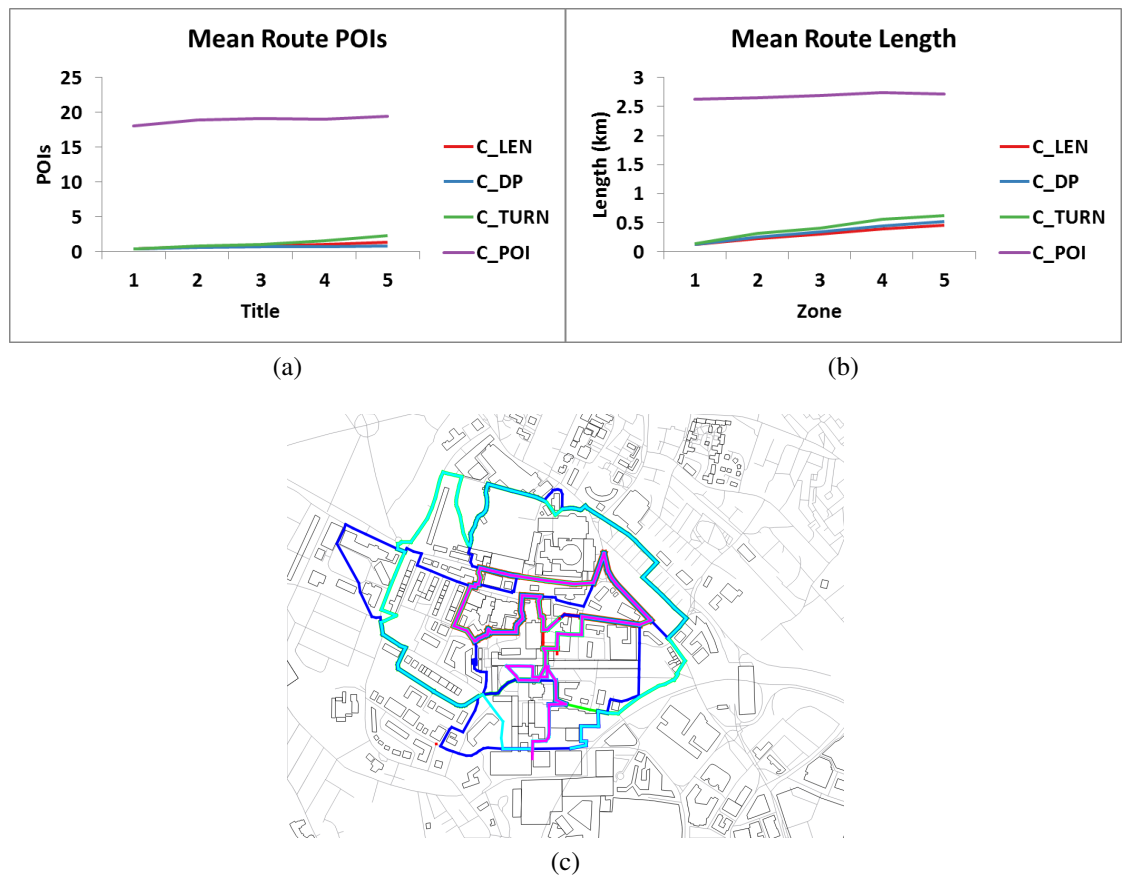
Figure 5.12: Attribute Means - showing the mean route (a) POIs and (b) length for the routes generated by the POI (C_{POI}) cost algorithm. (c) Shows example routes produced by C_{POI}, where the different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

5.3.2.2   Vegetation, Land Use and Dwellings

Despite being represented very differently, the continuous attributes (vegetation, land use and dwellings) have many of the same issues as points of interest. Initial testing indicates that they also suffer from the overwhelming effects of the links required to reach high attribute areas, and that the heuristics used to overcome this also encounter the same problems with the order in which routes are extended. The most appropriate cost functions were found to be those shown in Equation 5.8, Equation 5.9 and Equation 5.10. In common with C_{POI}, C_{VEG}, C_{LAND} and C_{DWEL} rely only on the proportion of the route not covered by the relevant attribute rather than being associated with links, but in this case have a quadratic relationship as indicated by Figure 3.2 (Section 3.4).

$$C_{VEG} = \left(1 - \frac{Vegetation_{Route}}{2 * Length_{Route}}\right)^2 \tag{5.8}$$

$$C_{LAND} = \left(1 - \frac{LandUse_{Route}}{2 * Length_{Route}}\right)^2 \tag{5.9}$$

$$C_{DWEL} = \left(1 - \frac{Dwellings_{Route}}{2 * Length_{Route}}\right)^2 \tag{5.10}$$

Although Algorithm 4 is the most appropriate for selecting routes in terms of points of interest (as shown in Section 5.3.2.1), testing with the minimum length cost function C_{LEN} indicates that Algorithm 4 does not produce a solution for all routes and attributes in polynomial time (no routes were produced in over three hours runtime for C_{LEN}). Therefore, in order to select the most appropriate algorithm for each attribute, the performance of each of the four suggested algorithms using all single attribute cost functions was compared as shown in Table 5.3.

	C_{LEN}	C_{DP}	C_{TURN}	C_{POI}	C_{VEG}	C_{LAND}	C_{DWEL}	$C_{LEN^{-1}}$
Algorithm 1	✓*	✓*	✓*	✓	✓	✓	✓	✓
Algorithm 2	✓*	✓*	✓*	✓	✓*	✓*	✓	✓*
Algorithm 3	✗	✓*	✓*	✓	✓	✗	✓	✗
Algorithm 4	✗	✓*	✓*	✓*	✗	✗	✓*	✗

Table 5.3: Performance of the four test algorithms for each of the single attribute cost functions. ✓algorithm completes within the test time (174 routes in 15 minutes), * algorithm produces good results, ✗algorithm does not complete within the test time.

These results indicate that Algorithm 2 works well for length, decision points, turns, vegetation, and land use, but Algorithm 4 is more suitable for points of interest and dwellings (the attributes which have a large number of zero value links). Running the combination of cost functions and algorithms described for vegetation, land use and dwellings produces the results shown in Figure 5.13, Figure 5.14 and Figure 5.15.

Figure 5.13 suggests that more vegetation (Figure 5.13a) is present on the majority of routes than either parkland (Figure 5.13b) or dwellings (Figure 5.13c). It also suggests that attempting to increase the presence of one attractiveness attribute substantially reduces the presence of the others, although overall C_{VEG} may be argued to give the best results for the continuous attributes. One surprising result is how low the number of points of interest are for C_{DWEL} and C_{LAND} (Figure 5.13d). Many points of interest normally occur in parks and gardens (such as statues or fountains) or may be dwellings themselves; however, for this environment this generalisation is not true as illustrated in Chapter 4.
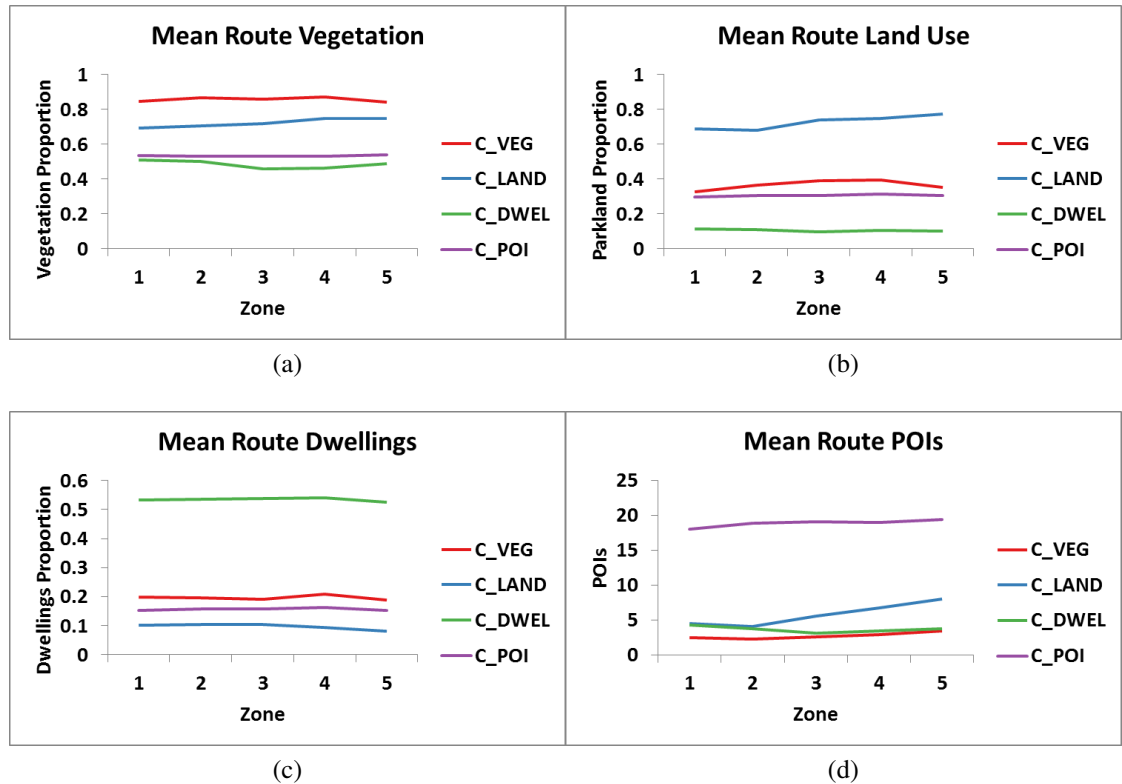
Figure 5.13: Attribute Means - showing the mean route (a) vegetation, (b) parkland, (c) dwellings and (d) POIs for the routes generated by the vegetation (C_{VEG}), land use (C_{LAND}), dwellings (C_{DWEL}) and POIs (C_{POI}) cost algorithms. Proportion indicates the proportion of the route covered by an attribute.

Figure 5.14 shows example routes produced by the cost functions C_{VEG}, C_{LAND} and C_{DWEL}. They clearly indicate that the routes selected follow areas containing high vegetation (Figure 5.14a), parkland (Figure 5.14b) and dwellings (Figure 5.14c) despite this requiring large detours in some cases.

As would be expected from the examples shown in Figure 5.14, the maximum value cost functions give better route similarity for the majority of attributes and zones, as shown in Figure 5.15, than the minimum value cost functions. As only a small number of links contain large values for these attractiveness attributes, it is logical that a large number of the routes selected would be attracted to these links producing high similarity across these routes. This is particularly true for the points of interest (Figure 5.15d), although the routes taken to reach these high value routes may be typically different.

(a)

(b)

(c)
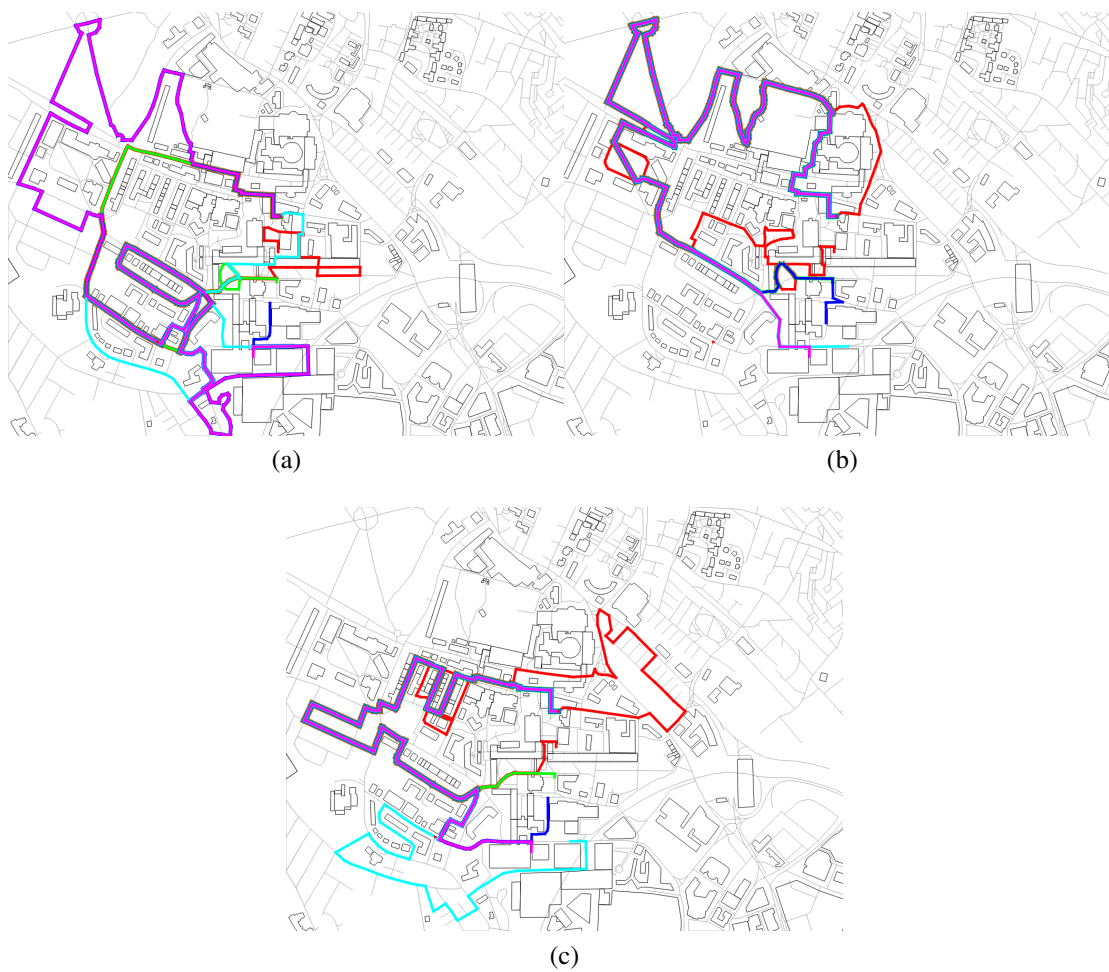
Figure 5.14: Example routes generated, for a single start point and sector, by the (a) vegetation (C_{VEG}), (b) parkland (C_{LAND}) and (c) dwellings (C_{DWEL}) cost algorithms. The different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

5.3.2.3 Length

The minimum length cost function has already been examined, but Figure 3.8b and Figure 3.8c indicate that longer routes are preferred for leisure and tourist routes respectively. Figure 3.8b indicates that although there is a positive relationship between leisure route preference and length, there is likely to be an upper limit to this length (and common sense would suggest that the same would be true of tourist routes). Converting this into a cost function produces Equation 5.11, shown below, where cost decreases as the route length tends towards the limit, and then increases as the length exceeds the limit.

$$C_{LENCAP} = abs(Length_{Limit} - C_{LEN})^2 \qquad (5.11)$$
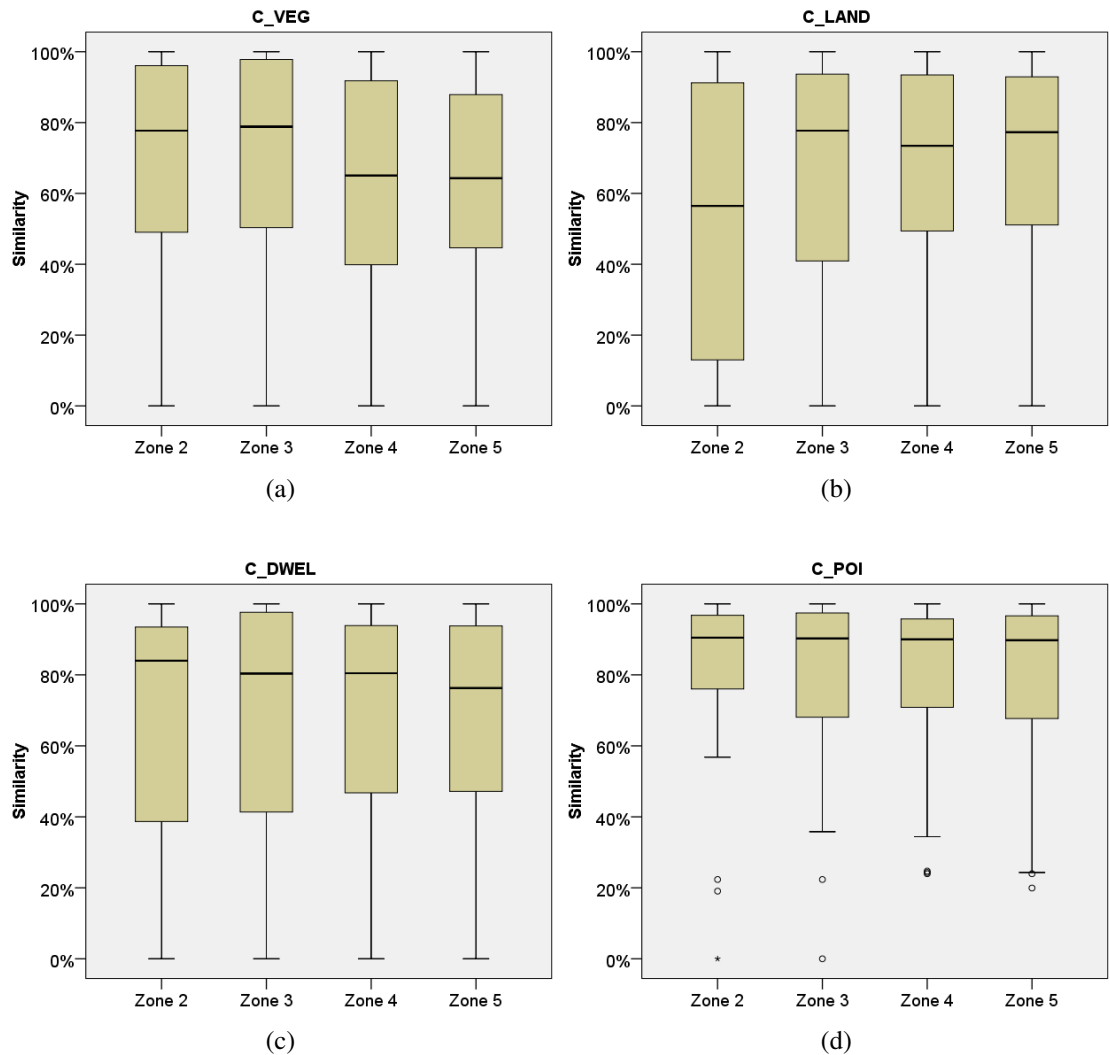
Figure 5.15: Route similarities for the routes generated by the (a) vegetation (C_{VEG}), (b) land use (C_{LAND}), (c) dwellings (C_{DWEL}) and (d) POIs (C_{POI}) cost algorithms. Boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range.

Determining the appropriate value for $Length_{Limit}$ will be discussed later, but for testing purposes values of 2km, 4km and 6km were chosen. Initial tests found that Algorithm 1 was the most appropriate for this attribute, and Figure 5.16 and Figure 5.17 shows the results of tests using this algorithm and the length limits indicated.

Figure 5.16a indicates that despite a tendency towards longer routes, this cost function does not always produce routes which are close to the length of the limit. This is particularly evident in the values for Zone 1 and Zone 2 for the 6km length limit. Two possible reasons for this are that there may only be a small number of routes available between these two points, or that the close proximity of the points may affect the order with which
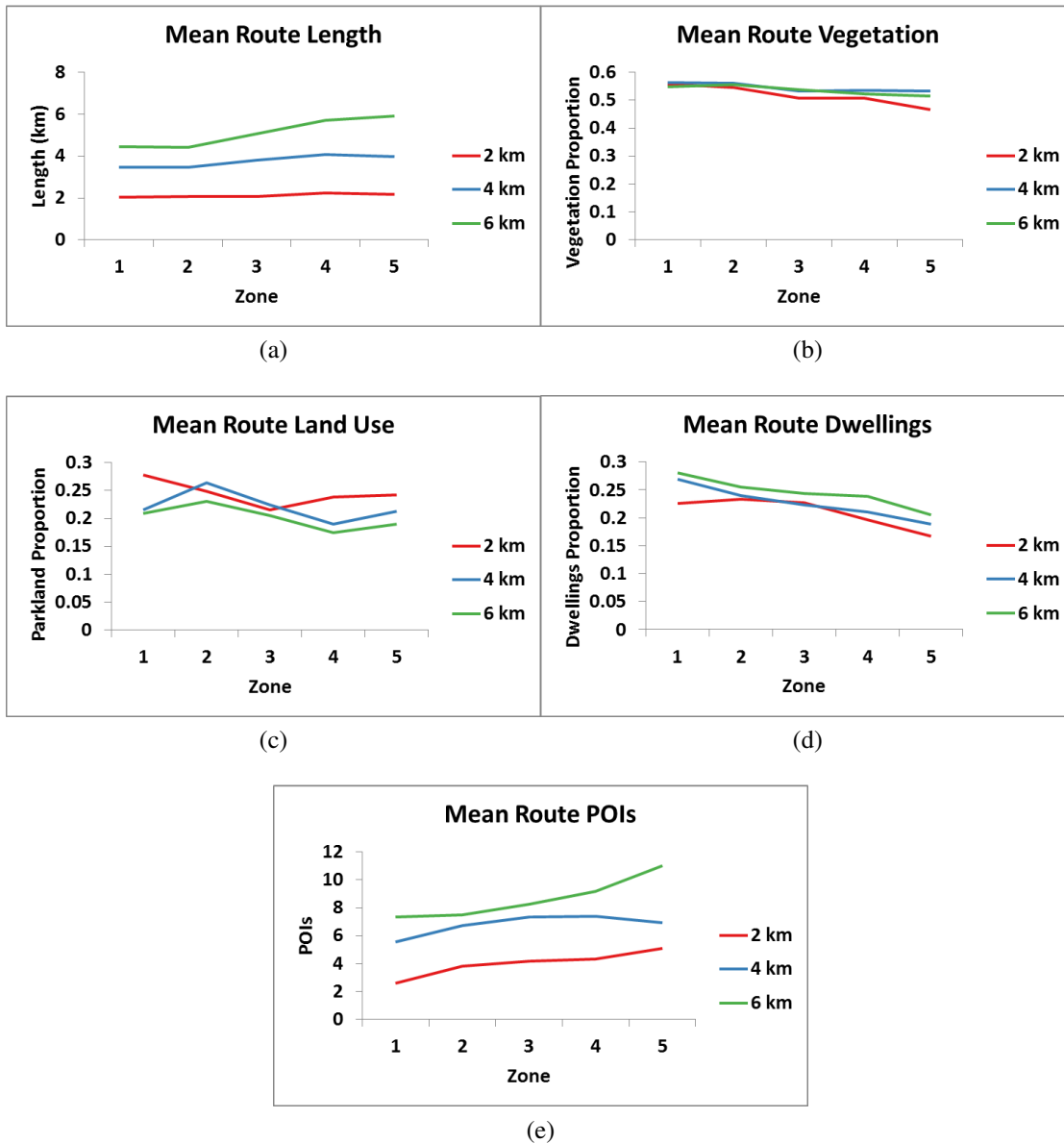
Figure 5.16: Mean route (a) length, (b) vegetation, (c) land use, (d) dwellings and (e) POIs for the capped length cost function C_{LENCAP} for 2km, 4km, and 6km length limits (*Length$_{Limit}$*). Proportion indicates the proportion of the route covered by an attribute.

routes are expanded (both of which are discussed in Section 5.3.2.1). This effect becomes more pronounced as the limiting length increases, but as a suitable length for pedestrian routes is likely to be below 6km, this cost function should be sufficient. In addition, for vegetation (Figure 5.16b) and land use (Figure 5.16c), increasing route length continues to produce comparable or mixed results, whereas as for dwellings (Figure 5.16d) it produces a small increase in performance, and a substantial one for points of interest (Figure 5.16e).
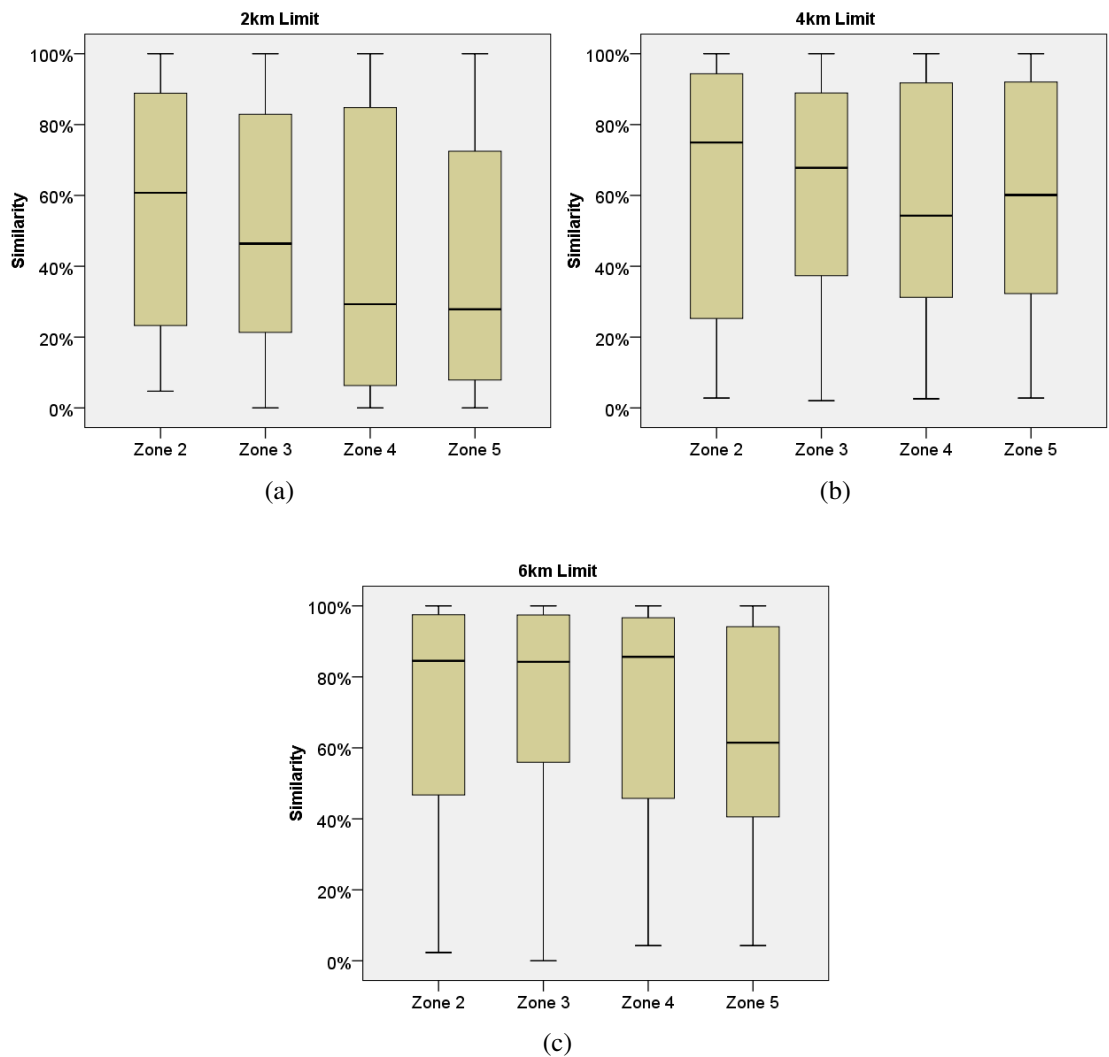
Figure 5.17: Route similarities for the routes generated by the capped length cost function C_{LENCAP} for (a) 2km, (b) 4km, and (c) 6km length limits ($Length_{Limit}$). Boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range.

Figure 5.17 indicates that route similarity increases (Figure 5.17a, Figure 5.17b and Figure 5.17c) as the limiting length increases, which is reflected in the example routes shown in Figure 5.18. As $Length_{Limit}$ increases, more of the map is covered by the routes produced, therefore increasing the route similarity.

(a)

(b)   (c)
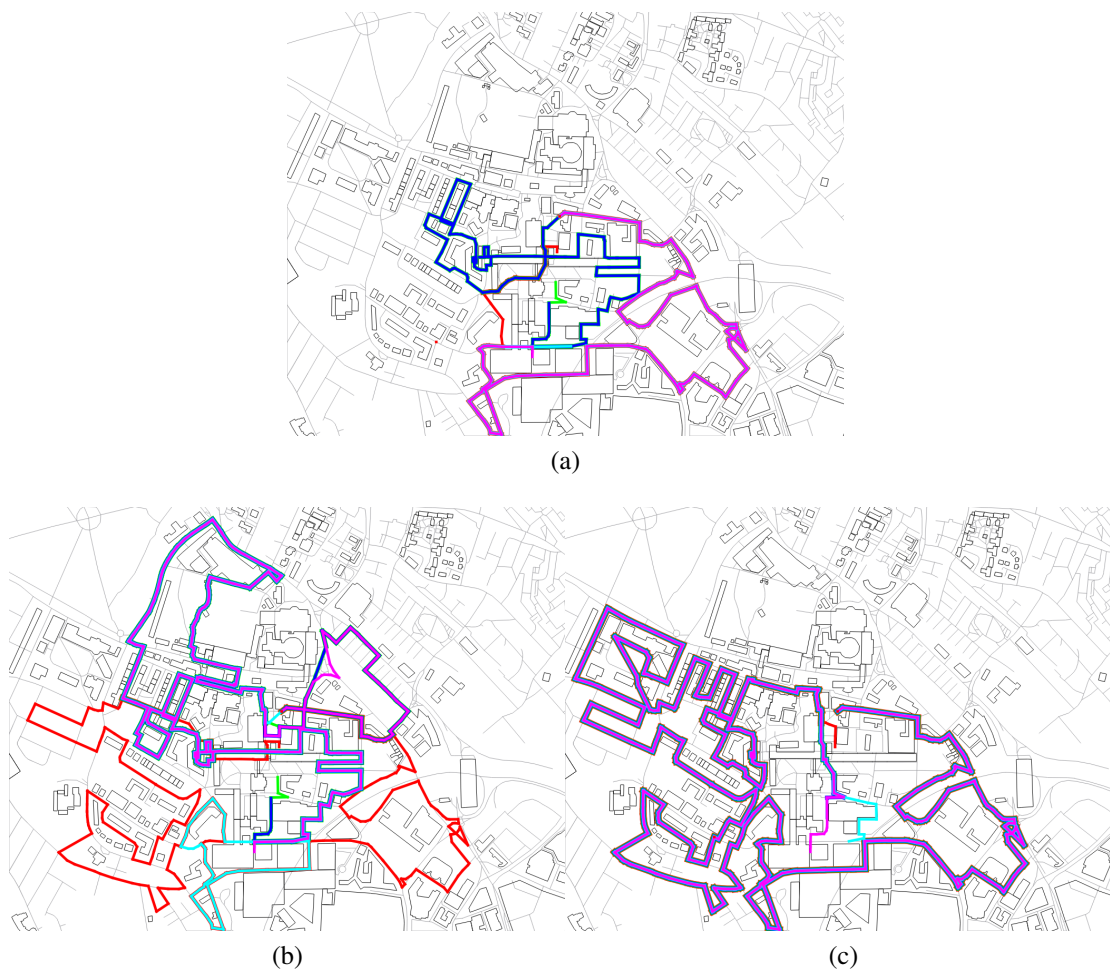
Figure 5.18: Example routes, for a single start point and sector, generated by the capped length cost function C_{LENCAP} for (a) 2km, (b) 4km, and (c) 6km length limits ($Length_{Limit}$). The different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

5.3.3 Summary

Table 5.4 shows a summary of the cost functions and algorithms found to be most appropriate for each of the attributes discussed in this section. Using a single algorithm for all attributes would be more straightforward; however, when using two, the increase in performance outweighs the extra inconvenience. Where indicated, the algorithms used represent heuristics, meaning that it is possible that better routes exist within the environment, but the results found here are considered sufficient.

Attribute	Cost Function	Algorithm	Heuristic
Length	$C_{LEN} = \sum_{1}^{Max} Length_{Link}$	2	✗
DPs	$C_{DP} = \sum_{1}^{Max} DP_{Route}$	2	✗
Turns	$C_{TURN} = \sum_{1}^{Max} Turns_{Route}$	2	✗
POIs	$C_{POI} = POI_{Max} - POI_{Route}$	4	✓
Vegetation	$C_{VEG} = \left(1 - \frac{Vegetation_{Route}}{2*Length_{Route}}\right)^2$	2	✓
Land Use	$C_{LAND} = \left(1 - \frac{LandUse_{Route}}{2*Length_{Route}}\right)^2$	2	✓
Dwellings	$C_{DWELL} = \left(1 - \frac{Dwellings_{Route}}{2*Length_{Route}}\right)^2$	4	✓
Capped Length	$C_{LENCAP} = abs(Length_{Limit} - C_{LEN})^2$	2	✗

Table 5.4: Summary of the single attribute cost functions and algorithms.

Which algorithm to use for these combined attribute route selections is an important question. However Table 5.3 suggests that Algorithm 4 only produces results in polynomial time for attributes in which many of the routes have equal length, which is unlikely to be the case where cost is determined by multiple attributes. With this in mind, the combined attribute algorithms for the differing attribute categories and journey types will all be based on Algorithm 2.

5.4 Optimisation Techniques

One drawback to varying the algorithms and cost calculations as described in the previous section is the increase in processing time over Dijkstra's Algorithm. Although Dijkstra's Algorithm is very efficient, the same is not necessarily true of all of the variants used by the present research. In this section simple speed tests were run to determine the time taken for each algorithm and cost function to select routes.

Tests on algorithm speed were performed, using single threaded code, on a Dell XPS, with a 1.90GHz Intel(R) Core(TM) i7-3517U CPU, and 8GB of RAM. Each algorithm was run used to generate routes between the same 174 start-end point pairs as used in Section 5.3, with the process being repeated five times (giving a total of 870 routes). The mean and maximum route selection times for the five runs of each attribute algorithm were collated, and the results of these tests are briefly discussed with possible improvements proposed.

5.4.1 Unoptimised Algorithms

Table 5.5 shows the running times for each of the single attribute algorithms and cost functions. The majority of these times are reasonable for an algorithm designed to be run in realtime; however, those of the dwellings algorithm may be longer than would normally be acceptable with a maximum of 0.19 seconds.

Attribute	Max (s)	Mean (s)
C_{LEN}	0.0156	0.0018
C_{DP}	0.0156	0.0017
C_{TURN}	0.0156	0.0021
C_{VEG}	0.0156	0.0023
C_{LAND}	0.0156	0.0025
C_{DWELL}	0.1872	0.0104
C_{POI}	0.0156	0.0034
C_{LENCAP} (2m)	0.0156	0.0022
C_{LENCAP} (6m)	0.0156	0.0027

Table 5.5: Mean and maximum running times for the unoptimised algorithms to select 870 routes.

In common with many label setting and label correcting algorithms (Section 2.5.1), the algorithms discussed in this Chapter spend much of their runtime storing the partial routes in ordered lists. As discussed previously, two of the most common forms of optimisation for improving the amount of time taken for these operations, are the use of heaps or buckets [41]. Heap optimisation usual involves the creation of linked lists which form tree-like structures to store the partial routes, whereas bucketing relies on arrays. Previous research [41] has indicated that the Fibonnaci and R-Level heaps, traditionally accepted as being the fastest approaches, perform poorly on complex and real-world graphs, and also indicates that bucketing provides a better solution in these cases. Here two different approaches to bucketing (one-dimensional and two-dimensional) are implemented in an attempt to improve the runtimes seen in Table 5.5.

5.4.2 One-Dimensional Bucket Optimisation

For the one-dimensional bucketing approach, an array of lists (buckets) is used to store the partial routes generated during the running of each algorithm. The cost of the partial route is used to determine which bucket the route will be placed in, with routes being added to the list in ascending order. Where partial routes have equal cost they are added in a last-in-first-out order, and a record is kept of the lowest value bucket currently in use

(to reduce the time taken to select the next route to be processed). The first entry in the lowest value bucket, is therefore always the lowest cost route. An example of how this process works is shown in Figure 5.19.
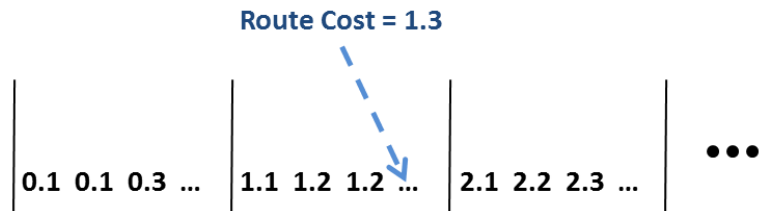
Figure 5.19

Figure 5.20 shows the results of speed tests, run in the same way and with the same equipment as those performed on the unoptimised algorithms. It indicates that five of the algorithms (C_{LEN}, C_{DP}, C_{TURN}, C_{DWELL} and C_{POI}) show improved mean runtimes, although in most cases these decreases are very small (as low as just 4.2% for turns and 4.1% for points of interest). Despite these decreases only C_{DWELL} showed an improvement in the maximum time taken to chose a single route, with the remaining attributes showing no difference between the unoptimised and optimised versions of the algorithms.

Looking at the partial routes generated by the algorithms gives two possible answers as to why the improvements are not greater. The first reason may be that the time spent searching through the list in each bucket is still substantial, especially if many of the partial routes have the same cost. This is particularly common in the C_{DWELL} and C_{POI} algorithms, where the majority of the environment contains a very small amount of the specific attribute or routes which encounter the same number of instances have the same cost. The second problem is the creation and destruction of the lists themselves. Each time a route is added to an empty bucket a new list is created, and each time a bucket is emptied the corresponding list is destroyed. This overhead is repeated many times in the case of algorithms which produce many differing route costs, such as those seen in the C_{LENCAP} algorithms. As the problem of many equal cost routes leads to the highest maximum times to find single routes, as indicated by the results for C_{DWELL}, then a second optimisation technique was tried.

5.4.3 Two-Dimensional Bucket Optimisation

For two-dimensional bucket optimisation the list within each bucket is extended to two dimensions, making the approach more similar to a bucketed heap. By doing this the length of the list within the bucket is reduced, therefore reducing the time to traverse
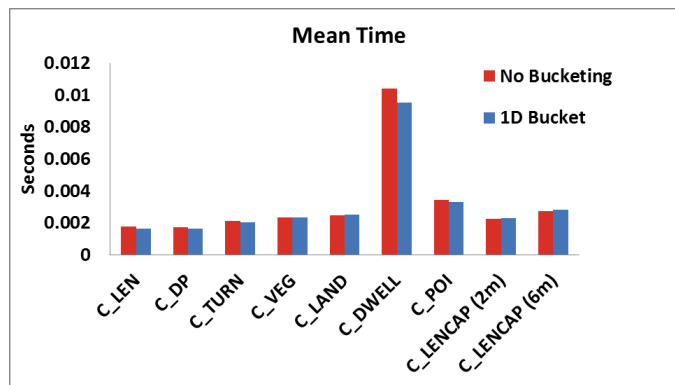
(a)
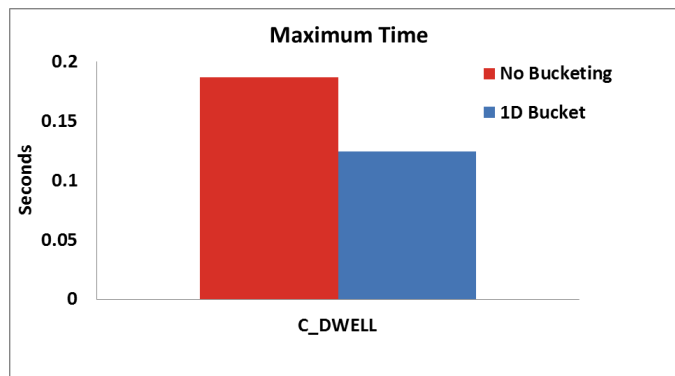
(b)

Figure 5.20: (a) Mean and (b) maximum running times for the unoptimised and 1D bucketed algorithms to select 870 routes.

it. Three different variations on this theme were tried; a variable sized two-dimensional bucket, and two lengths of fixed size two-dimensional buckets. In all cases, the same approach is taken as in one-dimensional bucketing to select the bucket into which the route is to be added, but here each entry in the first list held within a bucket represents the head of a second list. Figure 5.21 shows an example of the variable and fixed length approaches.

For the variable sized bucket approach partial routes with differing lengths are placed in ascending order in the first list in the bucket, forming the heads of the a series of second dimension of lists. All partial routes of equal cost are inserted into the top of the second dimension of the list, replacing the original entry in the first dimension list, therefore pushing any existing entries down.

As with the variable sized bucket, in the fixed sized bucket approach the initial partial route is inserted into the first dimension of the bucket (forming the head of a second dimension list). However, rather than following routes being inserted into the first dimension of the list, they are instead added in ascending order to the second dimension (see

Route Cost = 1.3

0.1 0.2 0.3 ...	1.1 1.2 1.3 ...	2.1 2.2 2.3 ...	•••
0.1	1.3	2.1 2.3	
0.1		2.3	

(a)

Route Cost = 1.3

0.1 0.1 ...	1.1	2.1 2.4	
0.1 0.2	1.1	2.1	•••
0.1 0.2	1.2	2.2	
0.1 0.3	1.2	2.2	
0.1 0.4		2.3	

(b)
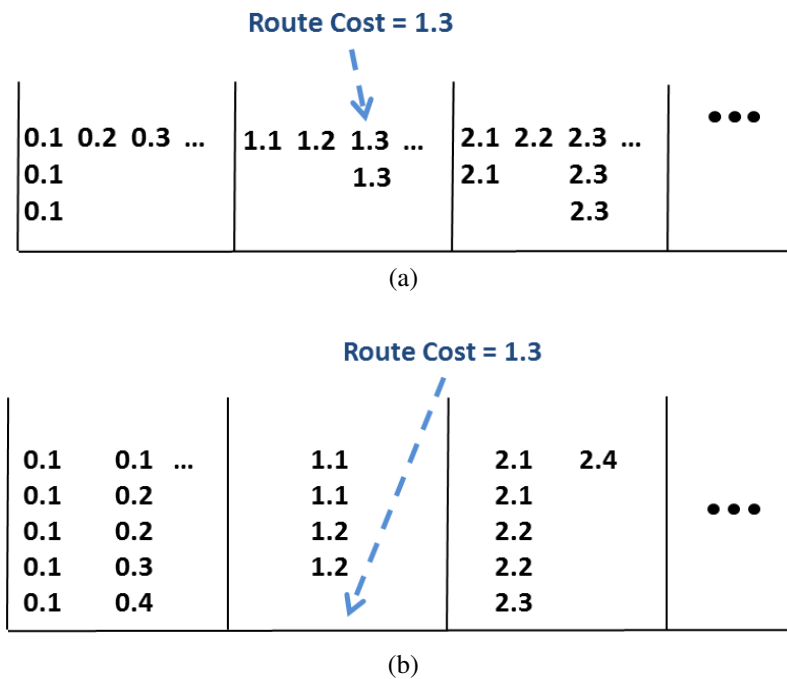
Figure 5.21

Figure 5.21b for an example of this). Once the list reaches a specified size the algorithm moves onto the next list in the bucket, and the process is repeated. In this case, list lengths of five and ten entries were tested.

Again the plots in Figure 5.22, indicate that the results are mixed. In all cases except C_{TURN}, all approaches (bucketed and non-bucketed) outperform the variable two-dimensional approach. This is probably due the overheads of creating and destroying lists as discussed previously. The fixed length bucketing algorithms on the other hand perform better than their variable length counterparts, especially in terms of the mean time taken for C_{DWELL} to find a route. A fixed bucket size of five produces the lowest maximum time (only one sixth of that taken by the unbucketed code) to find a route using the C_{DWELL} algorithm, but performs slightly worse than an increased bucket size of ten in terms of overall mean times (although it takes up to twice as long to select a route for C_{DWELL}). This suggests that the fixed length ten bucket is no longer affected substantially by the time to taken to search over the list, is more prone to performance issues generated by creating and destroying lists, but in general it offers the best two-dimensional bucketing option for reducing processing times.
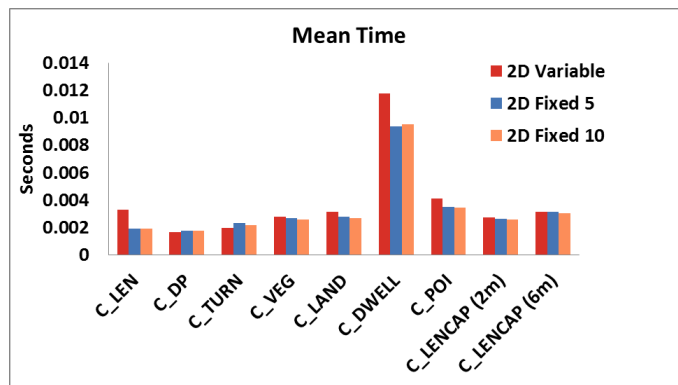
(a)
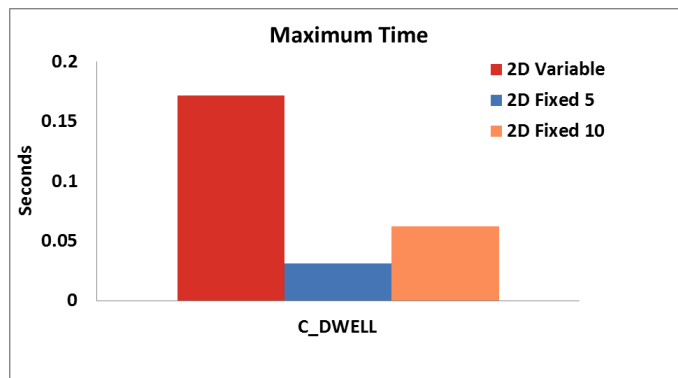
(b)

Figure 5.22: (a) Mean and (b) maximum running times for the three 2D bucketed algorithms to select 870 routes.

5.4.4 Summary

Figure 5.23 shows a comparison of the unbucketed, one-dimensional bucket and fixed length ten two-dimensional bucket approaches. Overall the one-dimensional bucket approach produces the best results, other than the maximum time to find a route with high dwelling values. The algorithms required to generate the simplest and most attractive routes, along with those to select routes for everyday, leisure and tourist route types, are combinations of attributes rather than single attribute cost functions. It is therefore unlikely that the partial routes generated will have the same characteristics as the C_{DWELL} algorithm. For this reason the one-dimensional bucket approach will be taken to find multi-attribute cost functions in the following chapter, although the three approaches shown in Figure 5.23 will again be tested for the resulting algorithms.
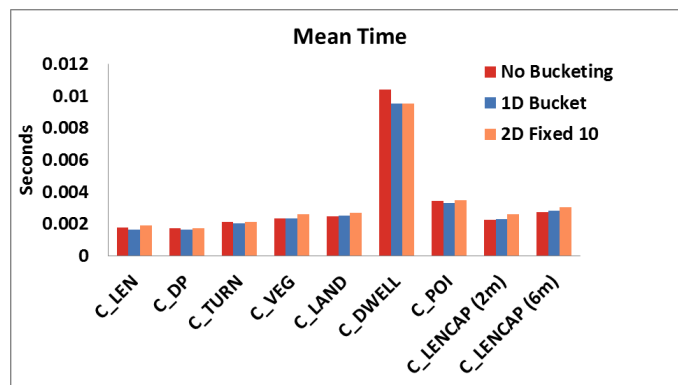
(a)
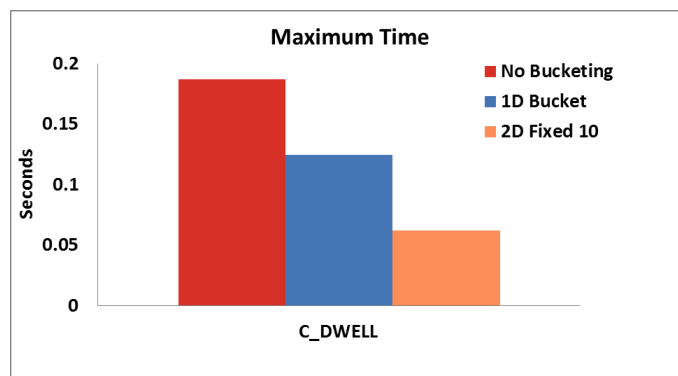
(b)

Figure 5.23: (a) Mean and (b) maximum running times for the unoptimised and 1D bucketed and 2D fixed size 10 bucketed algorithms to select 870 routes.

Chapter 6

Determining the Multi-Attribute Cost Functions

The cost functions have now been determined for each of the individual attributes, but Chapter 3 indicates that several attributes are influential for each of the attribute categories and journey types being considered by this research. For multi-attribute minimising algorithms there are many possible approaches to calculating cost, but a straightforward approach is to represent the cost (C) with a weighted cost function combining each of the individual attribute values (A,B,C,D...):

$$C = w_1 A + w_2 B + w_3 C + w_4 D + \cdots \tag{6.1}$$

Although a linear approach to combining the attribute costs is not the only possibility, it is the method used by four of the existing systems discussed in Section 2.2 [31, 33, 187, 228]. It also fits with the present research's aim to produce a simple algorithm.

Combining these single attribute costs into a cost function for each of the categories and journey types involves determining the 'best' weights for each cost as well as the number of attributes to include. For simplicity and attractiveness all attributes found to be influential (see Table 3.2) will be incorporated into the relevant cost functions, but for everyday, leisure and tourist journeys the number of attributes is large. However, as it is not clear how many of these attributes are required to select suitable routes, combinations produced according to the rankings shown in Table 6.1 will all be compared in order to

select those with the best performance.

Simple	Attractive		Everyday	Leisure	Tourist
Length[1]	Vegetation[2]		Length[1]	Vegetation[2]	POIs[2]
Turns[1]	POIs[2] =		Vegetation[2]	POIs[2]	Dwellings[2] =
DPs[1]	Land Use[2]		POIs[2] =	Land Use[2] =	Vegetation[2]
	Dwellings[2]		Turns[1]	Dwellings[2] =	Land Use[2]
			Dwellings[2] =	Length[2]	Length[2]
			DPs[1] =	DPs[1] =	
			Land use[2]	Turns[1]	

Table 6.1: Importance rankings for route attributes showing simplicity, attractiveness and each of the different journey types. [1] Indicates a negative relationship, [2] indicates a positive relationship. (POIs - points of interest, DPs - decision points)

This chapter will discuss the approach used to find the 'best' weights, investigate the performance of the resulting cost functions and determine which of the suggested cost functions produce the most appropriate routes. Initially focusing on the simplicity, it will empirically derive a method for evaluating weight combinations and use these weights find the most appropriate algorithm. Once this method has been derived, it is then applied to the attractiveness category followed by the everyday, leisure and tourist journeys.

6.1 The 'Simplest' Routes

The experiments conducted in Section 3.4 and previous research [98, 182, amongst many others] have shown that length, turns and decision points all play a part in how simple a route is perceived to be. Therefore an algorithm that suggests simple routes through an environment must consider all three of these attributes. A straightforward approach of combining the effects of these attributes is to construct a simple weighted summation equation for calculating route cost as shown in Equation (6.2):

$$C_{SIMP} = w_1 C_{LEN} + w_2 C_{DP} + w_3 C_{TURN} \tag{6.2}$$

Here C_{SIMP} is route cost, C_{LEN} is the length cost (Equation 5.1), C_{DP} is the decision point cost (Equation 5.2) and C_{TURN} is the turns cost (Equation 5.3) of the partial route, and W_1, W_2 and W_3 are their corresponding weights. One approach is to apply equal weighting to each of the length, turns and decision points, i.e. $w_1 = w_2 = w_3 = 1.0$ giving Equation 6.3.

$$C1_{SIMP} = C_{LEN} + C_{DP} + C_{TURN} \tag{6.3}$$

Figure 6.1 shows the metrics which result from creating routes with the same 174 start-end pairs as used previously. Figures 6.1a, 6.1b and 6.1c indicate that although higher than the means found by each of the respective single attribute algorithms, the simplicity algorithm selects routes which are reasonable for each of the tested attributes. However, Figure 6.1d shows that the route similarity for this algorithm is generally lower than C_{LEN}, C_{DP} and C_{TURN} (See Figure 5.5, Section 5.3.1).

Figure 6.1: Results for the routes generated by the equally weighted simplicity cost ($C1_{SIMP}$) algorithm (1st iteration). Shown are mean route (a) length, (b) DPs and (c) turns, and (d) route similarity. For (d) boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range.

Section 3.5 indicated that not all of attributes carry the same amount of influence when selecting a route. So how should we determine the 'best' values of these weights? We know that the weight values will be in the range of 0 and 1, although they are assumed to not equal 0, which gives us a starting range for evaluation. By selecting an increment size, it is possible to test each of the weight combinations automatically and then analyse

the results. However we must first find a way of defining a 'best' routes and, therefore, the corresponding weights.

6.1.1 Selecting Evaluation Metrics

Assuming that the metrics discussed in Section 5.2 can be used to define a 'good' route, we need to examine how they vary as different weights are applied to the cost function. Using the cost equation (Equation (6.2)), w_1, w_2 and w_3 were varied from 0.0 to 1.0 with increment steps of 0.25 across the same 174 routes as used previously. This produced a series of route sets, one for each weight combination.

In our search for the 'best' simplicity weights, we are aiming to remove any weight combinations which produce excessively long routes, or those with many turns or decision points. Figure 6.2 shows the percentage increase in route length, turns and decision points as the weight combinations vary between 0.0 and 1.0. Measures such as the mean increase were considered, but increase percentiles of over 80% were shown to provide more variation and therefore a better indication of performance. To avoid the divide by zero issue for turns, any entry with zero turns in C_{TURN} and more than zero turns in C_{SIMP} was assigned a value equal to the maximum increase for all weight combinations plus 100% (in this case 500%). To ensure that an attribute is influential when choosing weights, its metric's value should vary as the weight values change. For length (Figure 6.2a), decision points (Figure 6.2b) and turns (Figure 6.2c), percentiles from 80% to 100% all produce variation across the weight ranges, meaning that they all fulfill this requirement. In addition, we want to penalise high attribute value routes but without being overly affected by outliers. For this reason, the 95th percentile was selected as the most appropriate measure. These 95th percentile values were then standardised using Z-score scaling ($(x - x_{mean})/x_{std}$), giving the values shown in Figure 6.3a.

Summing these values into a total attribute score, as shown in Figure 6.3b, produces a high score value for routes which have excessive length, decision points, turns or a combination of these, and removal of these high scoring weights will therefore leave only better performing candidates. For the first stage of route evaluation, the top 10% of attribute score values is calculated (score threshold in Figure 6.3b), and is then used to reject weight combinations with attribute scores above this threshold.
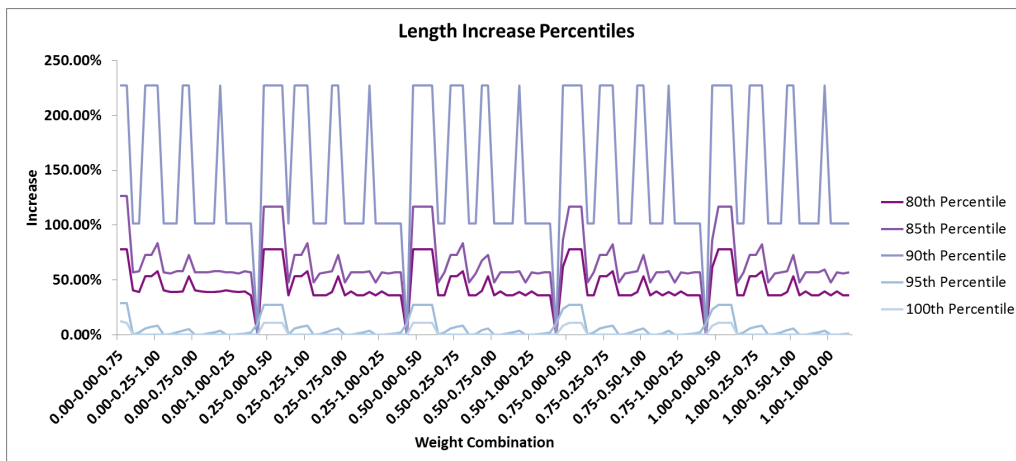
(a)
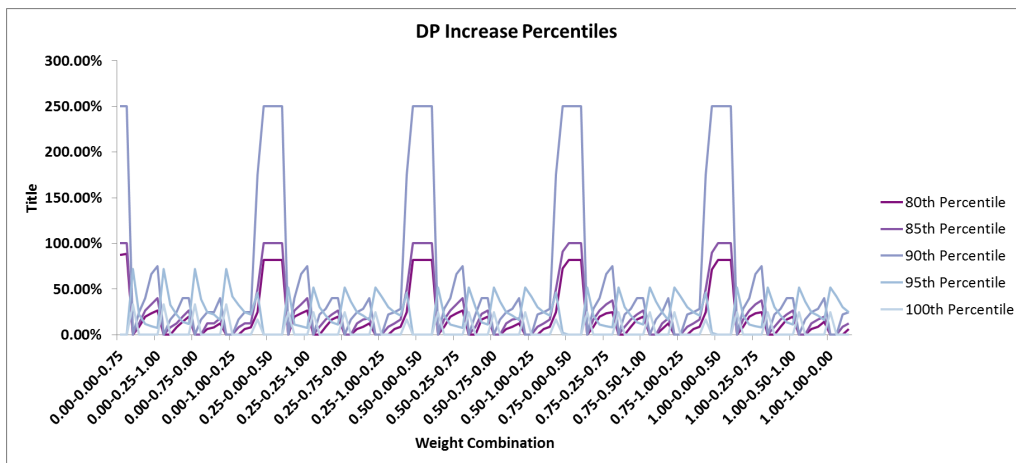
(b)
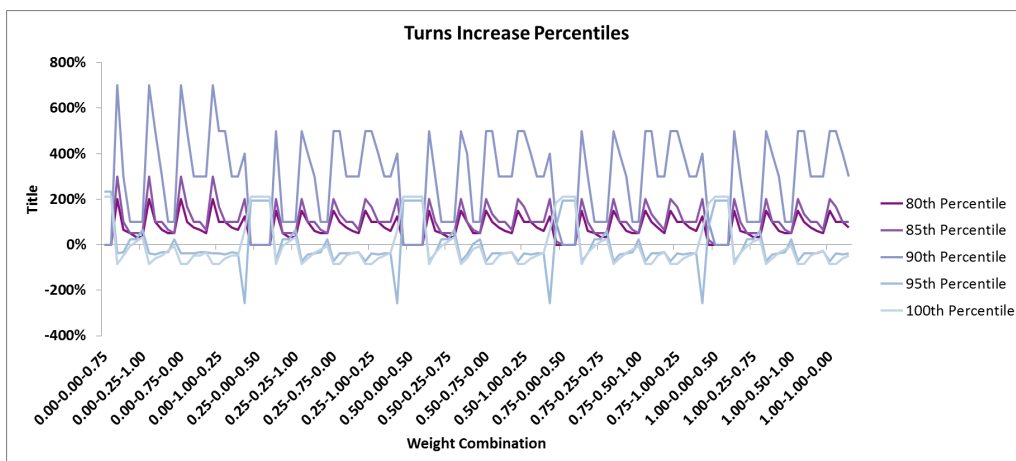
(c)

Figure 6.2: Attribute increase percentiles (the increase of route attribute cost compared to the minimum) of (a) length, (b) DPs and (c) turns as weight combinations vary (1st iteration). The weights are shown as $w_1 - w_2 - w_3$ and are sorted by w_3 then w_2 then w_1.

In addition, Section 5.2 has discussed the need for similar routes, suggesting a second stage of evaluation based on the similarity metric. This metric needs to enable weights producing routes with very low similarity to be removed, and therefore requires variation across the tested weight combinations. The route similarities generated by the single attribute cost functions typically have have a minimum and maximum of 0% and 100% respectively (see Figure 5.5 and Figure 5.15) ruling them out as an evaluation metric. However, the 25th percentile similarity shows suitable variation across the zones and
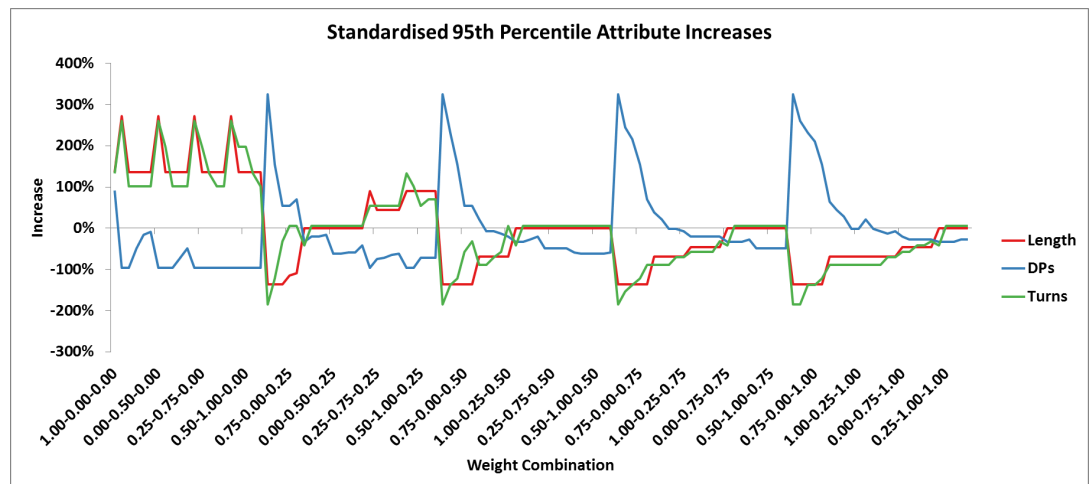
(a)
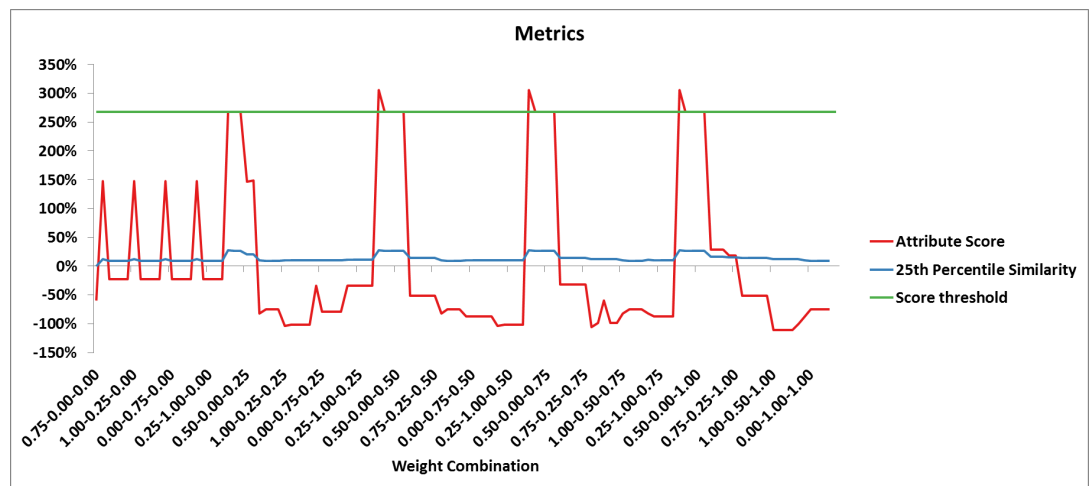
(b)

Figure 6.3: (a) Standardised attribute increases and (b) evaluation metrics (total attribute score, 25th percentile similarity and score threshold) for C_{SIMP} as weight combinations vary (1st iteration). The weights are shown as $w_1 - w_2 - w_3$ and are sorted by w_3 then w_2 then w_1 to show the patterns in the data.

attributes, and across differing weight combinations (Figure 6.3b). This percentile was therefore chosen to avoid combinations producing routes with particularly low similarity.

Figure 6.3b also shows how the attribute score and 25th percentile similarity compare, and how using these two metrics it should be possible to focus in on the best weights. It is important to note that more than one weight combination may have the same 25th percentile similarity, either due to the routes produced being identical, or where the similarity is 0% or 100%. Furthermore, where several weight combinations produce the same routes, the attribute score will also be identical. This effect will produce a range of weights, all equally good, and selecting any one of these weight combinations would produce the most appropriate routes.

6.1.2   Finding the 'Best' Weights

To determine the 'best' weights the algorithm was run several times, adjusting the range and step size each iteration according to the routes generated by the previous one. Initially the range 0.0 - 1.0 was selected, and 5 steps giving a step size of 0.25. At the end of the first iteration, the metrics for each weight combination were generated and compared using a four stage process:

1. Any weight combinations with attribute scores in the top 10% are rejected.

2. Weight combinations with all but the highest 25th percentile similarity are rejected.

3. If the 25th percentile similarities are equal [1], then of the weights remaining all but those with the minimum attribute score are rejected.

4. If a range of weights with equal similarity and attribute score remain, the set of weights closest to the mean within that range is chosen as the best. This is designed to make the algorithm robust to small variations in each weight.

This process produces a weight combination for w_1, w_2 and w_3, deemed to be the 'best' for that iteration. For the next iteration the 'best' weight combination is added to the list of weights, along with a new series of weights generated in the following way:

- If there are a range of equally good weight combinations, then these are used as the next range of weights.

- If there is a single 'best' weight combination, then these weights are used as the centre point for the next iteration's range. To increase the resolution of the search, the width of the range is halved from the original and capped at 0.0 and 1.0.

---

[1] see Section 6.1.1 for an explanation of equal similarities

The new range is then divided by the original number of steps to determine the increment required, the new series of weights are added to the weights list, and the algorithm is run again. This process is repeated until a stopping criterion is reached. In this case the stopping criterion was set to be a minimum of three iterations, and thereafter when three identical results had been returned.
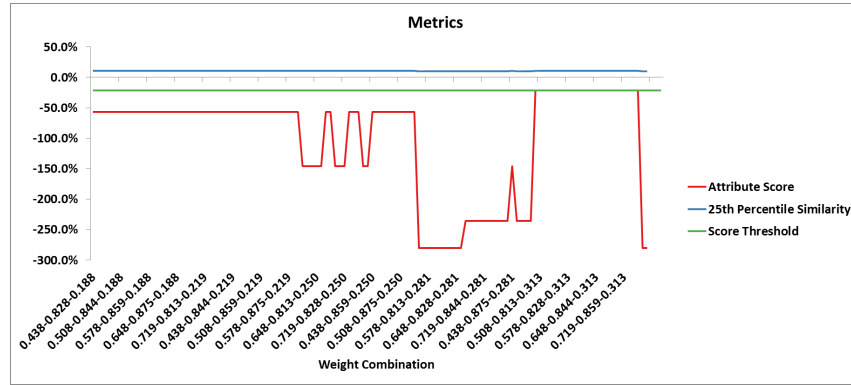
Figure 6.4: Evaluation metrics (total attribute score, 25th percentile similarity and score threshold) for C_{SIMP} as weights vary (4th iteration). Weights are sorted by w_3 then w_2 to show the pattern in the data.

For he simplicity journey type the process took four iterations, with the metrics for the fourth iteration being much more stable over the different weight combinations than for the first iteration, as shown in Figure 6.4. This produces a wide range of weight combinations which would produce equally good routes. In fact, there are eight weight combinations which could all be considered the 'best' (having identical 25th percentile similarity and attribute scores), with w_1 ranging from 0.4375 to 0.71875, w_2 from 0.8125 to 0.84375 and w_3 equal to 0.25. Selecting the weight combination closest to the mean value for each weight range produces the best cost function shown in Equation 6.4.

$$C_{SIMP} = 0.578125 * C_{LEN} + 0.828125 * C_{DP} + 0.25 * C_{TURN} \qquad (6.4)$$

6.1.3 The Performance of the 'Best' Weight Combination

To determine the performance of the 'best' weights, the resulting route attribute values and similarity (Figure 6.5) were compared to those of the equally weighted cost functions (Figure 6.1). The most substantial difference is in the route similarity for the best weighted cost function (C_{SIMP}) compared to that of the equally weighted cost function ($C1_{SIMP}$) as shown in Figure 6.1d and Figure 6.5d. Zones 3, 4 and 5 show increases in mean similarity for C_{SIMP} over $C1_{SIMP}$ (with Zone 2 having equal values), and increases in the 25the
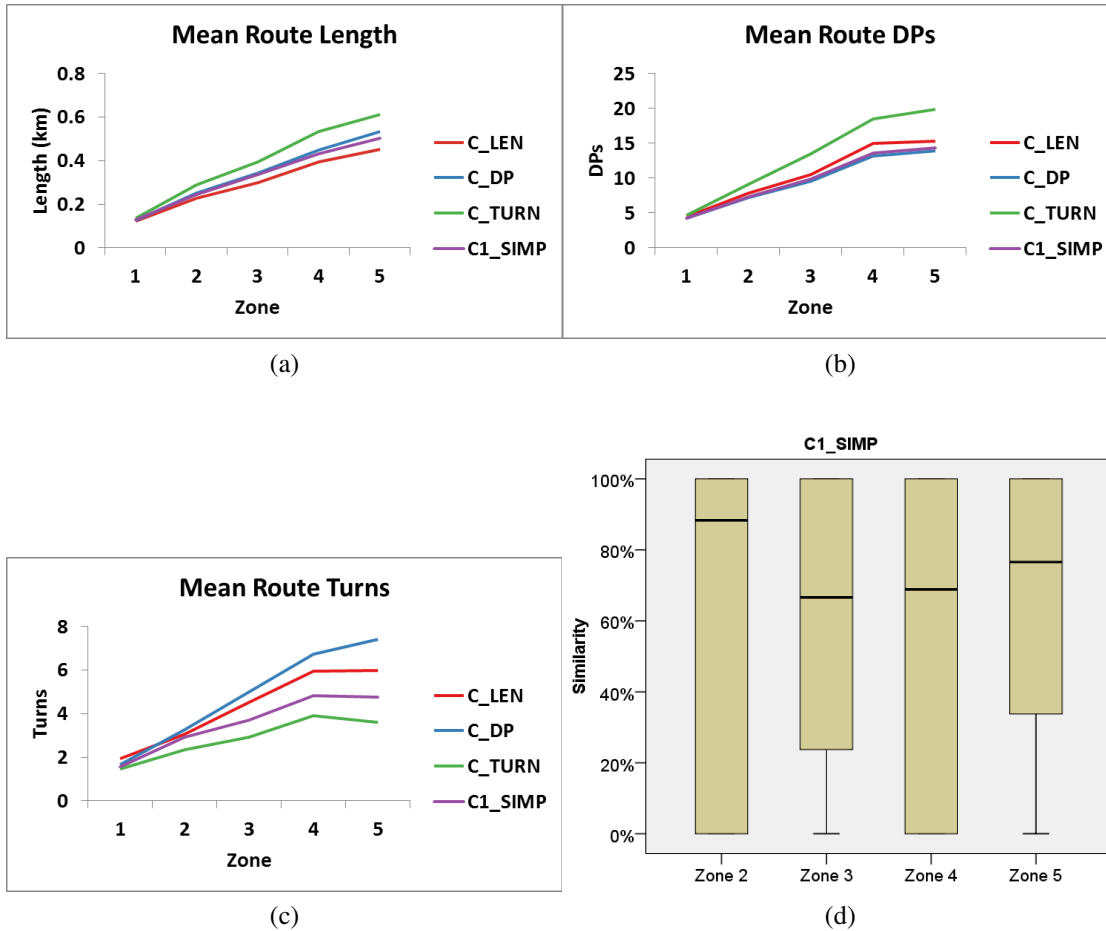
percentile similarity for Zone 4 and Zone 5.
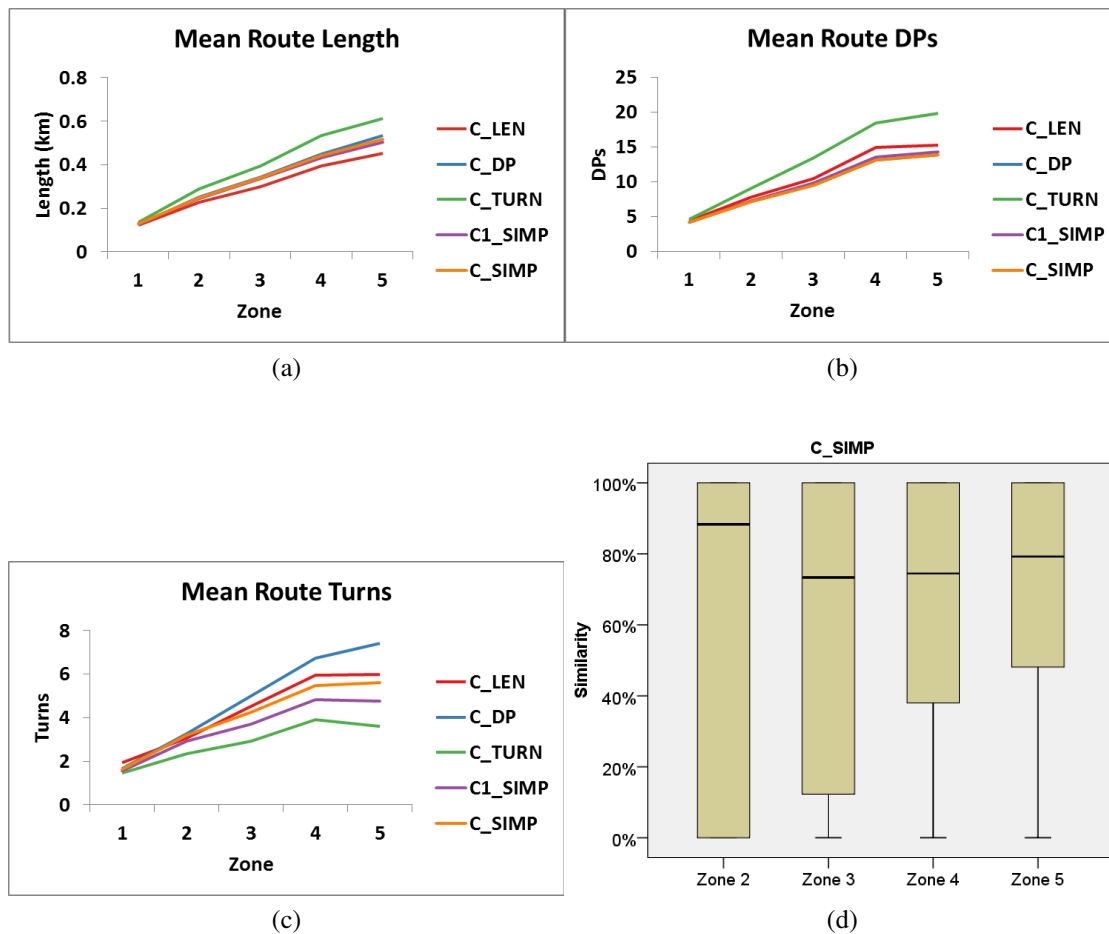
Figure 6.5: Results for the routes generated by the best weighted simplicity cost (C_{SIMP}) algorithm (4th iteration). Shown are mean route (a) length, (b) DPs and (c) turns, and (d) route similarity. For (d) boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range.

Figure 6.5a shows the mean route length of the best weighted cost function (C_{SIMP}) with $C1_{SIMP}$, with both being only slightly worse than that of the minimum cost function (C_{LEN}). For decision points (Figure 6.5b) C_{SIMP} performs better than C_{LEN} and slightly better than $C1_{SIMP}$, and being identical to C_{DP} for all but Zone 5. In contrast, more turns are encountered by C_{SIMP} than $C1_{SIMP}$, but it still outperforms C_{LEN} (Figure 6.5c). The results for route similarities and those for the attribute metrics indicate that the 'best' weight combination (C_{SIMP}) produces routes which are more appropriate than those found by the equally weighted simplicity cost function ($C1_{SIMP}$). The routes also offer a reasonable trade-off between length, number of decision points and number of turns, and the majority have a high similarity between routes from the same start point to end points that are

in approximately the same direction. Figure 6.6 shows examples of the routes generated by the best weighted simplicity algorithm for a single start point.
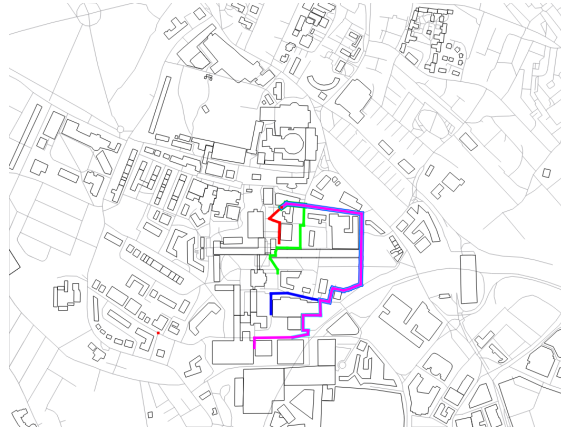
Figure 6.6: Example Routes generated by the best weighted simplicity cost algorithm for a single start point and sector. The different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

6.2 The 'Most Attractive' Route

Vegetation, land use, dwellings and points of interest all affect the attractiveness of a route (see Table 5.1), suggesting that a route for this category should consider all of these attributes. Figure 5.13 indicates that increasing one attribute decreases another, so as with simplicity, a good trade-off between the attributes must be found by combining them into a single cost function. As before, the straightforward approach of combining the effects of these attributes is to construct a simple weighted sum equation for calculating route cost as shown in Equation (6.5):

$$C_{ATTRACT} = w_1 C_{VEG} + w_2 C_{LAND} + w_3 C_{DWEL} + w_4 C_{POI} \qquad (6.5)$$

To confirm that the evaluation metrics can be calculated in the same way as simplicity, the attractiveness algorithm was run for one iteration with weights varying between 0.0 and 1.0, and in increment steps of 0.25. The attribute costs (C_{VEG}, C_{LAND}, C_{DWEL} and C_{POI}) were used to calculate the cost increase percentiles, from the same 174 start-end pairs as used previously. These were then examined, and the 95th percentile was found to be a sufficient metric for evaluation purposes.

These increases were combined as described in Section 6.1.1 to create the attribute metrics shown in Figure 6.7. To determine the 'best' weights the algorithm was again run
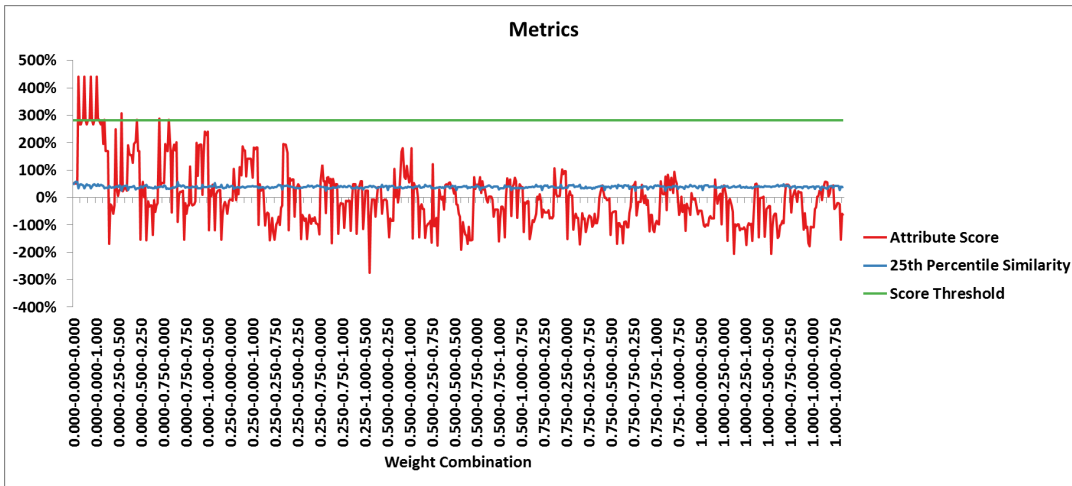
Figure 6.7: (Evaluation metrics (attribute score, 25th percentile similarity and score threshold) for $C1_{ATTRACT}$ as weight combinations vary (1st iteration)

to convergence, adjusting the range and step size each iteration according to the routes generated by the previous one, with the same four step process used to determine the best weights (Section 6.1.2):

1. Any weight combinations with attribute scores in the top 10% are rejected.

2. Weight combinations with all but the highest 25th percentile similarity are rejected.

3. If the similarity is equal, then of the weights remaining all but those with the minimum attribute score are rejected.

4. If a range of weights with equal similarity and attribute score remain, the weight combination closest to the mean within that range is chosen as the best.

The attractiveness weights converge after eight iterations and, as with simplicity, the vegetation, land use, dwellings and POI cost increases stabilise for the 8th iteration. The best weights were found to be $w_1 = 0.814453$, $w_2 = 0.53418$, $w_3 = 0.935547$ and $w_4 = 0.312012$, giving the cost function shown in equation 6.6.

$$C_{ATTRACT} = 0.814453 * C_{VEG} + 0.53418 * C_{LAND} + 0.935547 * C_{DWEL} + 0.312012 * C_{POI}$$
$$(6.6)$$

For comparison purposes, the algorithm was also run with the equally weighted cost function given in Equation 6.7, and the results are shown in Figure 6.8 and Figure 6.9.

$$C1_{ATTRACT} = C_{VEG} + C_{LAND} + C_{DWEL} + C_{POI} \qquad (6.7)$$

Comparing Figure 6.8a to Figure 6.8b, indicates that the route similarity for $C1_{ATTRACT}$ is generally higher that $C_{ATTRACT}$, but this is only one evaluation metric.
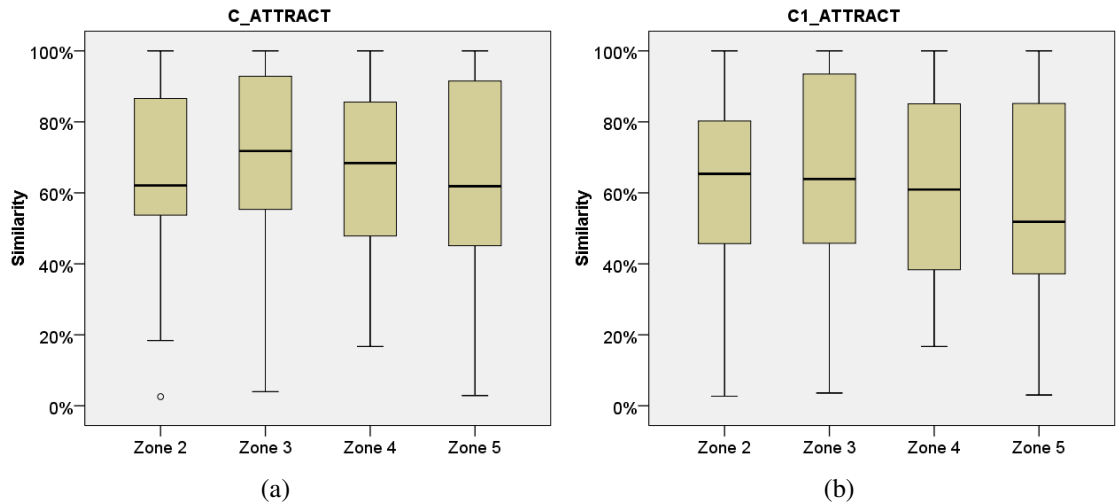
Figure 6.8: Similarities for the (a) best weighted attractiveness ($C_{ATTRACT}$) and (b) equally weighted attractiveness ($C1_{ATTRACT}$) algorithms. Boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range.

Figure 6.9 shows the performance of $C_{ATTRACT}$ and $C1_{ATTRACT}$, compared to that of each of the single attribute algorithms. For vegetation (Figure 6.9a) and points of interest (Figure 6.9d) $C_{ATTRACT}$ has comparable or higher means than $C1_{ATTRACT}$ for the majority of zones; however, $C_{ATTRACT}$ performs slightly worse for parkland (Figure 6.9b) and slightly better for dwellings (Figure 6.9c). Overall, $C_{ATTRACT}$ provides a good trade-off between the attractiveness attributes, and good similarity for the routes produced.
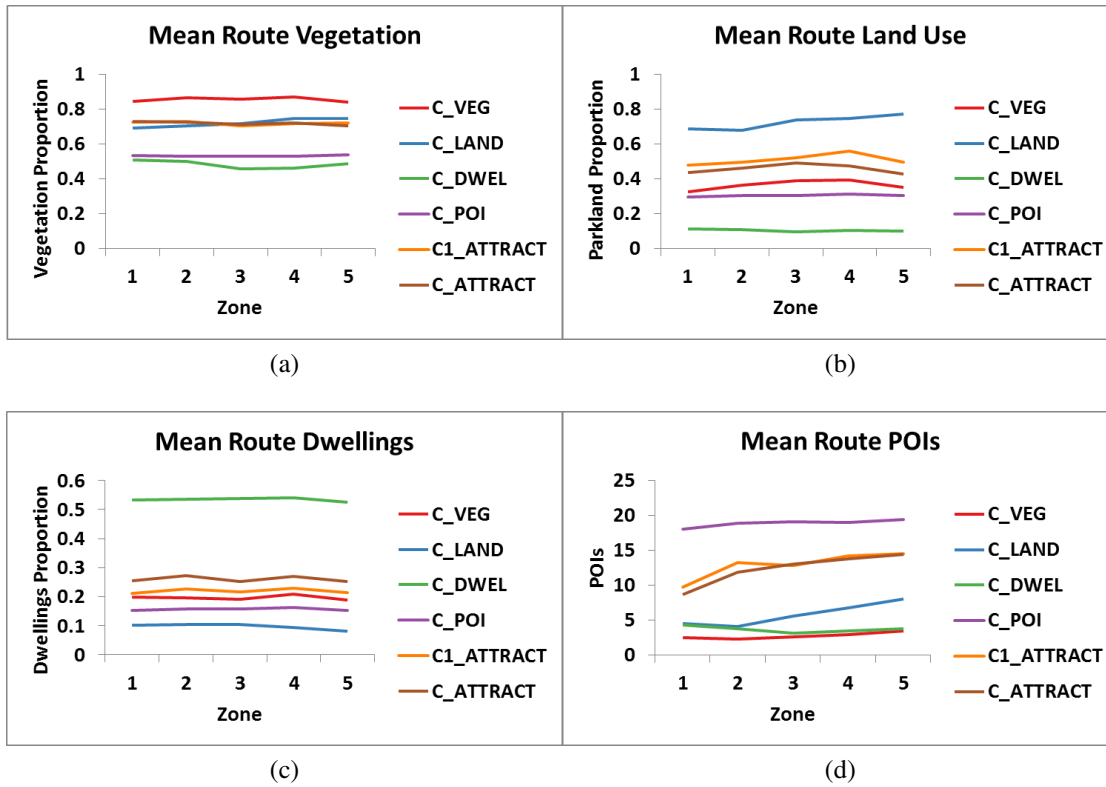
Figure 6.9: Attribute Means - showing the mean route (a) vegetation, (b) parkland, (c) dwellings and (d) POIs for the routes generated by the vegetation (C_{VEG}), land use (C_{LAND}), dwellings (C_{DWEL}), POIs (C_{POI}), and best ($C_{ATTRACT}$) and equally ($C1_{ATTRACT}$) weighted attractiveness cost algorithms.

An example of the routes produced by the best weighted algorithm is shown in Figure 6.10a. As expected, these routes avoid the lower right quadrant of the environment where possible, which was an area shown to have low values of the attractiveness attributes in Section 4.2. Another observation that can be made from Figure 6.10a is that some of the routes selected appear to be very long. Figure 6.10b confirms this observation. The minimum length route produced by this algorithm for each zone is acceptable at 0.1km to 1.1km, but with a maximum of 7.5km, some of these routes are unlikely to be suitable for pedestrian travel. This is not surprising as length is not included in the evaluation attributes for this category. However, as the attractiveness attributes are ranked high for leisure and tourist journeys, the length of the routes selected may have to be considered when selecting cost functions for these journey types.
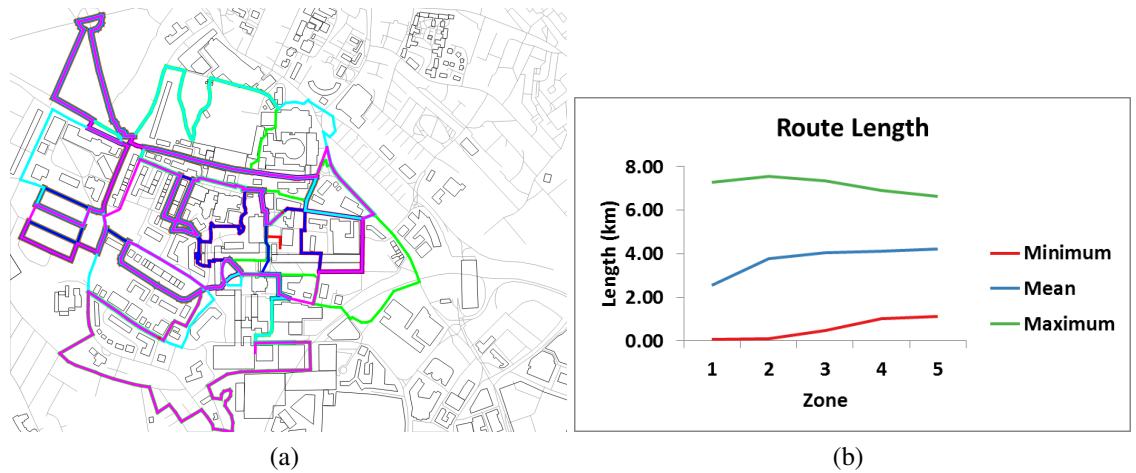
Figure 6.10: (a) Example routes for a single start point and sector, and (b) route lengths for best weighted attractiveness ($C_{ATTRACT}$) algorithm. The different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

6.3 The Everyday Journey Type

The experiments presented in Chapter 3 indicates that all seven tested attributes (length, decision points, turns, vegetation, land use, dwellings and points of interest) influence the choice of routes for everyday journeys as shown in column three of Table 5.1. The number of attributes which should actually be included in these tests is however somewhat questionable. Dwellings, decision points and land use are all ranked equal last by Experiment 4 (Section 3.7 and Experiment 6 (Section 3.9), indicating that they have less effect on route choice for everyday journeys, but should that influence be completely excluded? In order to determine this, a suitable series of cost functions need to be created and their performance for each of the metrics tested. The most straightforward approach to finding the most appropriate cost function for everyday routes is to add a single attribute at a time to the everyday cost function, and evaluate the performance of each algorithm produced. This allows each attribute combination to be evaluated, and leads to the twelve cost functions shown in Equation 6.8 to Equation 6.19.

$$C1_{EVER} = C_{LEN} \tag{6.8}$$

$$C2_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} \tag{6.9}$$

$$C3_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} \tag{6.10}$$

$$C4_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{TURN} \tag{6.11}$$

$$C5_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} + w_4 * C_{TURN} \tag{6.12}$$

$$C6_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} + w_4 * C_{TURN} + w_5 * C_{DWEL} \tag{6.13}$$

$$C7_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} + w_4 * C_{TURN} + w_5 * C_{DP} \tag{6.14}$$

$$C8_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} + w_4 * C_{TURN} + w_5 * C_{LAND} \tag{6.15}$$

$$C9_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} + w_4 * C_{TURN} + w_5 * C_{DWEL} \\ + w_6 * C_{DP} \tag{6.16}$$

$$C10_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} + w_4 C_{TURN} + w_5 * C_{DWEL} \\ + w_6 * C_{LAND} \tag{6.17}$$

$$C11_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} + w_4 * C_{TURN} + w_5 * C_{DP} \\ + w_6 * C_{LAND} \tag{6.18}$$

$$C12_{EVER} = w_1 * C_{LEN} + w_2 * C_{VEG} + w_3 * C_{POI} + w_4 * C_{TURN} + w_5 * C_{DWEL} \\ + w_6 * C_{DP} + w_7 * C_{LAND} \tag{6.19}$$

The first stage in determining which one of these cost functions selects the most appropriate routes for everyday travel is to find the best weights for each attribute combination with the exception of $C1_{EVER}$, which has only a single attribute and therefore no weights. Running the same evaluation process and 174 points as described in Section 6.1.1 pro-

duces a set of weights for each cost function as shown in Equation 6.20 to Equation 6.30.

$$C2_{EVER} = 0.5 * C_{LEN} + 0.757813 * C_{VEG} \tag{6.20}$$

$$C3_{EVER} = 0.245117 * C_{LEN} + 0.866943 * C_{VEG} + 0.624634 * C_{POI} \tag{6.21}$$

$$C4_{EVER} = 0.839844 * C_{LEN} + 0.804688 * C_{VEG} + 0.595703 * C_{TURN} \tag{6.22}$$

$$C5_{EVER} = 0.625 * C_{LEN} + 0.3125 * C_{VEG} + 0.96875 * C_{POI} + 0.3125 * C_{TURN} \tag{6.23}$$

$$C6_{EVER} = 0.810547 * C_{LEN} + 0.643555 * C_{VEG} + 0.949219 * C_{POI} + \\ 0.394531 * C_{TURN} + 0.928955 * C_{DWEL} \tag{6.24}$$

$$C7_{EVER} = 0.578125 * C_{LEN} + 0.941406 * C_{VEG} + 0.742188 * C_{POI} + \\ 0.148438 * C_{TURN} + 0.726563 * C_{DP} \tag{6.25}$$

$$C8_{EVER} = 0.585938 * C_{LEN} + 0.710938 * C_{VEG} + 0.9375 * C_{POI} + \\ 0.628906 * C_{TURN} + 0.132813 * C_{LAND} \tag{6.26}$$

$$C9_{EVER} = 0.265625 * C_{LEN} + 0.640625 * C_{VEG} + 0.958984 * C_{POI} + \\ 0.226563 * C_{TURN} + 0.757813 * C_{DWEL} + 0.958984 * C_{DP} \tag{6.27}$$

$$C10_{EVER} = 0.828857 * C_{LEN} + 0.983696 * C_{VEG} + 0.889526 * C_{POI} + \\ 0.5 * C_{TURN} + 0.92675 * C_{DWEL} + 0.343292 * C_{LAND} \tag{6.28}$$

$$C11_{EVER} = 0.757813 * C_{LEN} + 0.273438 * C_{VEG} + 0.882813 * C_{POI} +$$
$$0.179688 * C_{TURN} + 0.882813 * C_{DP} + 0.648438 * C_{LAND}$$

(6.29)

$$C12_{EVER} = 0.878906 * C_{LEN} + 0.261719 * C_{VEG} + 0.507813 * C_{POI} +$$
$$0.570313 * C_{TURN} + 0.507813 * C_{DWEL} + 0.507813 * C_{DP} + 0.507813 * C_{LAND}$$

(6.30)
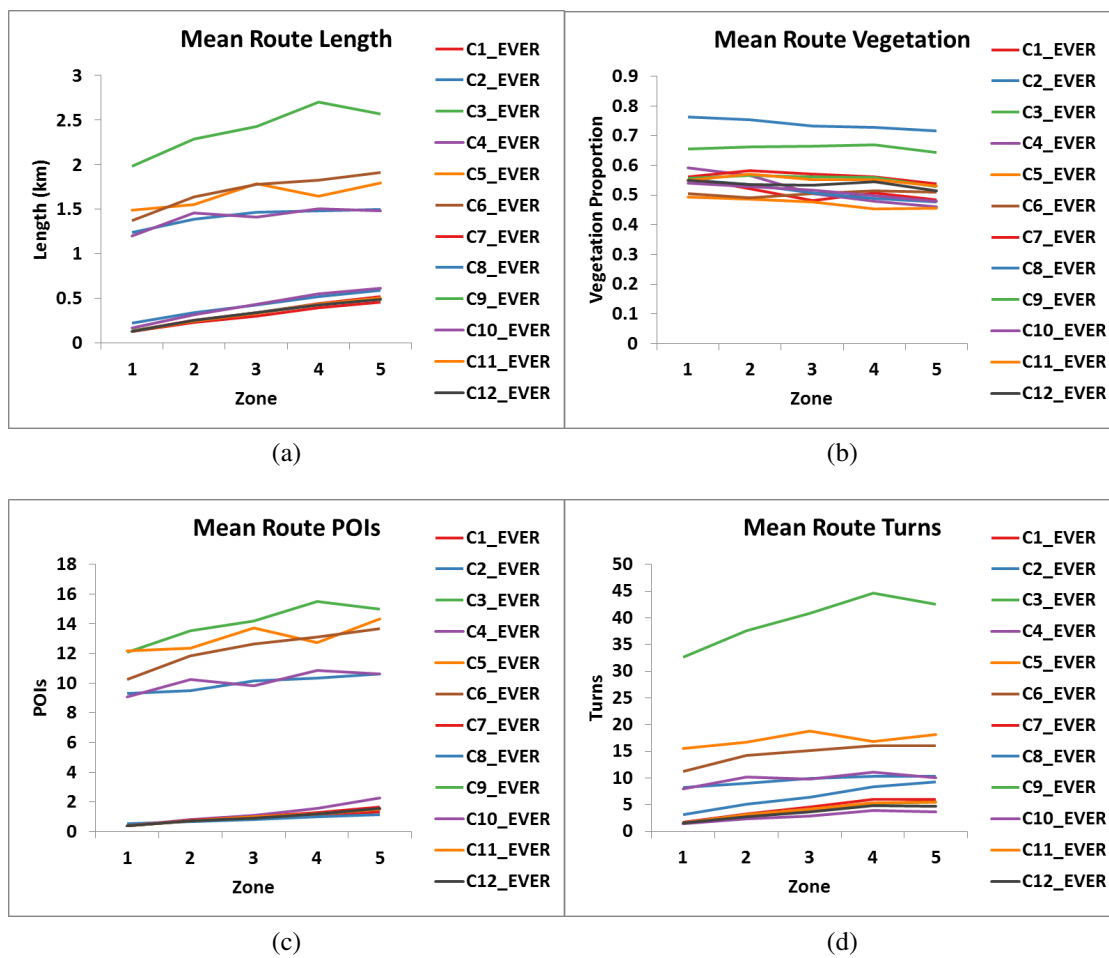
Figure 6.11: Evaluation metrics - showing the mean route (a) length, (b) vegetation, (c) POIs and (d) turns for the routes generated by the everyday cost algorithms.
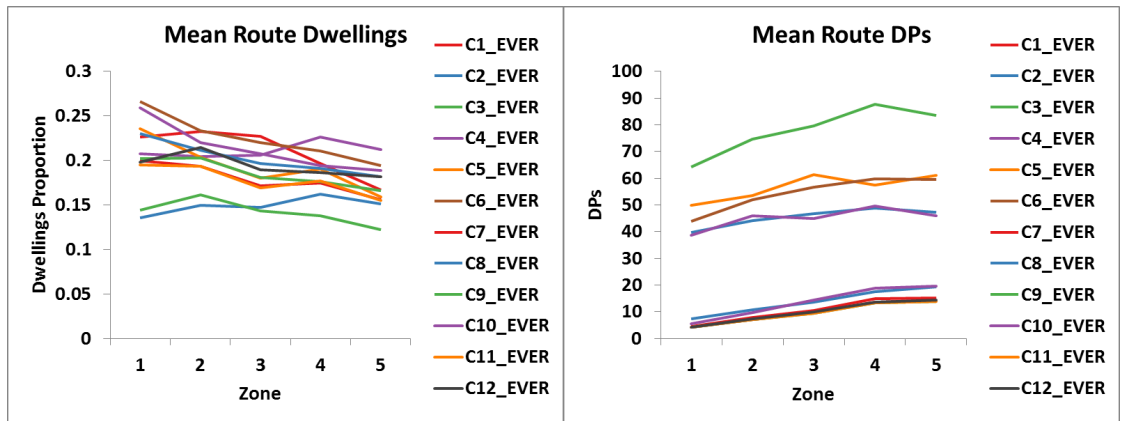
(a)

(b)
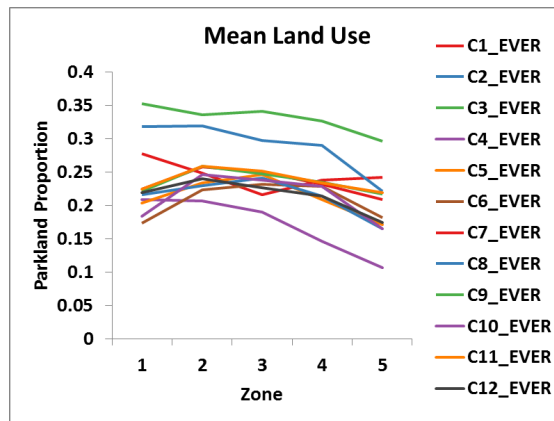
(c)
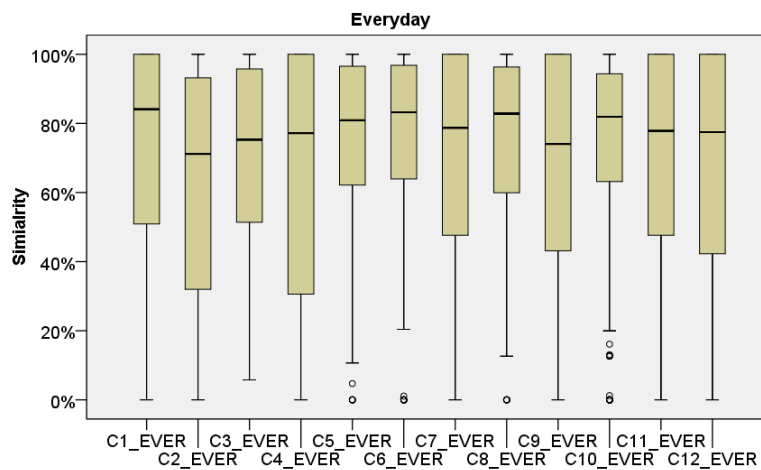
(d)

Figure 6.12: Evaluation metrics - showing the mean route (a) dwellings, (b) DPs, (c) land use and (d) similarity for the routes generated by the everyday cost algorithms. For (d) boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range.

The next stage of the investigation is to examine the performance of the algorithms for each cost function. Figure 6.11 and Figure 6.12(a-d) show the attribute means and Figure 6.12d shows the route similarities for $C1_{EVER}$, to $C8_{EVER}$. Although these give some indication of the cost functions performance, they do not clearly indicate which is likely to be the best.

Cost	Difference							Simil-
Function	Length	Veg	POIs	Turns	Dwel	DPs	Land	arity
$C1_{EVER}$	0.0%	40.1%	95.6%	49.8%	60.7%	10.4%	66.3%	50.9%
$C2_{EVER}$	41.2%	13.6%	95.9%	125.0%	72.1%	41.9%	60.1%	32.0%
$C3_{EVER}$	737.4%	23.1%	27.2%	1330.7%	73.5%	710.0%	54.4%	51.2%
$C4_{EVER}$	38.4%	38.5%	94.0%	0.2%	60.5%	41.8%	76.4%	30.5%
$C5_{EVER}$	480.9%	44.5%	32.1%	526.7%	63.8%	488.7%	70.7%	62.1%
$C6_{EVER}$	495.8%	41.3%	36.7%	424.8%	58.0%	464.9%	71.3%	63.9%
$C7_{EVER}$	13.1%	34.2%	95.0%	37.4%	66.5%	2.4%	67.7%	47.5%
$C8_{EVER}$	397.4%	39.8%	48.4%	246.4%	62.2%	371.4%	70.6%	59.6%
$C9_{EVER}$	10.3%	35.2%	95.4%	36.6%	65.3%	1.4%	67.5%	43.1%
$C10_{EVER}$	396.2%	39.8%	47.6%	256.3%	60.0%	374.4%	70.7%	63.1%
$C11_{EVER}$	9.3%	35.5%	95.3%	37.4%	66.7%	1.3%	67.2%	47.5%
$C12_{EVER}$	8.5%	37.5%	95.3%	21.6%	63.7%	4.7%	70.4%	41.4%

Table 6.2: Attribute and 25th similarity overall evaluation metrics for each of the everyday cost functions. Attributes are given as the percentage difference compared to the respective single attribute cost functions (percentage increase for length, turns and DPs, percentage decrease for vegetation, POIs, dwellings and land use). Grey boxes indicate the best values for each metric, and black the worst.

Table 6.2 shows the evaluation metrics for each of the tested attributes, combining zones to give an overall value. It indicates that although a cost function may perform well for a single attribute, it does not necessarily perform well for all. From these values, a three stage process was used to determine the best performing cost function.

1. **Cost functions which produce excessively long routes are discounted.** In this case, although all routes are produced are within acceptable length for pedestrian travel, $C3_{EVER}$, $C5_{EVER}$, $C6_{EVER}$, $C8_{EVER}$ and $C10_{EVER}$ clearly produce routes which are far longer than the other remaining cost functions (with an increase of 396% to 737%). These five cost function were therefore removed from the candidate pool.

2. **The worst performing cost function in terms of each attribute was identified and discounted.** From Table 6.2 this suggests that $C2_{EVER}$ and $C4_{EVER}$ should also be removed from the pool of possible candidates.

3. **The overall performance of the remaining cost functions was compared, and the best performing was selected.** The performance of the five remaining candidate cost functions ($C1_{EVER}$, $C7_{EVER}$, $C9_{EVER}$, $C11_{EVER}$ and $C12_{EVER}$) is shown in Table 6.3. Of these, $C1_{EVER}$ outperforms the others in three of the seven attributes, and has the best similarity. In addition, if the attribute differences are combined, $C1_{EVER}$ also produces the best overall performance.

Cost Function	Difference							Similarity
	Length	Veg	POIs	Turns	Dwel	DPs	Land	
$C1_{EVER}$	0.0%	40.1%	95.6%	49.8%	60.7%	11.97%	66.3%	50.9%
$C7_{EVER}$	13.1%	34.2%	95.0%	37.4%	66.5%	3.81%	67.7%	47.5%
$C9_{EVER}$	10.3%	35.2%	95.4%	36.6%	65.3%	1.37%	67.5%	43.1%
$C11_{EVER}$	9.3%	35.5%	95.3%	37.4%	66.7%	1.35%	67.2%	47.5%
$C12_{EVER}$	8.5%	37.5%	95.3%	21.6%	63.7%	4.74%	70.4%	41.4%

Table 6.3: Attribute and 25th similarity overall evaluation metrics for each of the everyday cost functions. Grey boxes indicate the best values for each metric.

(a) (b)
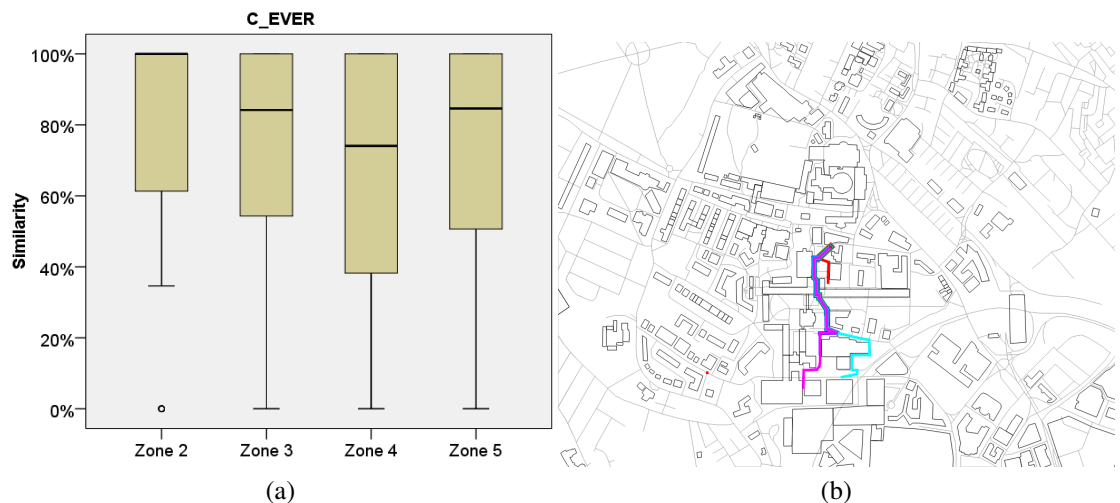
Figure 6.13: Results for the routes generated by the best weighted everyday (C_{EVER}) algorithm. Shown are mean route (a) route similarity and (b) example routes a single start point. For (a) boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range. For (b) the different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

(a)

(b)

(c)

(d)

(e)

(f)

(g)
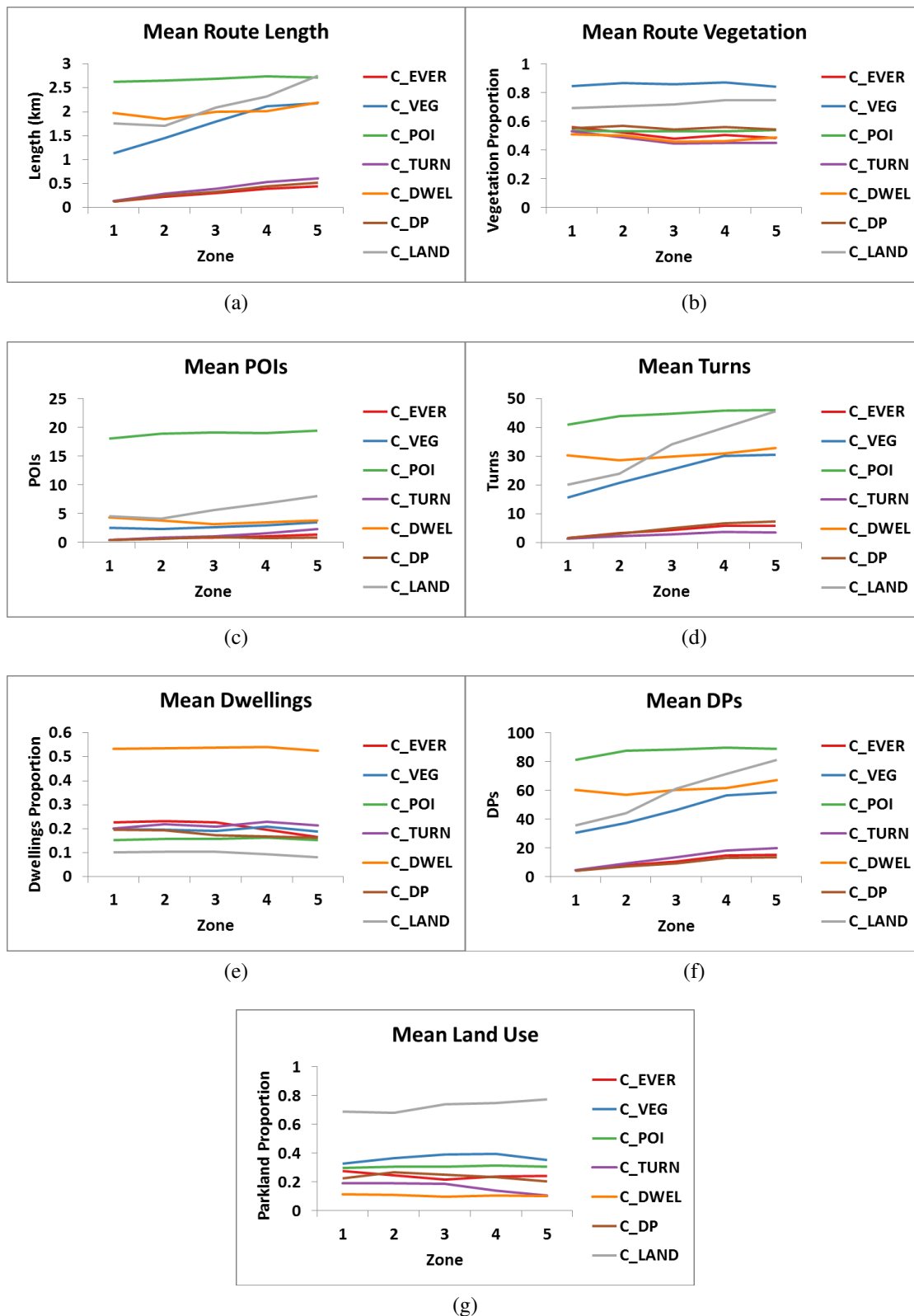
Figure 6.14: Results for the routes generated by the best weighted everyday (C_{EVER}) algorithm. Shown are mean route (a) length, (b) vegetation, (c) POIs, (d) turns, (e) dwellings, (f) DPs and (a) land use.

The result of this process suggests that $C1_{EVER}$ should be selected as the best everyday cost function, and this will therefore be referred to as C_{EVER} from this point forward as indicated by Equation 6.31.

$$C_{EVER} = C_{LEN} \tag{6.31}$$

Figure 6.13 shows the the similarities and example routes Figure 6.13 shows the similarities and example routes for C_{EVER}, and Figure 6.14 shows a comparison to the appropriate single attribute cost functions. However, although C_{EVER} has been chosen as the most appropriate for everyday travel, there is one obvious potential drawback. As C_{EVER} is identical to the shortest length algorithm (C_{LEN}), it considers only one of the attributes found to be influential for everyday journeys in Chapter 3. Whether or not this lack of influential attributes is justified, will be determined by comparing the routes chosen by the algorithm with those chosen by human participants in the following chapter (Chapter 7).

6.4 The Tourist Journey Type

The final column of Table 5.1 indicates that points of interest, dwellings, vegetation, land use and length are all important attributes for tourist journeys. Combining the attributes according to rank, produces the six possible tourist cost functions shown in Equations 6.32 to Equation 6.37 below:

$$C1_{TOUR} = C_{POI} \tag{6.32}$$

$$C2_{TOUR} = w_1 * C_{POI} + w_2 * C_{DWELL} \tag{6.33}$$

$$C3_{TOUR} = w_1 * C_{POI} + w_2 * C_{VEG} \tag{6.34}$$

$$C4_{TOUR} = w_1 * C_{POI} + w_2 * C_{DWELL} + w_3 * C_{VEG} \tag{6.35}$$

$$C5_{TOUR} = w_1 * C_{POI} + w_2 * C_{DWELL} + w_3 * C_{VEG} + w_4 * C_{Land} \tag{6.36}$$

$$C6_{TOUR} = w_1 * C_{POI} + w_2 * C_{DWELL} + w_3 * C_{VEG} + w_4 * C_{Land} + w_5 * C_{LENCAP} \quad (6.37)$$

As with everyday travel, the first stage in determining which one of these cost functions selects the most appropriate routes for tourist travel is to find the best weights for each attribute combination. Before investigating the effect of adding attributes to the leisure cost function, a complication in determining the most appropriate cost function for tourist journeys remains - what is the most appropriate value of $Length_{Limit}$ for C_{LENCAP}? Given the size of the environment being considered, a time limit of 30 minutes was chosen for test purposes (this was also reflected in the time limit given in the evaluation user study discussed in a later chapter), and with an average human walking speed of approximately 4km/h, this gives a $Length_{Limit}$ of 2km. Inserting this value into the cost functions which include C_{LENCAP}. Using this threshold and running the same evaluation process and 174 points as described in Section 6.1.1 produces a set of weights for each cost function as shown in Equation 6.38 to Equation 6.42.

$$C2_{TOUR} = 0.654297 * C_{POI} + 0.070313 * C_{DWELL} \quad (6.38)$$

$$C3_{TOUR} = 0.8125 * C_{POI} + 0.982422 * C_{VEG} \quad (6.39)$$

$$C4_{TOUR} = 0.464844 * C_{POI} + 0.515625 * C_{DWELL} + 0.992188 * C_{VEG} \quad (6.40)$$

$$C5_{TOUR} = 0.312012 * C_{POI} + 0.935547 * C_{DWELL} + 0.814453 * C_{VEG} + 0.53418 * C_{LAND}$$
$$(6.41)$$

$$C6_{TOUR} = 0.875 * C_{POI} + 0.375 * C_{DWELL} + 0.972656 * C_{VEG} + 0.5625 * C_{Land} +$$
$$0.59375 * C_{LENCAP}$$

where:

$$Length_{Limit} = 2.0$$

$$(6.42)$$
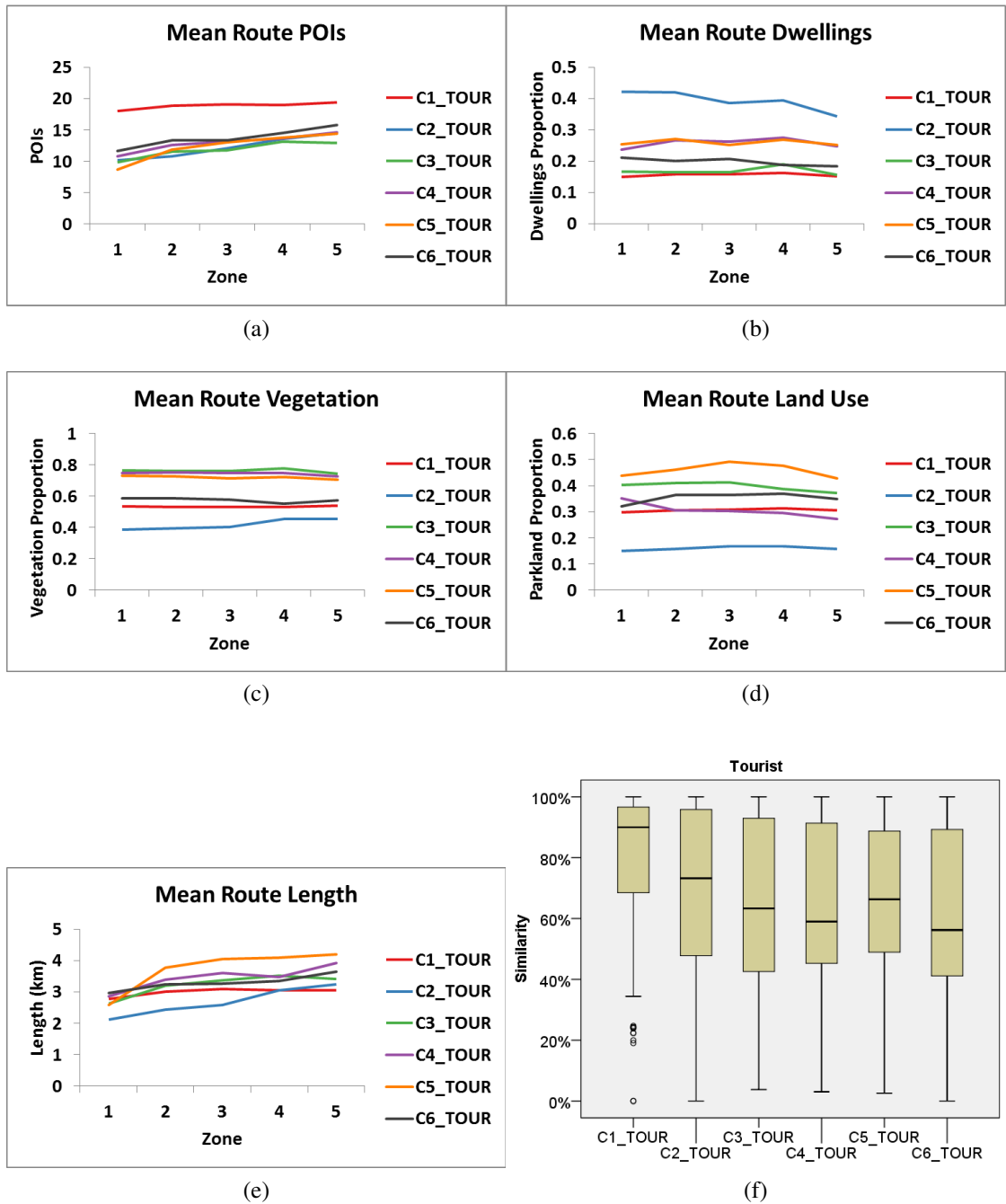
Figure 6.15: Results of different tourist cost functions showing mean route (a) POIs, (b) dwellings, (c) vegetation, (d) land use, (e) length and (f) similarity. For (f) boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range.

Figure 6.15 shows a comparison of the attribute and similarity results produced by these six tourist cost functions, indicating that each function has its advantages and disadvantages. For example, $C2_{TOUR}$ (Equation 6.40) has a particularly high proportion of

dwellings (Figure 6.15b), but is poor in terms of vegetation (Figure 6.15c) and parkland (Figure 6.15d). However, as with the everyday results, Figure 6.15 does not give a clear indication of which cost function performs best.

Cost Function	Difference					Similarity
	POIs	Dwellings	Vegetation	Land Use	Length	
$C1_{TOUR}$	0.0%	70.7%	37.8%	57.5%	142.3%	68.5%
$C2_{TOUR}$	36.3%	25.7%	51.6%	77.7%	124.9%	47.2%
$C3_{TOUR}$	37.8%	68.4%	11.1%	44.5%	152%	42.6%
$C4_{TOUR}$	32.0%	51.6%	13.1%	57.1%	162.1%	45.3%
$C5_{TOUR}$	35.7%	51.3%	15.9%	36.1%	174.9%	48.9%
$C6_{TOUR}$	28.2%	63.0%	32.9%	50.3%	155.1%	41.1%

Table 6.4: Attribute and 25th similarity overall evaluation metrics for each of the tourist cost functions. Grey boxes indicate the best values for each metric, and black the worst.

Table 6.4 shows the overall results for all six tourist cost functions, and indicates that cost functions that produce the best results for one attribute are likely to produce the worst for others. To select the best performing cost function, the same three stage process as in everyday routes was followed.

1. **Cost functions which produce excessively long routes are discounted.** Unlike everyday journeys, not all of the cost functions for tourist journeys produce routes which could be considered appropriate for pedestrian journeys. $C2_{TOUR}$, $C3_{TOUR}$ and $C4_{TOUR}$ all produce routes of over 5km, and were therefore discounted at this stage.

2. **The worst performing cost function in terms of each attribute was identified and discounted.** From Table 6.4 this suggests that $C1_{TOUR}$ and $C6_{TOUR}$ should also be removed from the pool of possible candidates.

The first two stages of the three stage selection process therefore rule out all of the cost functions tested so far. As it leads to the removal of all but two candidates, this may indicate that length plays a greater part in the selection of tourist routes than suggested in Experiment 6 (Section 3.9). One possible solution to this problem could therefore be to include C_{LENCAP} much earlier in the process. This produces the additional equations shown in equation 6.43 to equation 6.46, all with $Length_{Limit} = 2.0$.

$$C7_{TOUR} = 0.84375 * C_{POI} + 0.46875 * C_{LENCAP} \qquad (6.43)$$

$$C8_{TOUR} = 0.921875 * C_{POI} + 0.03125 * C_{DWELL} + 0.953125 * C_{LENCAP} \qquad (6.44)$$

$$C9_{TOUR} = 0.75 * C_{POI} + 0.867188 * C_{VEG} + 0.796875 * C_{LENCAP} \qquad (6.45)$$

$$C10_{TOUR} = 0.796875 * C_{POI} + 0.314545 * C_{DWELL} + 0.953503 * C_{VEG} +$$
$$0.598206 * C_{LENCAP} \qquad (6.46)$$

Table 6.5 shows the overall results for the four new tourist cost functions, plus $C1_{TOUR}$ and $C6_{TOUR}$ which produce routes with a maximum length of 5km or less.

Cost	Difference					Similarity
Function	POIs	Dwellings	Vegetation	Land Use	Length	
$C1_{TOUR}$	0.0%	70.7%	37.8%	57.5%	142.3%	68.5%
$C6_{TOUR}$	28.2%	63.0%	32.9%	50.3%	155.1%	41.1%
$C7_{TOUR}$	35.7%	65.5%	40.9%	65.54%	145.7%	31.6%
$C8_{TOUR}$	35.9%	64.2%	42.2%	65.88%	142.8%	31.2%
$C9_{TOUR}$	29.2%	56.6%	32.4%	58.85%	148.7%	35.7%
$C10_{TOUR}$	25.9%	61.1%	35.2%	57.55%	156.1%	41.3%

Table 6.5: Attribute and 25th similarity overall evaluation metrics for each of the tourist cost functions. Grey boxes indicate the best values for each metric, and black the worst.

With the first stage of the selection process already having been completed by the removal of routes that are longer than those considered appropriate for pedestrian journeys, the Rea mining two stages were then repeated.

2. **The worst performing cost function in terms of each attribute was identified and discounted.** From Table 6.5 this suggests that $C1_{TOUR}$, $C8_{TOUR}$ and $C10_{TOUR}$ should also be removed from the pool of possible candidates.

3. **The overall performance of the remaining cost functions was compared, and the best performing was selected.** The performance of the three remaining candidate cost functions ($C6_{TOUR}$, $C7_{TOUR}$ and $C9_{TOUR}$) is shown in Table 6.6. Of these, $C6_{TOUR}$ outperforms the others in two of the five attributes and has the best similarity. In addition, if the attribute differences are combined, $C6_{TOUR}$ also produces the best overall performance.

Cost	Difference					Similarity
Function	POIs	Dwellings	Vegetation	Land Use	Length	
$C6_{TOUR}$	28.2%	63.0%	32.9%	50.3%	155.1%	41.1%
$C7_{TOUR}$	35.7%	65.5%	40.9%	65.54%	145.7%	31.6%
$C9_{TOUR}$	29.2%	56.6%	32.4%	58.85%	148.7%	35.7%

Table 6.6: Attribute and 25th similarity overall evaluation metrics for each of the tourist cost functions. Grey boxes indicate the best values for each metric.
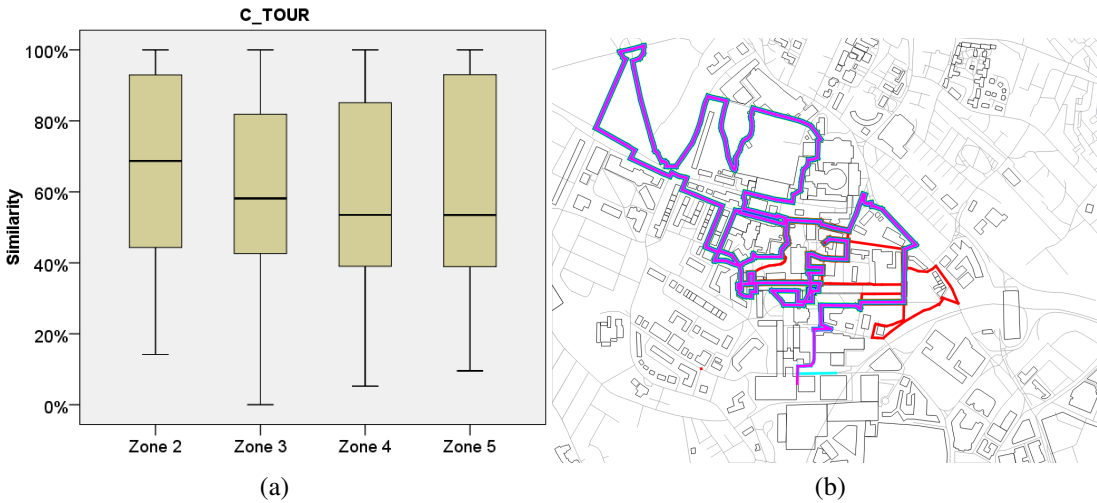
Figure 6.16: Results for the routes generated by the best weighted tourist (C_{TOUR}) algorithm. Shown are mean route (a) route similarity and (b) example routes for a single start point and sector. For (a) boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range. For (b) the different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

The result of this process suggests that $C6_{TOUR}$ should be selected as the best everyday cost function, and this will therefore be referred to as C_{TOUR} from this point forward as indicated by Equation 6.47.

$$C_{TOUR} = 0.875 * C_{POI} + 0.375 * C_{DWELL} + 0.972656 * C_{VEG} + 0.5625 * C_{Land} +$$
$$0.59375 * C_{LENCAP}$$

where:

$$Length_{Limit} = 2.0$$

(6.47)

Figure 6.16 shows the route similarities and example routes for C_{TOUR}, and Figure 6.17 shows shows a comparison to the appropriate single attribute cost functions. Unlike

C_{EVER}, C_{TOUR} includes all of the attributes found to influence tourist routes in Chapter 3, but it is important to note that C_{TOUR} produces routes which may still be too long for pedestrian travel. This may mean that the length will actually need to be more closely restricted, and may indicate that length plays a greater role in route selection than suggested by Experiment 6 (Section 3.9). Again these results and hypothesis will be tested against human suggested routes in the following chapter.

(a)

(b)

(c)

(d)

(e)
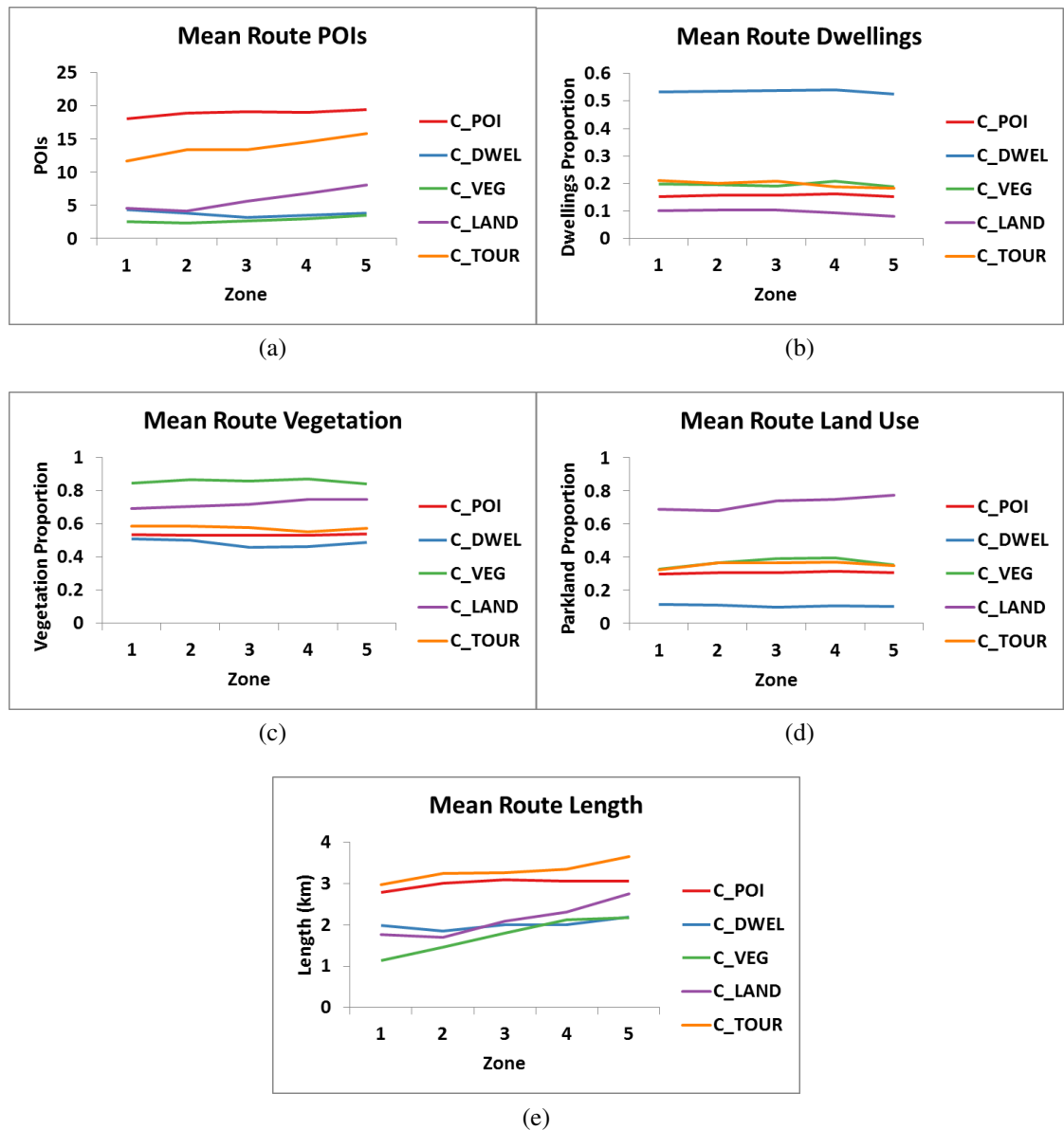
Figure 6.17: Results for the routes generated by the best weighted tourist (C_{TOUR}) algorithm. Shown are mean route (a) POIs, (b) dwellings, (c) vegetation, (d) land use and (e) length.

6.5   The Leisure Journey Type

For leisure journeys, Table 5.1 indicates that up to seven attributes influence route choice. As with everyday and tourist routes, adding one attribute at a time to the algorithm and evaluating the routes produced allows the best weight and attribute combinations to be found. This approach produces the cost functions shown in Equation 6.48 to Equation 6.59:

$$C1_{LEIS} = C_{VEG} \tag{6.48}$$

$$C2_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} \tag{6.49}$$

$$C3_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{LAND} \tag{6.50}$$

$$C4_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{DWEL} \tag{6.51}$$

$$C5_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{LENCAP} \tag{6.52}$$

$$C6_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{LAND} + w_4 * C_{DWEL} \tag{6.53}$$

$$C7_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{LAND} + w_4 * C_{LENCAP} \tag{6.54}$$

$$C8_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{DWEL} + w_4 * C_{LENCAP} \tag{6.55}$$

$$C9_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{LAND} + w_4 * C_{DWEL} + w_5 * C_{LENCAP} \tag{6.56}$$

$$C10_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{LAND} + w_4 * C_{DWEL} + w_5 * C_{LENCAP} \\ + w_6 * C_{DP} \tag{6.57}$$

$$C11_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{LAND} + w_4 * C_{DWEL} + w_5 * C_{LENCAP}$$
$$+ w_6 * C_{TURN}$$

(6.58)

$$C12_{LEIS} = w_1 * C_{VEG} + w_2 * C_{POI} + w_3 * C_{LAND} + w_4 * C_{DWEL} + w_5 * C_{LENCAP}$$
$$+ w_6 * C_{DP} + w_7 * C_{TURN}$$

(6.59)

Calculating the 'best' weights for each cost function be running the same evaluation process and 174 points as described in Section 6.1.1, gives the weight combinations shown in Equation 6.60 to Equation 6.70 below.

$$C2_{LEIS} = 0.991943 * C_{VEG} + 0.730469 * C_{POI}$$

(6.60)

$$C3_{LEIS} = 0.5 * C_{VEG} + 0.5 * C_{POI} + 0.734375 * C_{LAND}$$

(6.61)

$$C4_{LEIS} = 0.992188 * C_{VEG} + 0.464844 * C_{POI} + 0.515625 * C_{DWEL}$$

(6.62)

$$C5_{LEIS} = 0.867188 * C_{VEG} + 0.75 * C_{POI} + 0.796875 * C_{LENCAP}$$
where:
$$Length_{Limit} = 2.0$$

(6.63)

$$C6_{LEIS} = 0.814453 * C_{VEG} + 0.312012 * C_{POI} + 0.53418 * C_{LAND} +$$
$$0.935547 * C_{DWEL}$$

(6.64)

$$C7_{LEIS} = 0.766357 * C_{VEG} + 0.585938 * C_{POI} + 0.391846 * C_{LAND} +$$
$$0.526367 * C_{LENCAP}$$
where:
$$Length_{Limit} = 2.0$$

(6.65)

$$C8_{LEIS} = 0.953571 * C_{VEG} + 0.941589 * C_{POI} + 0.314484 * C_{DWEL} +$$
$$0.598175 * C_{LENCAP}$$
where:
$$Length_{Limit} = 2.0$$
(6.66)

$$C9_{LEIS} = 0.972656 * C_{VEG} + 0.875 * C_{POI} + 0.5625 * C_{LAND} + 0.375 * C_{DWEL} +$$
$$0.59375 * C_{LENCAP}$$
where:
$$Length_{Limit} = 2.0$$
(6.67)

$$C10_{LEIS} = 0.958984 * C_{VEG} + 0.480469 * C_{POI} + 0.835938 * C_{LAND} +$$
$$0.148438 * C_{DWEL} + 0.070313 * C_{LENCAP} + 0.523438 * C_{DP}$$
where:
$$Length_{Limit} = 2.0$$
(6.68)

$$C11_{LEIS} = 0.457031 * C_{VEG} + 0.962158 * C_{POI} + 0.28125 * C_{LAND} +$$
$$0.761719 * C_{DWEL} + 0.570313 * C_{LENCAP} + 0.828125 * C_{TURN}$$
where:
$$Length_{Limit} = 2.0$$
(6.69)

$$C12_{LEIS} = 0.792969 * C_{VEG} + 0.804688 * C_{POI} + 0.953125 * C_{LAND} +$$
$$0.140625 * C_{DWEL} + 0.84375 * C_{LENCAP} + 0.695313 * C_{DP} + 0.25 * C_{TURN}$$
where:
$$Length_{Limit} = 2.0$$
(6.70)

Figure 6.18 and Figure 6.19 indicate the performance of the cost functions with respect to each attribute, but as with the everyday and tourist journeys, they are not clear enough to determine which is the most appropriate.

Table 6.7 shows the overall evaluation metrics for the cost functions, allowing easier
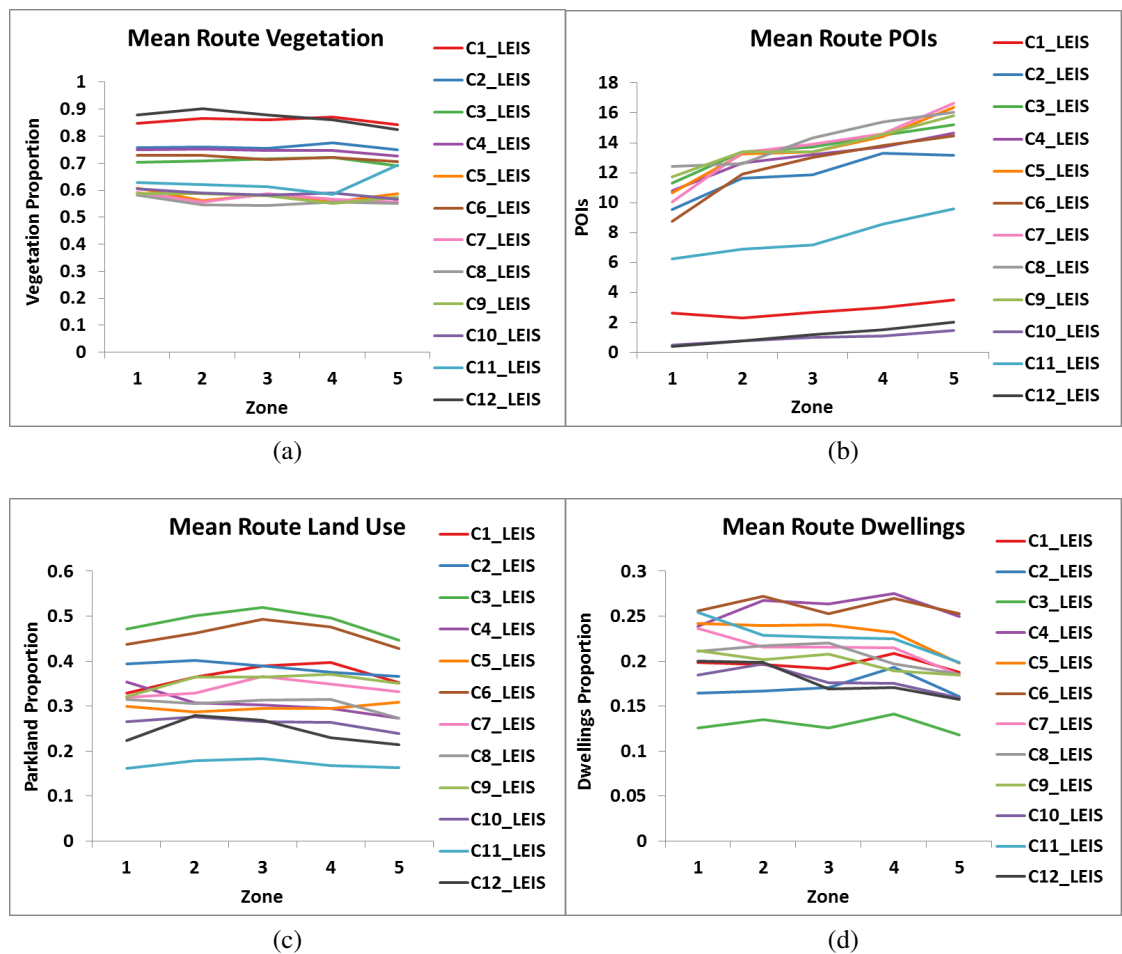
Figure 6.18: Evaluation metrics - showing the mean route (a) vegetation, (b) parkland, (c) dwellings and (d) POIs for the routes generated by the leisure cost algorithms.

comparison.

Selection of the best performing cost function, followed the same three stage process as was used for the everyday and tourist algorithms.

1. **Cost functions which produce excessively long routes are discounted.** As with tourist journeys, not all of the cost functions for leisure journeys produce routes which could be considered appropriate for pedestrian journeys. $C1_{LEIS}$, $C2_{LEIS}$, $C3_{LEIS}$, $C4_{LEIS}$ and $C6_{LEIS}$ all produce routes of over 5km, and were therefore discounted at this stage.

2. **The worst performing cost function in terms of each attribute was identified and discounted.** From Table 6.7 this suggests that $C5_{LEIS}$, $C10_{LEIS}$ and $C11_{LEIS}$ should also be removed from the pool of possible candidates.
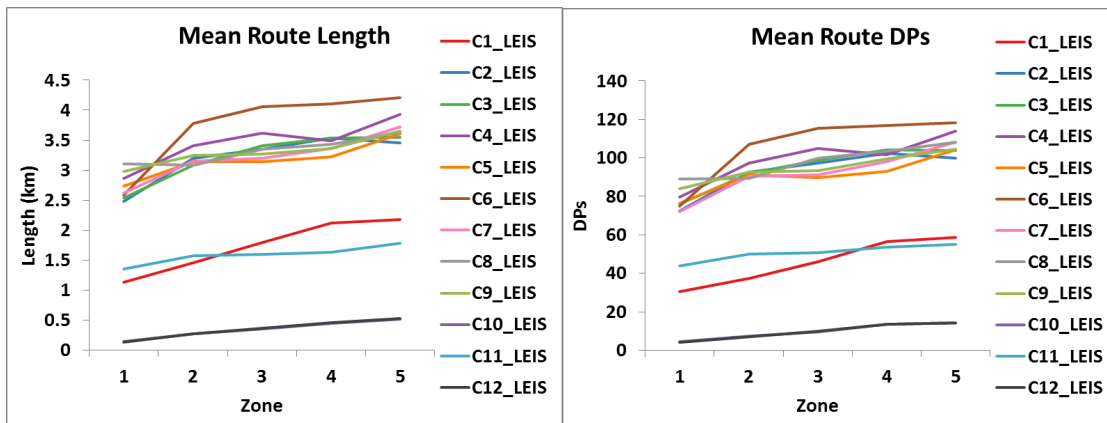
(a)

(b)
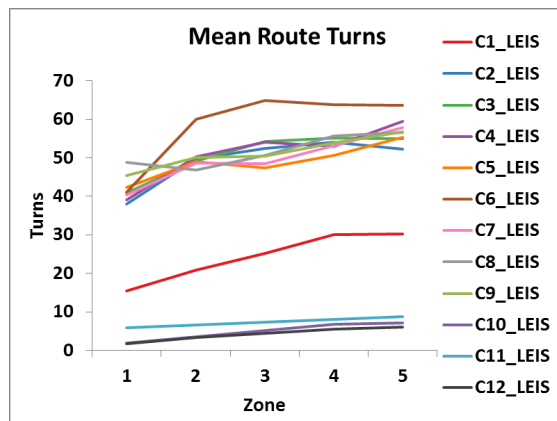
(c)
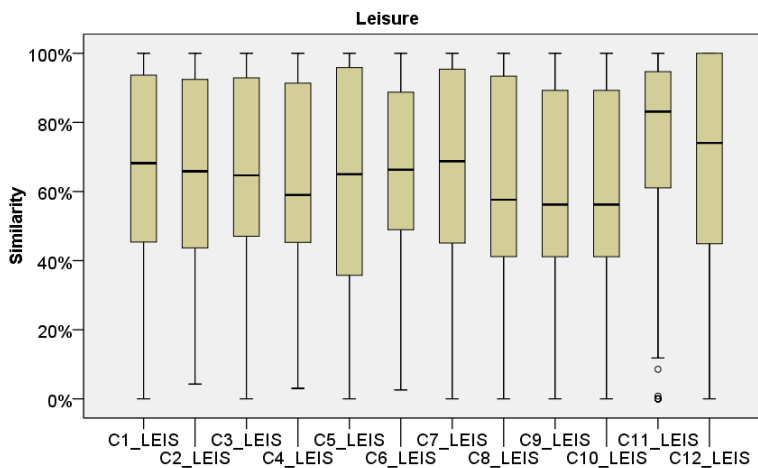
(d)

Figure 6.19: Evaluation metrics - showing the mean route (a) length, (b) DPs, (c) turns, and (d) similarity for the routes generated by the leisure cost algorithms. For (d) boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range.

| Cost | Difference | | | | | | | Simil- |
Function	Veg	POIs	Land	Dwel	Length	DPs	Turns	arity
$C1_{LEIS}$	0.0%	85.4%	49.3%	63.3%	20.2%	382.9%	754.3%	45.4%
$C2_{LEIS}$	11.5%	37.9%	46.2%	68.1%	50.2%	879.4%	1629.9%	43.5%
$C3_{LEIS}$	17.4%	28.8%	32.1%	75.8%	50.7%	901.6%	1684.3%	46.8%
$C4_{LEIS}$	13.1%	32.0%	57.1%	51.6%	62.1%	948.7%	1695.6%	45.3%
$C5_{LEIS}$	32.4%	29.2%	58.8%	56.6%	48.7%	857.9%	1616.9%	35.7%
$C6_{LEIS}$	15.9%	35.7%	36.1%	51.3%	74.9%	1021.5%	1957.9%	48.9%
$C7_{LEIS}$	33.2%	28.9%	53.0%	59.7%	50.0%	869.7%	1642.1%	45.1%
$C8_{LEIS}$	35.2%	26.2%	57.5%	61.1%	56.0%	931.9%	1714.2%	41.1%
$C9_{LEIS}$	32.7%	28.1%	51.0%	62.5%	55.1%	898.9%	1701.2%	41.1%
$C10_{LEIS}$	31.3%	95.1%	63.3%	66.3%	84.6%	3.4%	69.9%	41.1%
$C11_{LEIS}$	39.1%	60.2%	76.1%	57.2%	25.5%	433.8%	155.6%	60.5%
$C12_{LEIS}$	34.4%	94.2%	65.9%	66.1%	84.3%	3.6%	47.0%	44.9%

Table 6.7: Attribute and 25th similarity overall evaluation metrics for each of the leisure cost functions (where length is compared to $C_{LEN_{CAP}}$). Grey boxes indicate the best values for each metric, and black the worst.

3. **The overall performance of the remaining cost functions was compared, and the best performing was selected.** The performance of the four remaining candidate cost functions ($C7_{LEIS}$, $C8_{LEIS}$, $C9_{LEIS}$ and $C12_{LEIS}$) is shown in Table 6.8. Of these, $C7_{LEIS}$ outperforms the others in two of the seven attributes, and has the best similarity. However, if the attribute differences are combined, $C12_{EVER}$ produces the best overall performance (mostly due to the high values for DP and turns in $C7_{LEIS}$).

| Cost | Difference | | | | | | | Simil- |
Function	Veg	POIs	Land	Dwel	Length	DPs	Turns	arity
$C7_{LEIS}$	33.2%	28.9%	53.0%	59.7%	50.0%	869.7%	1642.1%	45.1%
$C8_{LEIS}$	35.2%	26.2%	57.5%	61.1%	56.0%	931.9%	1714.2%	41.1%
$C9_{LEIS}$	32.7%	28.1%	51.0%	62.5%	55.1%	898.9%	1701.2%	41.1%
$C12_{LEIS}$	34.4%	94.2%	65.9%	66.1%	84.3%	3.6%	47.0%	44.9%

Table 6.8: Attribute and 25th similarity overall evaluation metrics for each of the remaining leisure cost functions. Grey boxes indicate the best values for each metric.

This would suggest that either $C7_{LEIS}$ or $C12_{LEIS}$ should become C_{LEIS}, but as with tourist routes the removal of five cost functions due to route length may suggest that C_{LENCAP} should again be included in earlier cost functions. Given the position in terms of rank of length (see Table 6.1), this produces just one additional cost equation as shown below (Equation 6.71).

$$C13_{LEIS} = 0.0.5625 * C_{VEG} + 0.0.816406 * C_{LENCAP}$$

where: (6.71)

$$Length_{Limit} = 2.0$$

Table 6.7 shows the overall evaluation metrics for $C7_{LEIS}$, $C12_{LEIS}$ and $C13_{LEIS}$. Again it indicates that $C7_{LEIS}$ performs best, in this case outperforming $C12_{LEIS}$ and $C7_{LEIS}$ for three of the seven attributes and having the best similarity as before. However, if the attribute differences are combined $C12_{LEIS}$ produces the best overall performance as before (mostly due to the high values for DP and turns in $C7_{LEIS}$). In this case, and as turns and decision points are the lowest ranking attributes, $C7_{LEIS}$ is selected as the best performing cost function for selecting leisure routes, and it therefore becomes C_{LEIS} as shown in Equation 6.72.

Cost	Difference							Simil-
Function	**Veg**	**POIs**	**Land**	**Dwel**	**Length**	**DPs**	**Turns**	**arity**
$C7_{LEIS}$	33.2%	28.9%	53.0%	59.7%	50.0%	869.7%	1642.1%	45.1%
$C12_{LEIS}$	34.4%	94.2%	65.9%	66.1%	84.3%	3.6%	47.0%	44.9%
$C13_{LEIS}$	26.0%	86.9%	57.9%	60.7%	2.5%	422.4%	952.7%	33.5%

Table 6.9: Attribute and 25th similarity overall evaluation metrics for $C7_{LEIS}$, $C12_{LEIS}$ and $C13_{LEIS}$. Grey boxes indicate the best values for each metric.

$$C_{LEIS} = 0.766357 * C_{VEG} + 0.585938 * C_{POI} + 0.391846 * C_{LAND} +$$

$$0.526367 * C_{LENCAP}$$

            (6.72)

where:

$$Length_{Limit} = 2.0$$

Figure 6.20 shows the route similarities and example routes for C_{LEIS}, and Figure 6.21 shows shows a comparison to the appropriate single attribute cost functions. Despite this algorithm producing the most appropriate routes in terms of overall performance, it selects routes which may be considered too long for pedestrian travel (a maximum route length of almost 5km was recorded for the start and end points tested). One solution to this may be to restrict C_{LENCAP} further, but it is likely to compromise the algorithm in terms of the remaining attributes. In addition, the exclusion of three attributes found to be influential in Chapter 3 from this cost function may prove detrimental, as may the 33% reduction in
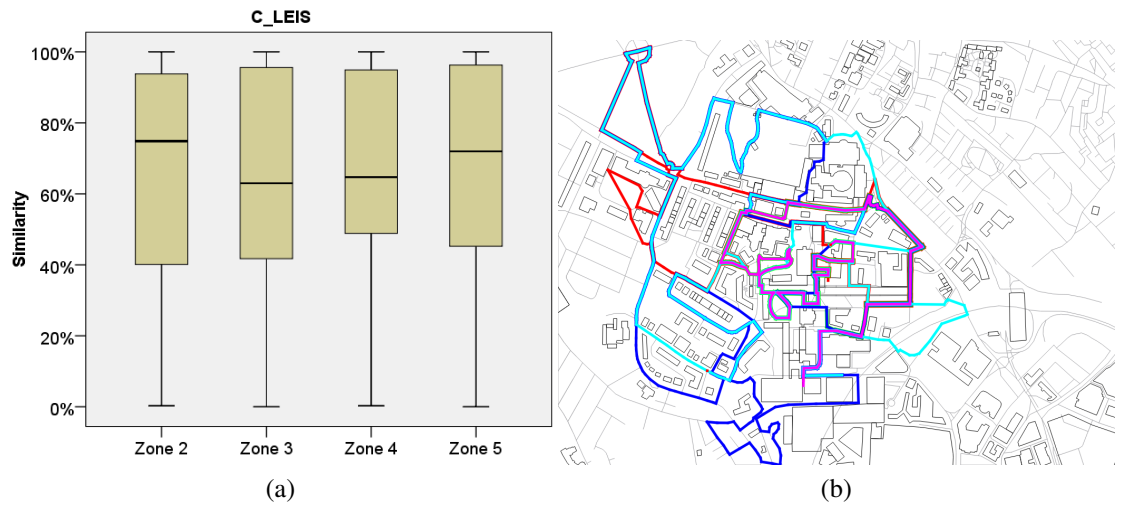
Figure 6.20: Results for the routes generated by the best weighted leisure (C_{LEIS}) algorithm. Shown are mean route (a) route similarity and (b) example routes for a single start point and sector. For (a) boxes indicate 25% to 75% and whiskers indicate these percentages $\pm$ 1.5*interquartile range. For (b) the different colours indicate routes end points in different zones (Zone 1 - red, Zone 2 - green, Zone 3 - blue, Zone 4 - cyan and Zone 5 - magenta).

vegetation coverage reported in Tables 6.9. Again, whether or not these factors prove to be problematic will be tested in the following chapter.

6.6 Algorithm Speed Performance

As with the single attribute cost functions, the multi-attribute cost functions speeds were also tested, using the same approach as in Section 5.4, with five iterations of the 174 start-end points. The tests were again performed, using single threaded code, on a Dell XPS, with a 1.90GHz Intel(R) Core(TM) i7-3517U CPU, and 8GB of RAM. Here only the unbucketed, one-dimensional bucket and fixed length ten two-dimensional bucket approaches were compared, and the mean running times are shown in Figure 6.22. All of the approaches produce a maximum runtime of 0.0156 seconds for all cost functions.

(a)

(b)

(c)

(d)

(e)

(f)

(g)
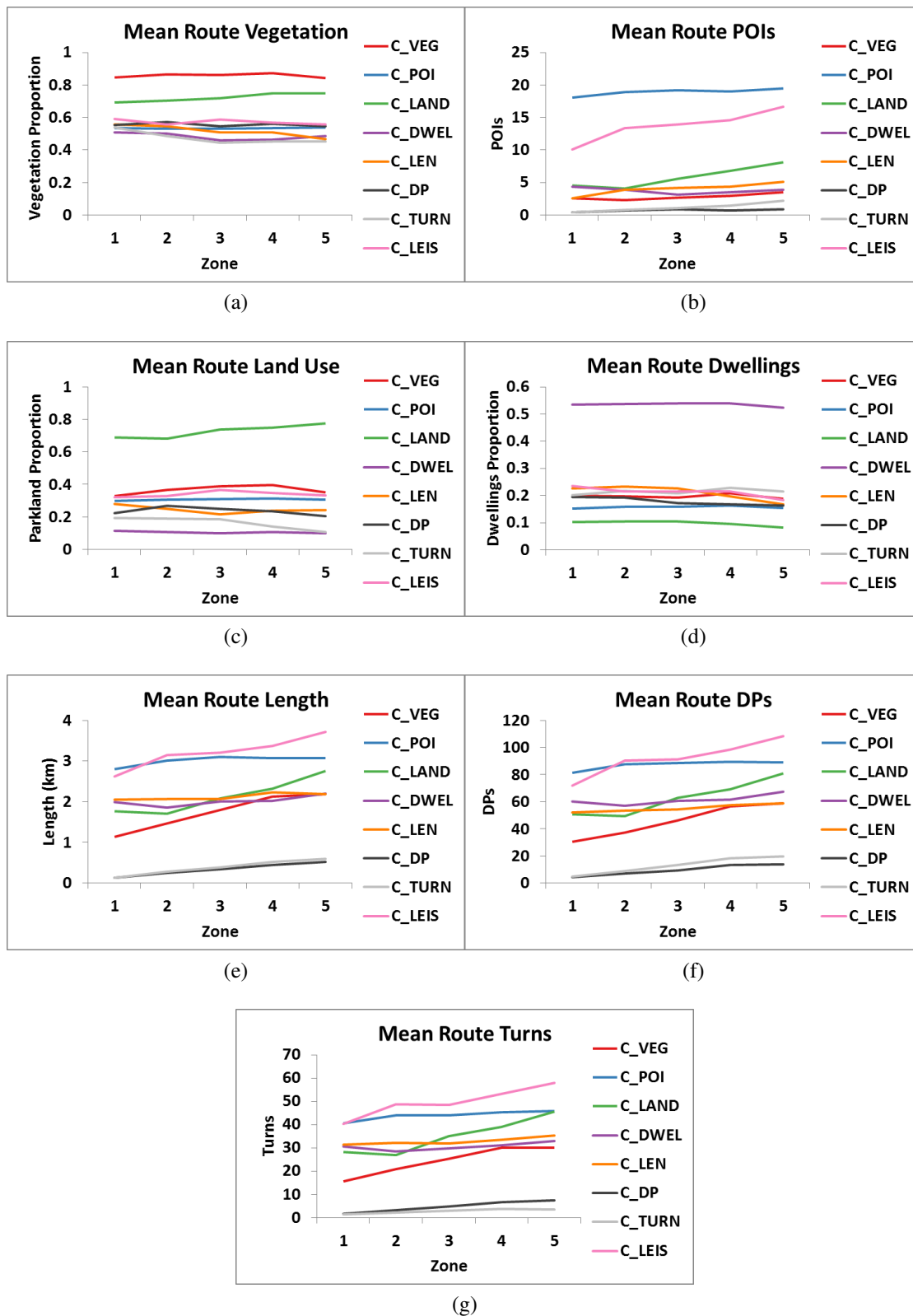
Figure 6.21: Results for the routes generated by the best weighted leisure (C_{LEIS}) algorithm. Shown are mean route (a) vegetation, (b) POIs, (c) land use, (d) dwellings, (e) length, (f) DPs and (g) turns.
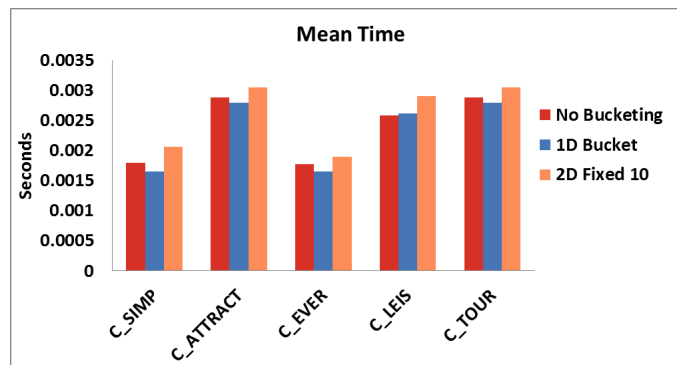
Figure 6.22: (a) Mean running times for the unoptimised and 1D bucketed and 2D fixed size 10 bucketed algorithms to select 870 routes.

As expected the one-dimensional bucketed approach produces the fastest speeds, with the unoptimised algorithm producing better mean times for only the leisure routes. With an overall mean running time of 0.0023 seconds across all of the multi-attribute cost functions, and a maximum of 0.0156 seconds to produce a single route, the algorithms would all be appropriate for realtime applications using this approach. This indicates that the one-dimensional bucketing is the most best optimisation technique, of those tested, for selecting routes in the present research.

6.7 Summary

Table 6.10 shows a summary of the multi-attribute cost functions for the simplicity and attractiveness categories, and everyday, tourist and leisure journey types. Each of these was determined by weight optimisation and evaluation, using metrics related to attribute value and similarity, based on the method that was empirically derived for simplicity first. This method produced convincing results, but it does have limitations, as do the metrics used to derive it. The most obvious of these is that there is no indication that the use of attribute value or similarity is the most appropriate approach to evaluating routes. The linear summation of attribute costs is also not the only possible approach to calculating route cost, and may not be the best solution. With no standard evaluation method or empirical evidence as to the correctness or otherwise of summing attribute costs, the approach taken by this chapter is considered satisfactory given the context of the present research. It is also possible that the weight combinations have been allowed to converge on a local maximum rather than the best possible solution, although it was hoped the use of ranges rather single values in the intermediate steps would avoid this.

Overall each of the algorithms produces the best performance with respect to its con-

Category / Journey Type	Cost Function
Simplicity	$C_{SIMP} = 0.578125 * C_{LEN} + 0.828125 * C_{DP} + 0.25 * C_{TURN}$
Attractiveness	$C_{ATTRACT} = 0.814453 * C_{VEG} + 0.53418 * C_{LAND} + 0.935547 * C_{DWEL} + 0.312012 * C_{POI}$
Everyday	$C_{EVER} = C_{LEN}$
Tourist	$C_{TOUR} = 0.875 * C_{POI} + 0.375 * C_{DWELL} + 0.972656 * C_{VEG} + 0.5625 * C_{Land} + 0.59375 * C_{LENCAP}$
Leisure	$C_{LEIS} = 0.766357 * C_{VEG} + 0.585938 * C_{POI} + 0.391846 * C_{LAND} + 0.526367 * C_{LENCAP}$

Table 6.10: Summary of the multi-attribute cost functions and algorithms. For $C_{LENCAP} Length_{Limit} = 2.0$

stituent attributes, and suitable levels of similarity. The algorithms also select routes in times suitable for realtime applications. However there may be several drawbacks to the chosen cost functions. C_{EVER} for example contains only one of the seven attributes found to be influential, which may not be enough to select suitable everyday routes. In contrast C_{TOUR} uses all five of the attributes considered influential for tourist journeys, but this may dilute the effects of the higher ranking ones. Additionally, there is no direct evidence that applying these cost functions to points other than those tested will select appropriate routes. In the majority of cases, the leisure and tourist cost functions also selected routes that may be considered too long for pedestrian journeys. This indicates that tighter restrictions on length may be required for these journey types. With only one area being used to generate and evaluate the performance of these algorithms, biases may also have been introduced by artifacts of this particular test area. If this proves to be the case then the weights generated may not be applicable to other areas, or alternative cost functions may provide better solutions.

To determine whether or not the limitations posed by the inclusion or or exclusion of attributes within the cost functions are detrimental, comparison with human suggested routes is required. By comparing not only the best cost function for each journey type but all of those suggested by the previous sections, the possible bias introduced by using a limited number of influential attributes can be considered more fully. This comparison against human suggested routes is the next step, with the approach taken and results produced detailed in the following chapter.

Chapter 7

Evaluating the Routes

In order to evaluate the routes selected by each algorithm, human suggestions for the same attribute categories and journey types must first be collected. To gather these route suggestions, a series of trials were conducted to collate the required information from a small number of participants. The study used the same area as was described in Section 4.1.4, and the attribute categories (simplicity and attractiveness) and journey types (everyday, leisure and tourist) as described in Section 3.6. This chapter gives the details of this study, an analysis of the resulting routes, and a comparison of the algorithm's suggested routes with participants' routes.

Prior to the start of this experiment, a short pilot was run to test the materials and procedure. Participants were presented with the scenario 'Imagine that someone has approached you and asked you for the nicest route from point A to point B. Please describe the route you would suggest', and for attractiveness proceeded to select routes which were very similar to those they chose for the simplicity category. In the majority of cases only a few links differed, despite vegetation, points of interest and parkland all being given as reasons for these variations. Participants indicated that the length of the route was a major consideration, found selecting attractive routes difficult, and seemed only to consider the immediate area around the start and end points rather than the whole map. For these reasons, the attractiveness category was removed from the main study. During the pilot, three important links were found to be missing from the map, and those errors were corrected and all of the cost functions and results regenerated. This pilot was also used to fine tune the questions that were asked, and to evaluate the materials and number of

trials for each participant. Comments made after the pilot suggested that the task became monotonous towards the end, and this may lead to a lack of participant concentration. By removing the attractiveness category these comments were also addressed.

7.1 Method

This experiment was designed to gather participant route suggestions for comparison to algorithm selected routes. A within participant random block design was used where the participants were supplied with five start/end pairs, and asked to suggest the simplest route and those suitable for each of the three journey types.

7.1.1 Participants

A total of 14 individuals (3 females; 11 males) aged 18 to 44 (mean 27.6 years, SD 7.5 years) participated in this experiment, and all were either university students or members of staff. This cohort was considered more appropriate than any others, as they were more likely to have at least some familiarity with the test area than the general population. The experiment was approved by the Faculty Ethics Committee, and informed consent was provided by participants completing the relevant consent forms. Posters were used to advertise the study and attract participants, as per ethical guidelines.

7.1.2 Materials

As the participant pool for this experiment was restricted to university students and staff, five pairs of points on campus were chosen as start and end points for gathering route suggestions. Three of the start points were taken directly from those described in Section 5.1. However two of the previous start points were outside of the campus area, and were therefore replaced by points that were within the prescribed area. The tool's automatic points selection facility (as described in Section 5.1), was then used to generate 148 end points. Of these, 64 points were rejected as either being placed outside campus, or generating routes which had more than 50% of their length outside of the chosen area. The remaining 84 end points were then divided according to zone, and one start-end pair was randomly selected per zone. This process resulted in the five point pairs shown in Figure 7.1, and detailed in Table 7.1.

Pairs of start and end points were displayed on a satellite image of the environment, with photos of each point and text explaining their location, as shown in Figure 7.2a. Participants were also provided with a high resolution laminated A2 map showing the test
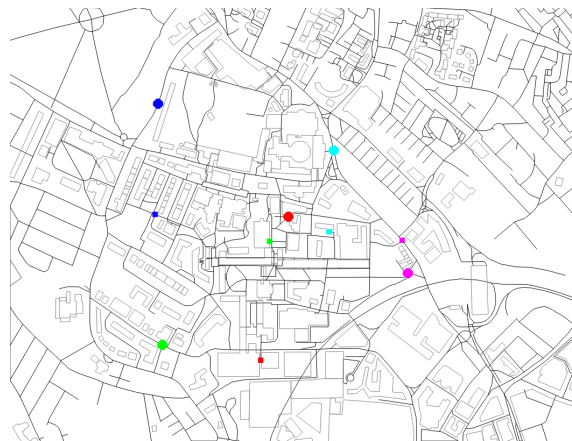
Figure 7.1: Evaluation start-end point pairs. Pair 1 - magenta, pair 2 - cyan, pair 3 - blue, pair 4 - green and pair 5 - red.

Zone	Colour	Point	Description
1	Magenta	A	Fenton Street Entrance
		B	Print Shop (Woodhouse Lane)
2	Cyan	A	Parkinson Steps
		B	Opposite Physics Deck (Hillary Place)
3	Blue	A	Henry Price Residences
		B	Outside Bright Beginnings Nursery
4	Green	A	Behind the Faversham
		B	Between Edward Boyle and Maths
5	Red	A	Corner of Ziff Building
		B	Between Worsley and Light Buildings

Table 7.1: Description of the evaluation point pairs. Colour refers to the points shown in Figure 7.1, point A is the start point and point B is the end point of each pair.

environment (taken from OpenStreetMap), as shown in Figure 7.2b. This had building and street names labelled, and each start-end point pair marked as coloured dots (with a larger dot used for the start point).

In order to provide their route suggestions, participants were also given paper maps (a separate map for each start-end pair, see Figure 7.3a), paper and writing materials, as well as the sessions being recorded on an audio device. A questionnaire was supplied (part of which is shown in Figure 7.3b), on which participants could indicate their familiarity (using a four point Likert Scale) with the points and the routes that they had chosen, and give reasons for their suggestions. A full copy of the materials used and an example of the series of images that were displayed are shown in Appendix B.
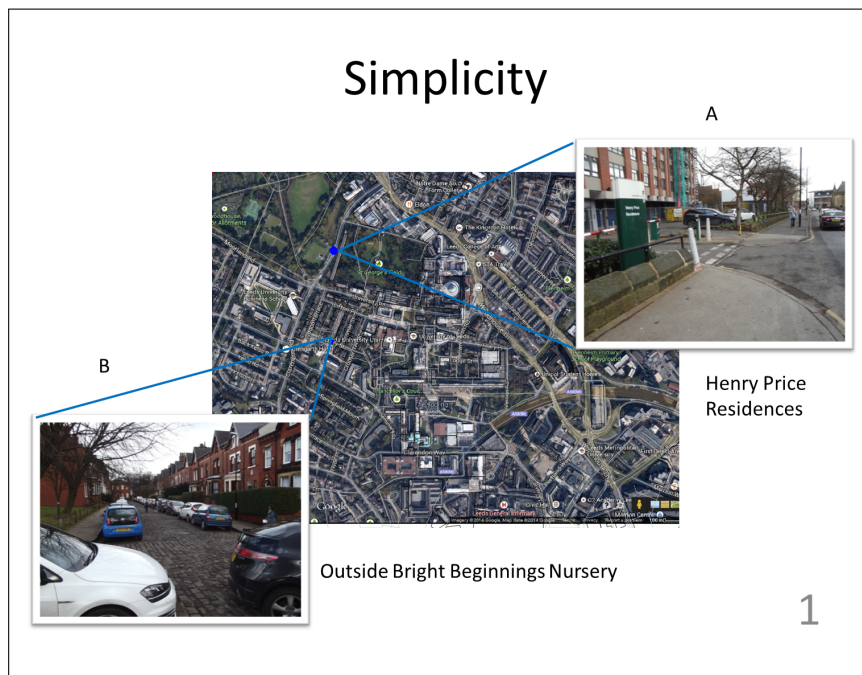
(a) Example screen. Category - simplicity, start point - Henry Price Residences, end point - Outside Bright Beginnings Nursery
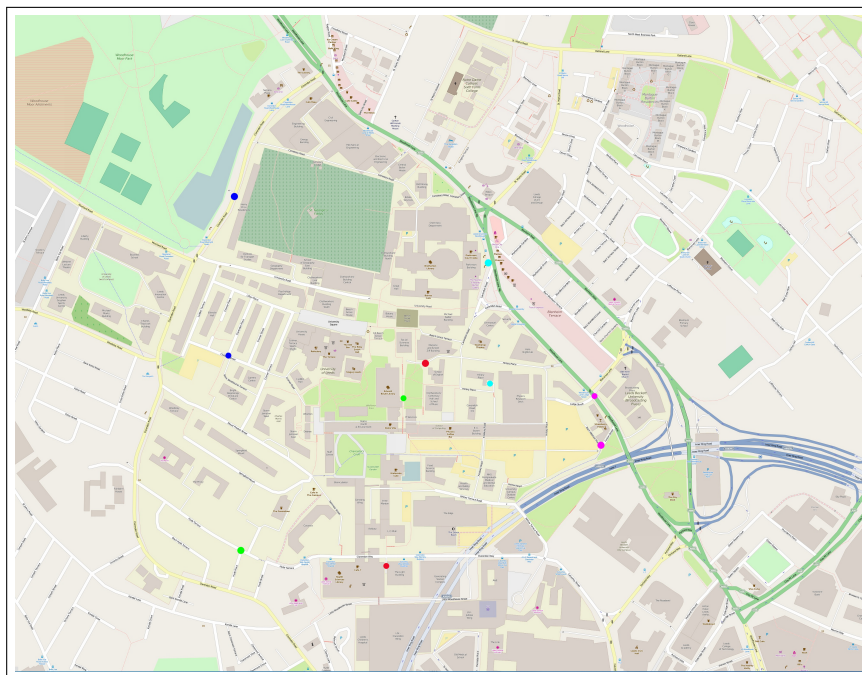
(b) Image used for the A2 laminated map.

Figure 7.2: Examples of the materials provided to the participants to describe the test area and start-end points.

7.1.3 Procedure.

Each participant was provided with a copy of a participant information sheet and an instruction sheet. Once they had read this information, and completed the consent form,

(a) Example of the paper maps provided on which the participants could draw their routes. The red line shown is a participant suggested route.

(b) Part of the questionnaire that was used to gather information on the familiarity of participants with the points and routes, and the reasons used for route choice for each journey type or attribute category.

Figure 7.3: Examples of the materials provided to the participants to elicit the required data.

the audio recorder was started and the experiment began. The questionnaire gave spaces for participants to record their gender and age, but they were instructed to not write their

name on any sheet that they wished to submit. Each participant was also asked to confirm that they were familiar with the University campus before proceeding with any trials.

A total of 20 trials per participant were carried out during the experiment (four journey types or categories, five point pairs), and a randomised block design was used to vary the order of display of both the types of journey and point pairs (i.e. the combined order of journey type and pairs were different for each participant in the evaluation). To familiarise the participant with the format of the experiment, the first trial was used as a practice during which the instructions were explained by the experimenter, and the participants were encouraged to ask questions. During this trial, participants were given the opportunity to change their selected route as many times as they wished, with extra materials being provided if required. They were also asked if they were satisfied with their final selection, and if they were happy for it to be included in their submission. With the practice trial completed, the procedure detailed below was then repeated for the remaining 19 trials.

For each trial, the appropriate screen was displayed and participants were asked one of the following questions, according to the appropriate attribute category or journey type:

Simplicity: 'Imagine that someone has approached you and asked you for the simplest route from point A to point B. Please describe the route you would suggest.'

Everyday: 'Imagine that you were giving instructions to someone who needed to travel from point A to point B every day, for example as part of their daily commute to work. Please describe the route you would suggest.'

Leisure: 'Imagine that you were giving instructions to someone who had some free time and wanted to go for a stroll from point A to point B. Please describe the route you would suggest. You can assume that they have approximately 30 minutes free'

Tourist: 'Imagine that you were giving instructions to a tourist who wanted to go on a tour of the campus sights, starting at point A and ending at point B. Please describe the route you would suggest.'

The locations of the start and end points on the laminated map were then indicated to the participant, and they were informed that they could write, speak or draw their suggestions using the materials provided. In addition they were asked to avoid passing through buildings, to stick to the pathways shown on the map and to avoid repeating any section of route wherever possible. Help in identifying or locating buildings or areas of the map was provided, and where participants chose to give verbal instructions, the experimenter drew the described route on a paper map, and then verified its accuracy

with the participant. No time limit was placed on the completion of each trial, and the participants were allowed to alter their route as many times as they wanted (with more materials being provided if required).

Once a route had been chosen, the participant was then asked to indicate on the questionnaire how familiar they were with the start point, end point and the route that they had chosen. A four-point Likert scale was provided for them to respond to these questions, ranging from 'Very familiar' to 'Not familiar'. In addition, at the end of each journey type, the participants were asked to write a few words on why they had chosen the routes suggested within that section. For tourist journeys, they were also asked to include any point of interest which they had wished their tourist to see.

7.2 Results and Analysis

All participants confirmed that they were familiar with the University campus prior to starting the trials, and that they understood the participant information sheet. All but one participant used the provided paper maps to record their routes exclusively, and this participant chose to switch to this approach after their second trial. Participants took between 29 and 63 minutes (mean 44.2, SD 10.5) to complete the entire 20 trials (including the practice trial), and eight asked for help in locating a building, or for confirmation of a road name. In addition, all gave permission to include test trials as submissions.

The questionnaires and route data were collected together, and examined for issues before the data was analysed. Of the 14 participants, one stated that they would rather 'chill on a bench' than spend their leisure time walking, and one gave 'shortest length' as their reason for choosing tourist routes. Both of these statements indicate that the participant has either misunderstood the task, or refused to complete it as instructed. For these reasons, their route suggestions were excluded for the relevant trials. In addition, one participant indicated on paper that they were very familiar with all points and routes, despite verbally contradicting this. Familiarity data for this participant were therefore also excluded.

Of the remaining route suggestions, a number contained either repeated or invalid (non-existent, passing through buildings, unconnected) links. In the case of repetitions, where possible the routes were used as described by the participant, despite contradicting the participant instructions and any effect that this may have on the resulting similarities (the route suggestion algorithms do not permit repeated links in routes). For all invalid links, the most obvious alternatives were chosen. In all cases, the net result was a valid and complete route from the start to the end point.
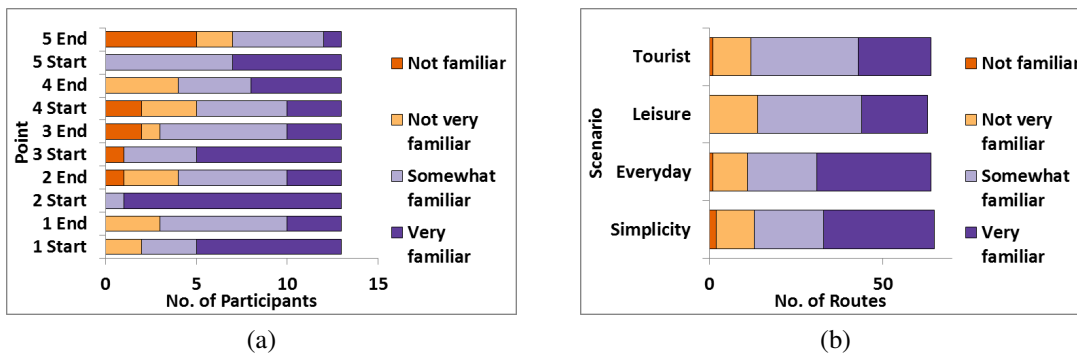
Figure 7.4: Participant familiarity with (a) start and end points, and (b) routes chosen (divided by journey type).

Analysis of the data collected from the participant questionnaire indicates that the majority of points were known to most participants (Figure 7.4a). The end point for Zone 5 is the obvious exception to this, with more participants having little or no familiarity with it. However, other points nearby were known to most, and therefore the levels of familiarity were considered sufficient for participants to be able to suggest suitable routes. In addition, Figure 7.4b indicates that although not all participants were familiar with the routes that they were suggesting, there is a good level of familiarity across all journey types, and overall more than 80% of the routes were at least somewhat familiar. This level of familiarity gives confidence that the participants were suggesting routes that they think are appropriate for the journey described.

The following sections give a more in-depth analysis of the data provided by the participants, and use it to evaluate the algorithm suggested solutions, one journey type at a time. To enable this, the similarity between the algorithm suggested route and those of each participant were calculated, and taken to be a measure of performance. For statistical analysis of these similarities, the Friedman and Wilcoxon's tests were chosen as the data are not independent (the same participants and routes were used to analyse the performance each algorithm), and does not have a normal distribution. Where appropriate, these tests will be used to give the order of performance (or rank) of each of the tested algorithms. As in previous chapters, route similarity was defined as the percentage of routes which overlap:

$$Similarity = Length_{overlap}/Length_{max}$$

where:

$Length_{overlap}$ = total length of links which occur in both routes

$Length_{max}$ = length of the longer of the two routes

(7.1)

Where available (and unless otherwise stated), the algorithms' routes were compared to the most preferred route which is the route chosen by the maximum number of participants (most preferred similarity). Mean and standard deviation were calculated across the differing participant routes with identical suggestions removed, and where appropriate, the similarity between the algorithm suggested route and the nearest matching participant route was given as the maximum similarity. Weighted similarity mean and standard deviation were calculated by including all participants routes (including identical routes), allowing more commonly chosen routes to have more weight. Finally, in this context, consistency refers to how often a route is chosen rather than the similarity of multiple routes.

7.2.1 Tourist Routes

The results of an analysis of the similarity between routes that participants suggested for tourist journeys are shown in Table 7.2. A total of 65 routes were suggested for the 5 point pairs (shown in Figure 7.5), and show very little similarity between the participants (see Table 7.2) which was contrary to expectations.

Zone	Routes Suggested	Similarity Mean	SD
1	13	20%	14%
2	13	20%	13%
3	13	23%	17%
4	13	22%	15%
5	13	19%	13%
Mean	13	21%	15%

Table 7.2: Analysis of the participant tourist route suggestions. Number of routes suggested, and mean and standard deviation similarity between all routes.

The attributes stated by participants to have influence on the choice of tourist routes are shown in Figure 7.6a. As expected, points of interest ranks highest, being mentioned by almost all participants. The lack of reference to other attributes seen here may be a

(a)

(b)

(c)

(d)

(e)
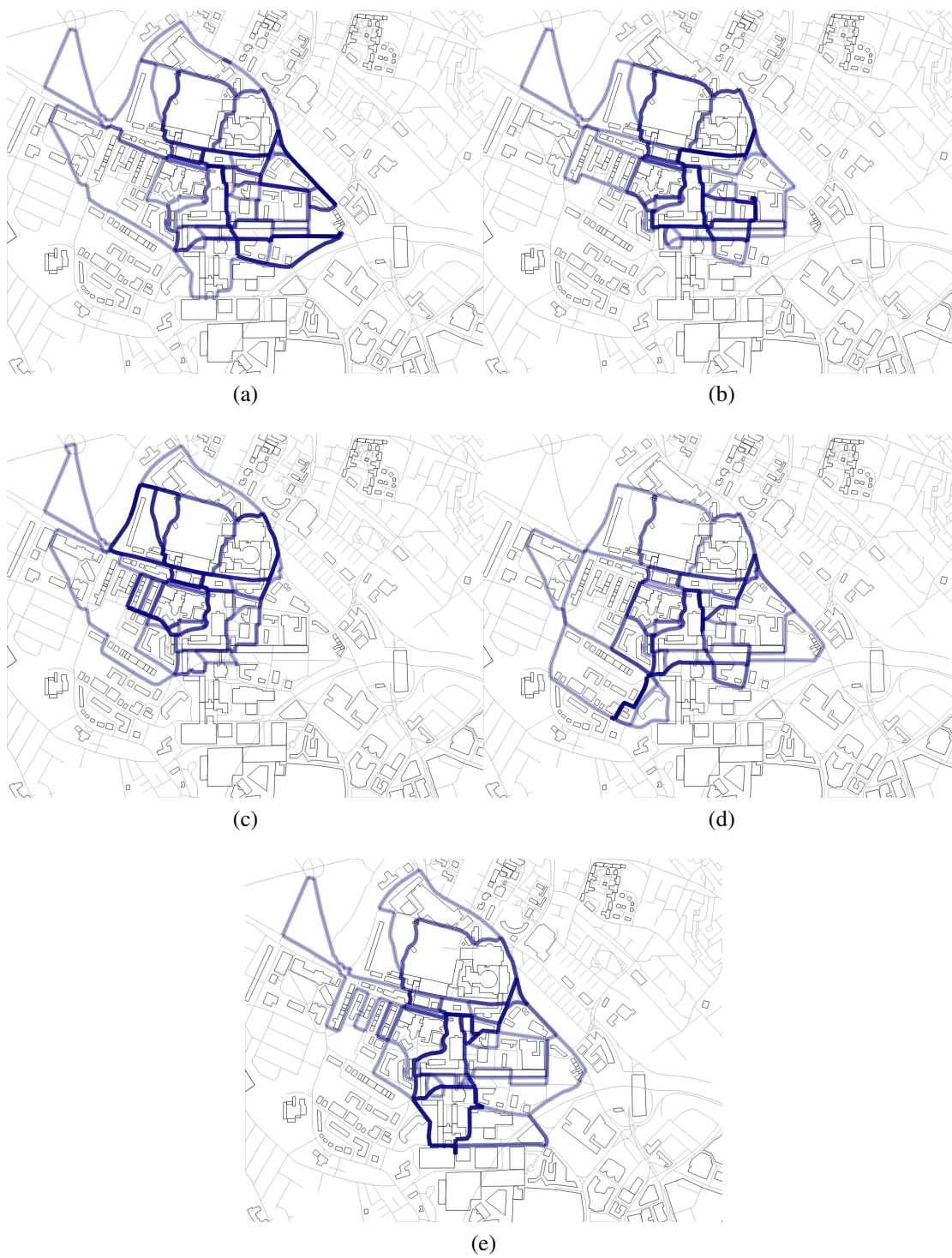
Figure 7.5: Participant suggested tourist routes for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5, where the more opaque the route section the more suggested routes it appears in.

direct response to the way that the task was set. In addition to being asked to give their reasons for choosing the tourist routes, they were asked to provide a list of 'sights' that they wished their tourist to see. In many cases this resulted in just a list of points of interest and no further details, even when participants were encouraged to offer additional information.
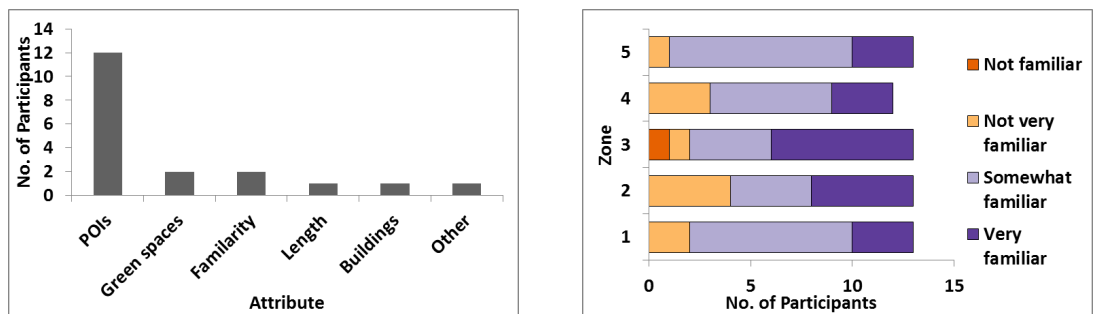
(a) Frequency of attributes stated as influences on tourist route choice.

(b) Familiarity of suggested tourist routes for point pairs in each zone.

Figure 7.6: Analysis of the data from the tourist section of the participant questionnaire.

Comments taken from the audio recordings of the experiments however, suggest that participants are considering more attributes than indicated in the questionnaire. 'Green space', 'pretty gardens' and 'nice building frontages' are all given as reasons for choosing particular links over others, suggesting that parkland, vegetation and possibly dwellings also have an influence on the routes that were selected.

Figure 7.6b and Figure 7.4b indicate that the overall familiarity is lower than expected. Seven participants stated that they were not very familiar with at least one of their suggested routes, and one participant had no familiarity at all with their suggested Zone 3 route. This lack of experience of travelling the chosen routes may offer an explanation for the variation in participant suggestions. However, the number of participants stating that they were very familiar with their suggested routes indicates that familiarity cannot account for all of the differences that were seen.

An initial examination of the participants' stated lists of points of interest indicated two important issues. Firstly, the list of points of interest supplied by the University was not comparable to those stated by the participants. Eleven of the 20 University points of interest were not mentioned by any participants, and a further six previously 'unknown' points of interest were identified. Given that the differences between these two sets of 'sights' would be expected to produce substantially different routes, this is likely to affect both route similarity and analysis of participant data. To remove this potential bias, the list of participant suggested points of interest was collated and used to construct an alternative

user map.

(a)   (b)

Figure 7.7: Screenshot showing the annotated points of interest levels for (a) the original map and (b) the user map.

Figure 7.7 shows the differences in the levels of points of interest in the original map (Figure 7.7a) and the user map (Figure 7.7b). The comparison of the two maps in Figure 7.8 shows that the points of interest percentage frequencies for the participant tourist routes (Figure 7.8a) is higher for the user map (mean of 46% POIs per route) than the original map (mean of 40% POIs per route), as expected. Although less pronounced, this increase is also seen for the participant leisure routes (which will be fully discussed later) in Figure 7.8b (original mean of 29% and user mean of 34%) .

(a)   (b)
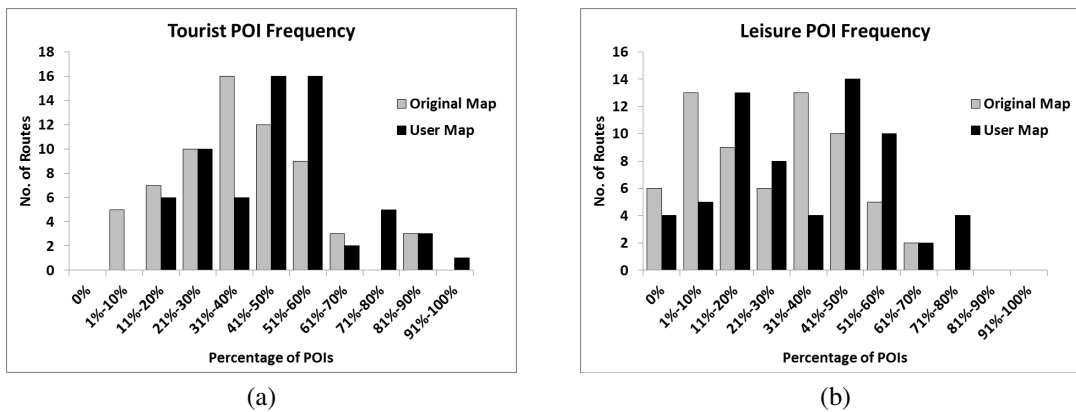
Figure 7.8: Comparison of the frequency of percentages of POIs for the participant routes on the original map and the user map, for (a) tourist routes and (b) leisure routes.

Secondly, although the points of interest also varied between participants (as suggested by previous research [2, 66, 206, amongst many others]), not all suggested routes

pass those indicated as important by either the participants or the university. This is contrary to expectations, and may indicate that the participants have not fully understood the task that they were asked to complete. Alternatively this may indicate that they are using different attributes to select routes, or lend weight to the idea that humans are willing to settle for good routes rather than seeking the 'best' route available. Examples of the latter are where a route which contains only five points of interest is chosen over a much longer alternative with 15, or only one of each type of 'sight' is included (one statue, one building, etc).

As mentioned previously, there are many possible explanations to the wide variation in route suggestions for tourist journeys, but one possibility is that all of these routes share similar characteristics. For example, this could occur where two routes pass on different sides of a point of interest, or where two very different routes pass through areas of high density vegetation such as gardens. To examine this possibility, an analysis of the user suggested tourist routes in terms of the individual attributes is shown in Figure 7.9 and Figure 7.10. To provide a point of reference the data for the shortest route algorithm (C_{LEN}) are also shown, although the conclusions that can be drawn are limited due to the small number of samples.

All seven attribute plots show high levels of variation, disproving the theory of differing routes displaying similar characteristics. Figure 7.9a shows that participants selected routes of different lengths, with only three suggesting routes which are at least as long as the anticipated 2km (see Section 6.4 for details of this). However the majority are substantially longer than those selected by the shortest path algorithm (C_{LEN}), as anticipated by the results of Experiment 5 (Section 3.8) Of all the participants, participant 13 suggested by far the longest routes (up to 3.1km compared with 2.2km for the other participants).

As with length, an increase in decision points (Figure 7.9b) and turns (Figure 7.9c) is observed within participant routes compared to the shortest routes. This indicates that minimising these attributes is unlikely to play a part in the selection of tourist routes, as predicted by Hypothesis H7 and the results of Experiment 3 (Section 3.6).

Figure 7.10a shows a relationship less clearly than the three attributes discussed so far. One reason for this is the high levels of vegetation encountered by the minimum length routes, and another is the within participant variation. However closer investigation reveals that eight participants (2, 3, 4, 9, 10, 11, 12 and 13) all have mean vegetation proportions higher than that of C_{LEN}. This may indicate that vegetation plays a part in selecting for tourist journeys, as suggested by Hypothesis H7 and the results of Experiment 3, but that it is unlikely to be the most influential attribute (it is ranked only equal second by Experiment 6, Section 3.9).

(a)

(b)

(c)
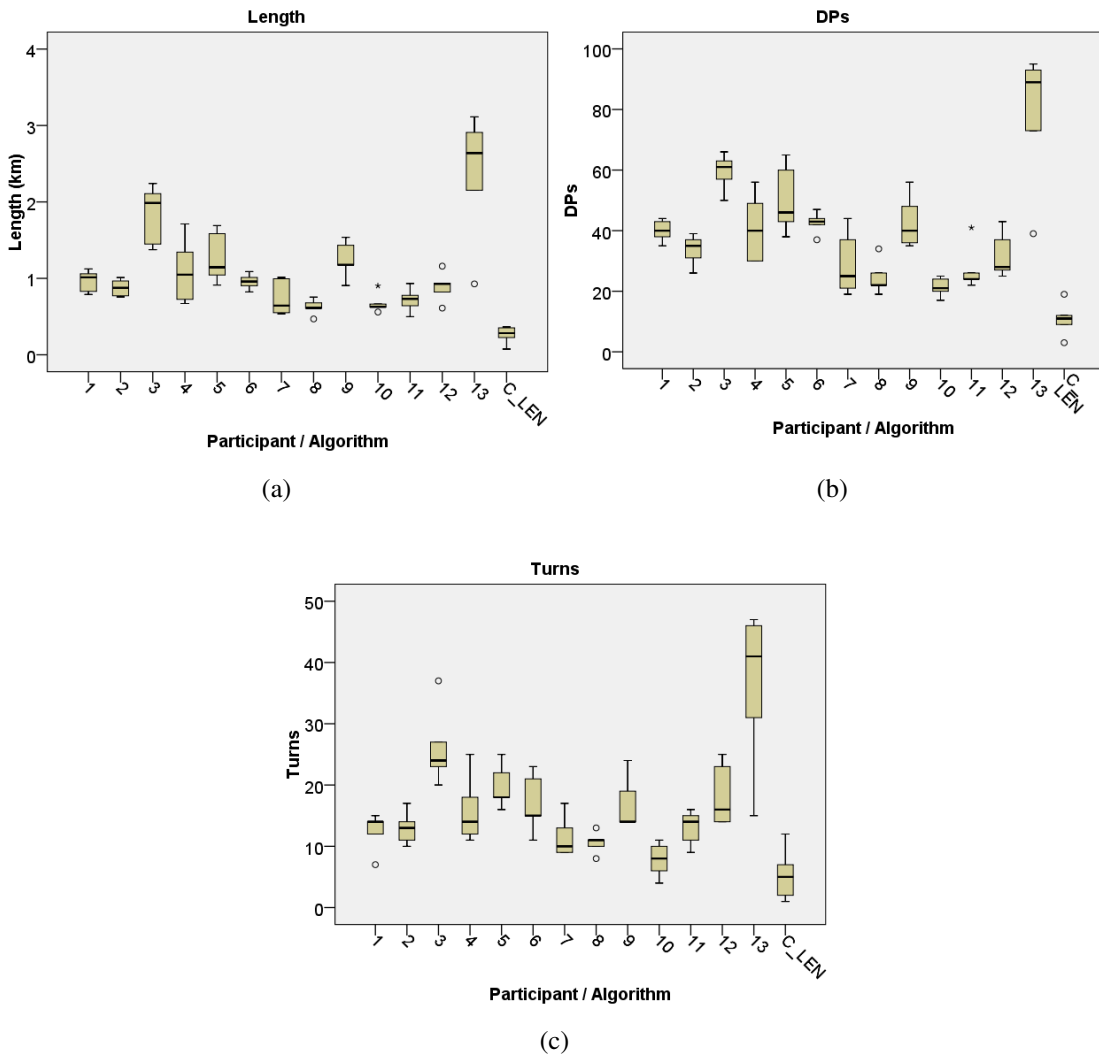
Figure 7.9: Individual attribute analysis of the participant suggested tourist routes for start-end pairs showing (a) length, (b) DPs and (c) turns. Boxes indicate 25% to 75% and whiskers indicate the $\pm$ 1.5*interquartile range.

The proportion of parkland shown in Figure 7.10b also shows some variation, but over half of the participant suggestions have a higher mean value than the minimum length routes generated by C_{LEN}. The substantially higher proportion illustrated for participant 13 indicates that parkland is likely to play a large part in their selection for this type of journey. This theory is also backed up by comments during the relevant trials. Although not as pronounced, the increases shown by participants 3, 4, 5 and 9 are also likely to show an inclination towards the intentional inclusion of this attribute for tourist journeys. This appears to confirm part of the theory proposed by Hypothesis H7 and the findings of Experiment 3.

Figure 7.10: Individual attribute analysis of the participant suggested tourist routes for start-end pairs showing (a) vegetation, (b) land use, (c) dwellings and (d) POIs. Boxes indicate 25% to 75% and whiskers indicate the ± 1.5*interquartile range.

For dwellings (Figure 7.10c) and points of interest (Figure 7.10d), the relationships are somewhat clearer. In both cases, the participant means are all higher than those from the routes produced by C_{LEN}. This is likely to indicate that they both play a part in the selection of tourist routes, as suggested by Hypothesis H7 and Experiment 3 (Section 3.6). Although no participant included all points of interest in their suggested routes, the relatively high numbers that are encountered may lend weight to Hypothesis H10.

Looking at the participants individually also provides a possible explanation for the inconsistencies seen between their suggested routes. The best illustration of this is a comparison of the attribute values of participant 10 and participant 13. Participant 10 has

suggested relatively short routes with fewer turns and points of interest, little parkland, but a larger proportion of dwellings than shown by the other participants. Little explanation was given as to why these routes were suggested, but it may indicate that length and dwellings have more influence on route choice for this participant. In contrast participant 13 suggested much longer routes, with far higher levels of parkland and many points of interest, likely suggesting that these attributes were more influential. Smaller variations are seen across the other participants, which is may indicate that different participants are using differing attributes to select routes appropriate for tourist travel, and that a single fixed algorithm may not produce the 'best' routes for all users.

7.2.2 'Simplest' Routes

The results of an analysis of the similarity between routes that participants suggested for the simplicity category are shown in Table 7.3. Unlike the routes suggested for tourist journeys, the majority of participants suggested the same routes for all but two point pairs (see Table 7.3 and Figure 7.11). The start-end pair in Zone 3 gave an even split of participants between two routes (two participants suggested a third possibility) with the majority of participants indicating that two of these routes were thought to be equivalent. This is unsurprising as the two most suggested routes vary in length by only 2.7m, although one has fewer turns than the other. In addition, although the seven participants suggested a single route for pair 4, the mean similarity for the routes chosen by the participants as a whole is low. This is due to a small number of participants who considered routes with substantial differences to be more appropriate.

Zone	Routes Suggested	Participants for Most Preferred	Similarity	
			Mean	**SD**
1	1	14	100%	N/A
2	1	14	100%	N/A
3	3	6	41%	16%
4	5	7	19%	25%
5	6	8	40%	36%
Mean	3.2	9.8	44%	38%

Table 7.3: Analysis of the participant simplicity route suggestions. Number of routes suggested, number of participants for the most preferred route, and mean and standard deviation similarity between most preferred and remaining routes.

Despite having high consistency for simplicity route choice, participants gave a variety of different reasons for their choices, and Figure 7.12a shows the attributes associated

Figure 7.11: Participant suggested simplicity routes for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5. The more opaque the route section the more suggested routes it appears in, and the most preferred route is shown in red.

with these reasons. In addition to the simplicity attributes anticipated to have influence in Section 3.4 (length, turns and decision points), well defined landmarks (which were discussed in Section 2.1.1), familiarity and routes which could be easily converted into simple directions were all mentioned.

Further discussions with participants after the end of the experiment provided one possible reason for the route choice variation. A small number of participants stated that a lack of familiarity with either the start or end point would lead to known routes being used as far as possible, even if that increased the complexity. The start and end points for Zone 4 and Zone 5 have a lower combined familiarity than the remaining start-end pairs (see Figure 7.4a), suggesting that this may also play a part in the low mean similarities observed. Figure 7.12b also indicates that the participants were less familiar with the routes that suggested for the start-end pairs in Zone 4 and Zone 5, meaning that they may have insufficient knowledge about alternative routes to be able to choose the simplest.
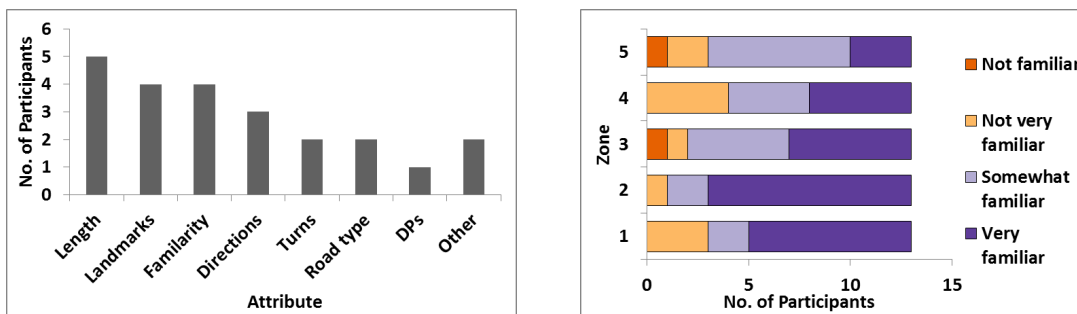
(a) Frequency of attributes stated as influences on simplicity route choice.

(b) Familiarity of suggested simplicity routes for point pairs in each zone.

Figure 7.12: Analysis of the data from the simplicity section of the participant questionnaire.

One surprising observation is that not all participants gave length as a reason for simplicity route choice. Previous research has indicated that length is an important factor in perceived route simplicity, and it was expected that all participants would include this in their list of reasons. Furthermore the participant routes for the Zone 4 and Zone 5 point pairs show substantial variations in length, with some being up to twice as long as others.

An analysis of the user suggested simplicity routes, in terms of the individual attributes, is shown in Figure 7.13 and Figure 7.14. To provide a point of reference the data for the shortest route algorithm (C_{LEN}) are also shown, although the conclusions that can be drawn are limited due to the small number of samples. From an initial examination of these plots, one very obvious observation can be made - the routes suggested by participant 12 display substantial differences to those of the other participants. They are longer,
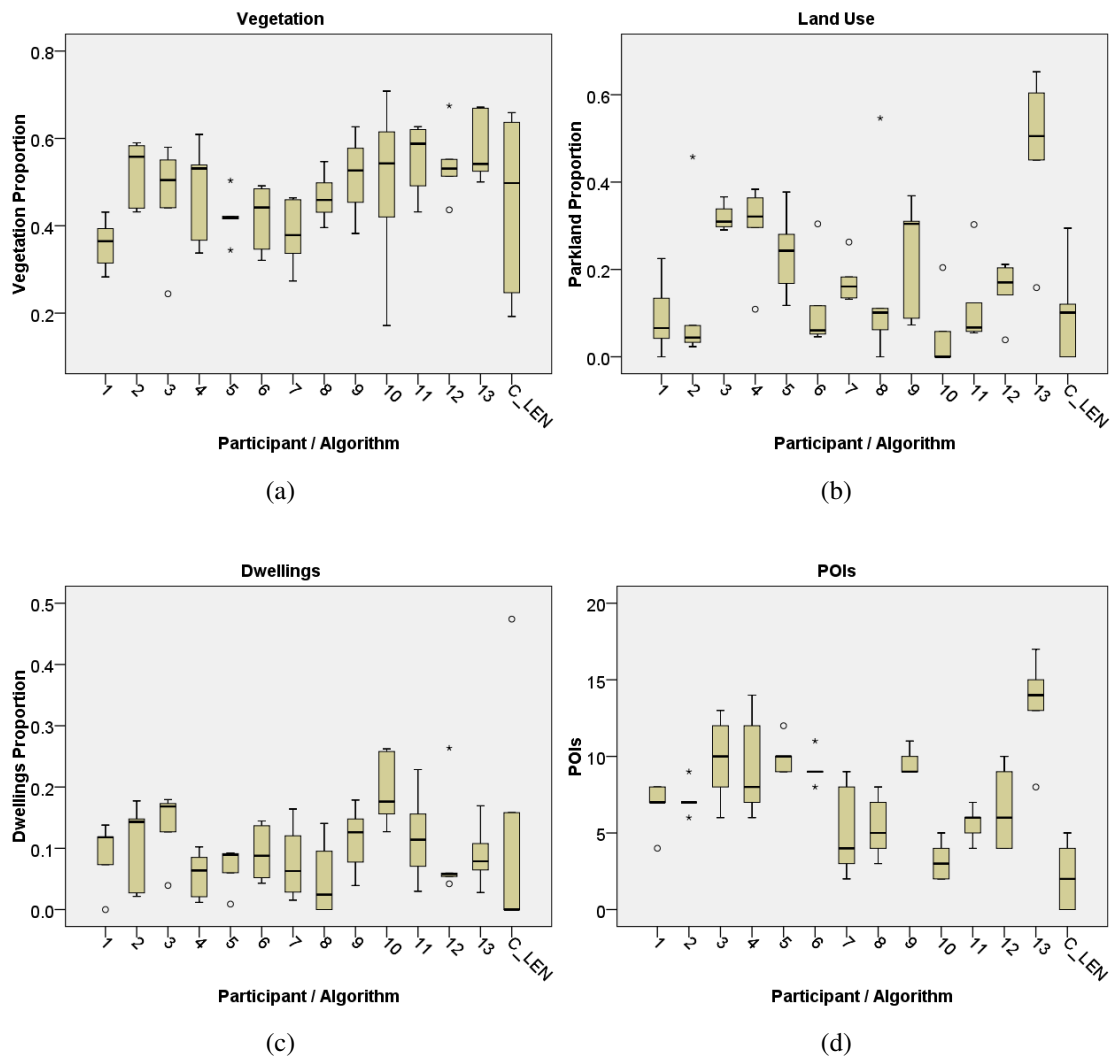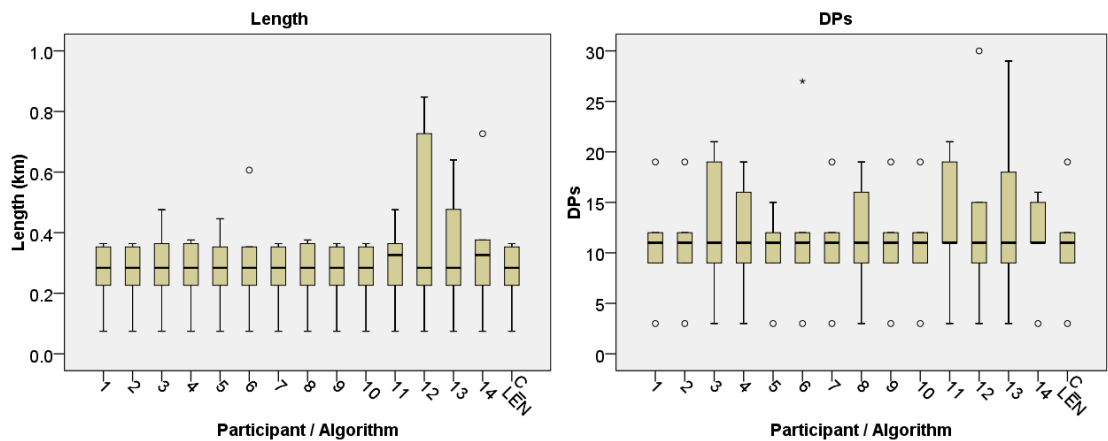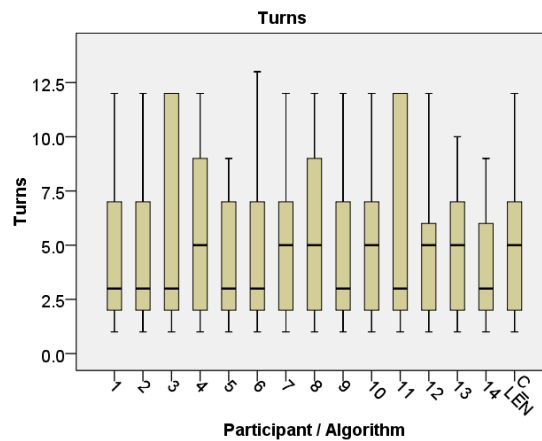
Figure 7.13: Individual attribute analysis of the participant suggested simplicity routes for start-end pairs showing (a) length, (b) DPs and (c) turns. Boxes indicate 25% to 75% and whiskers indicate the $\pm$ 1.5*interquartile range.

and contain a higher proportion of dwellings than the other suggested routes. This participant also stated they they had less familiarity with the test area than the other participants, and comments from this trial suggests that the participant used clearly recognisable buildings to guide their route choice. Although a corresponding increase in points of interest is not seen in Figure 7.14d, several of these building were included in the points of interest given by the University. These variations in route are contrary the results of earlier research [98, 182, amongst many others] which indicates that length is the most important attribute for simplicity, but it does correspond to the results of Experiment 2 (Section 3.5) which suggested that landmarks are ranked highest of the tested attributes for selecting

simple routes.
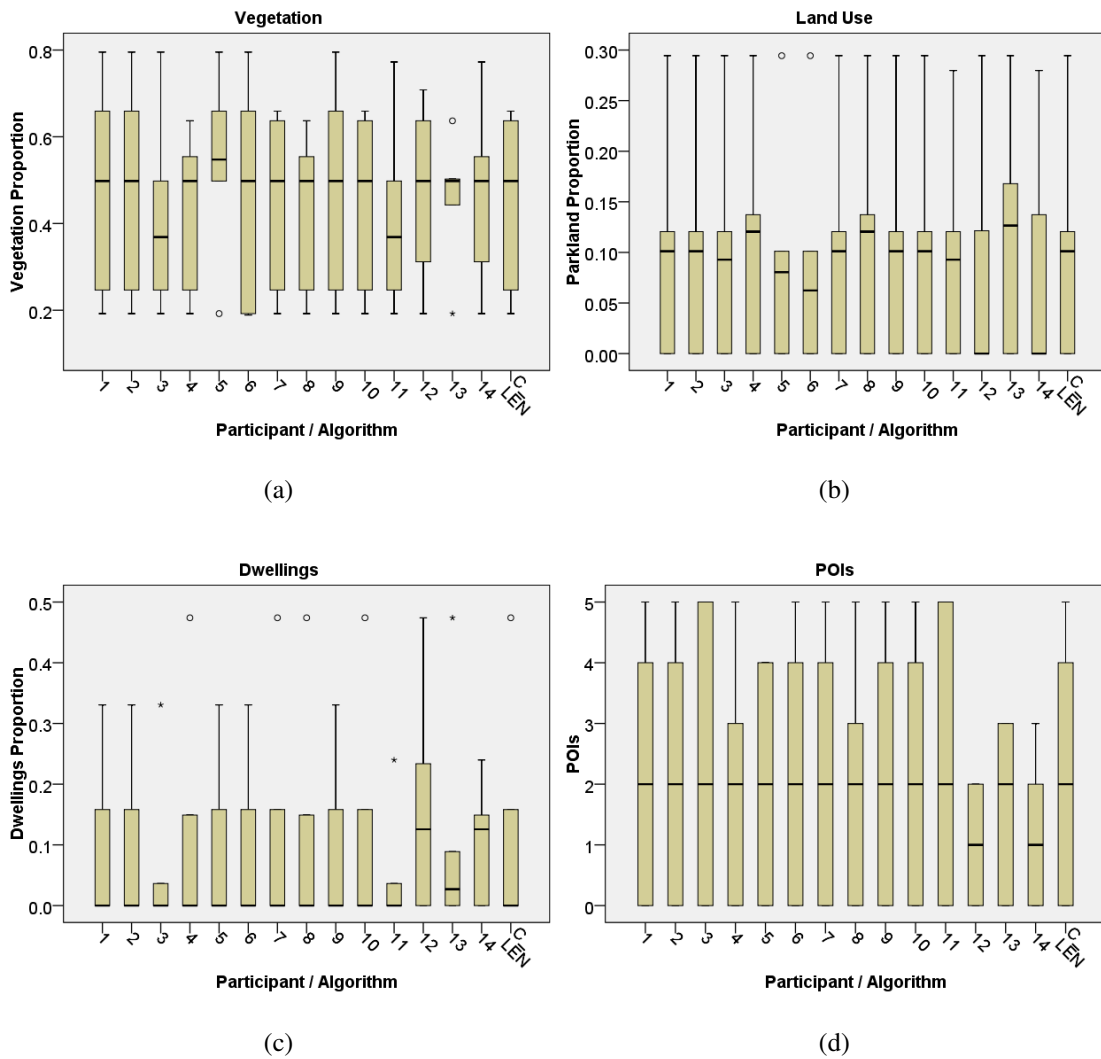
Figure 7.14: Individual attribute analysis of the participant suggested simplicity routes for start-end pairs showing (a) vegetation, (b) land use, (c) dwellings and (d) POIs. Boxes indicate 25% to 75% and whiskers indicate the $\pm$ 1.5*interquartile range.

The routes suggested by participant 13 also show a slight increase in length, and a substantial increase in decision points. In addition they have a slight decrease in points of interest and dwellings, although there is nothing within their comments or stated reasoning to indicate why this would be the case. Participant 5 also shows a variation in the amount of vegetation which was present on the 'simple' route which they suggested. Although this may indicate that vegetation played a small part in the selection of these routes, there was again no explanation in the comments or reasoning that they gave.

Other than the differences stated here, the routes suggested as being the simplest be-

tween to two stated points not only show similar characteristics between the participants, but also to those selected by the shortest path algorithm (C_{LEN})

7.2.3 Everyday Routes

A total of 14 everyday routes were suggested for the 5 point pairs, as shown in Figure 7.15 (a-e) and Table 7.4, with the majority of participants suggesting the same routes for all but two start/end pairs - Zone 3, and Zone 4. For Zone 3 there was an even split of participants between two routes, as with simplicity. In contrast for Zone 4 seven participants suggested one route, five another, and the two remaining participants ech suggested a different route entirely. For comparison purposes the Zone 3 route with more vegetation was chosen as the most preferred, as this attribute ranks second in importance (the length of the routes, which was the highest ranked attribute, is almost identical). These results indicate that there is good consistency (the same route is chosen by the majority of participants) for the all but one (Zone 3) of the suggested routes, but the longer Zone 4 and Zone 5 routes show more variation (the number of routes chosen was higher) between participants.

Zone	Routes Suggested	Participants for Most Preferred	Similarity	
			Mean	**SD**
1	1	14	N/A	N/A
2	2	13	25%	0%
3	2	7	52%	0%
4	4	7	49%	37%
5	5	10	23%	14%
Mean	2.8	10.2	49%	35%

Table 7.4: Analysis of the participant everday route suggestions. Number of routes suggested, number of participants for the most preferred route, and mean and standard deviation similarity between most preferred and remaining routes.

Analysis of the reasons given for selecting these routes (shown in Figure 7.16a) indicates that, in addition to the anticipated inclusion of length, participants were also considering other aspects including familiarity, traffic and footfall. This may be an answer as to why a small minority of participants suggested such different routes. Notably, none of the participants mentioned that points of interest, vegetation or turns played a part in choosing these everyday routes; however, one included decision points in their reasoning. One explanation for this may be the format of the experiment - length is far easier to determine for the maps provided than any of the other anticipated attributes. Alternatively, the omission of points of interest, vegetation and turns may indicate that these three attributes are
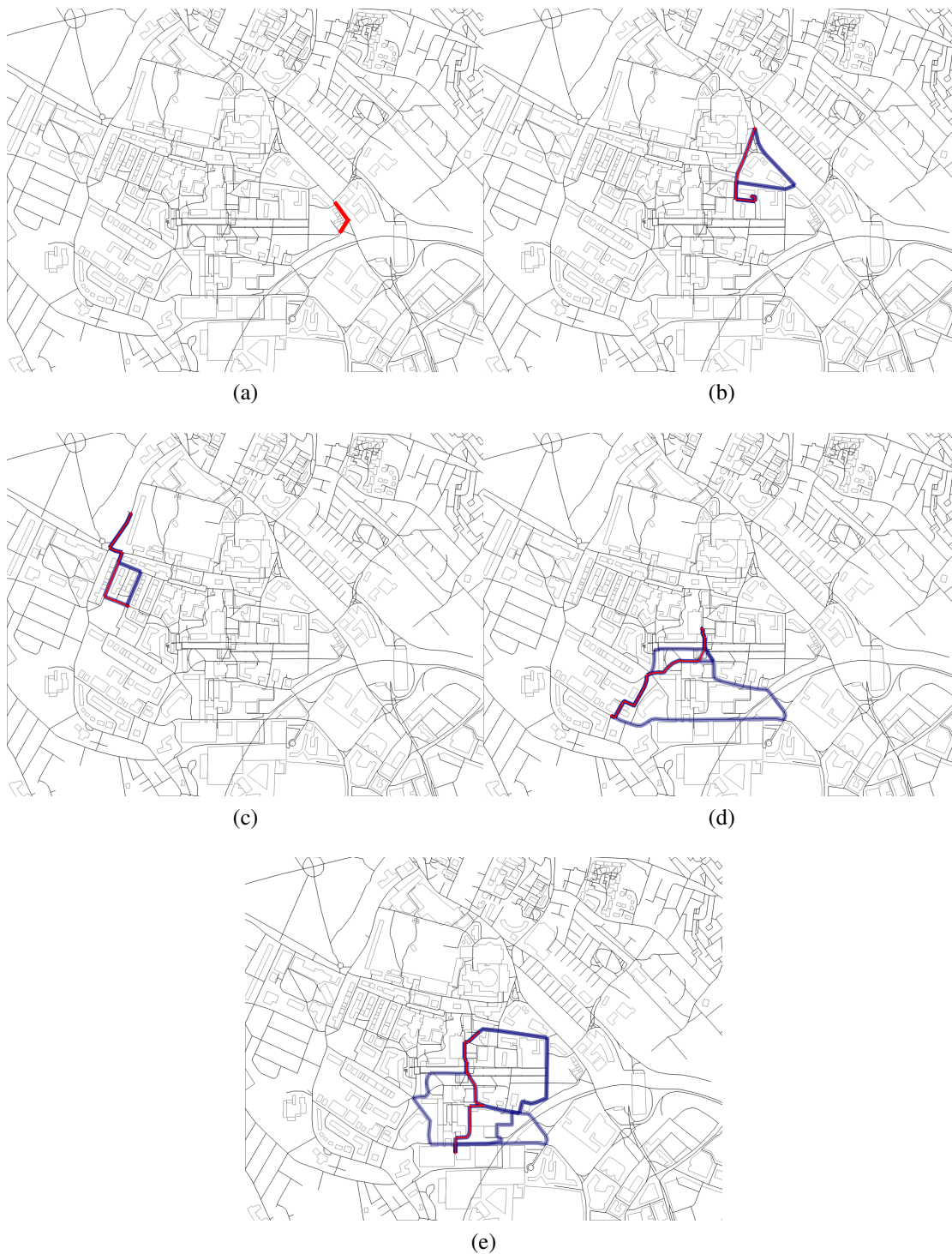
Figure 7.15: Participant suggested everyday routes for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5. The more opaque the route section the more suggested routes it appears in, and the most preferred route is shown in red.

not as important as suggested by the user study detailed in Section 3.7, or simply that they are outweighed by other attributes which were not tested. An example of this is the two routes suggested by participants for Zone 3. One route has higher vegetation coverage and fewer turns than the other, and it is possible that these differences are preferred by 7/14 participants, or that they are just choosing randomly between the two alternatives.

(a) Frequency of attributes stated as influences on everyday route choice.

(b) Familiarity of suggested everyday routes for point pairs in each zone.
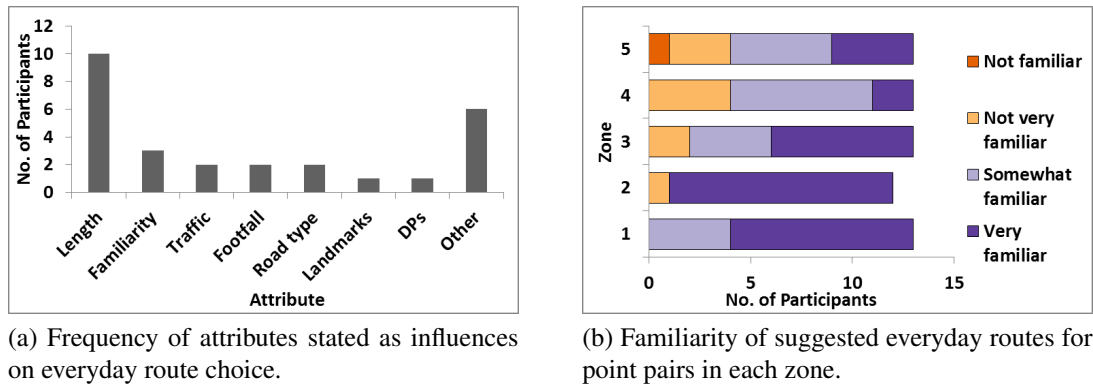
Figure 7.16: Analysis of the data from the everyday section of the participant questionnaire.

Figure 7.16b shows the familiarity of the participants with the everyday routes that they suggested, and indicates that as with simplicity, familiarity may also offer an explanation for the low mean similarity value observed for the Zone 5 participant routes. In addition, although the Zone 2 routes (Figure 7.16b) and points (Figure 7.4a) were relatively familiar to the participants, the single participant who suggested an alternative route for this start-end pair was not familiar with either their route or the end point. Analysis of the audio recording for this participant indicated that they therefore chose a route which followed main roads, rather than considering any of the expected attributes.

An analysis of the user suggested everyday routes, in terms of the individual attributes, is shown in Figure 7.17 and Figure 7.18. To provide a point of reference the data for the shortest route algorithm (C_{LEN}) are again shown, although the conclusions that can be drawn are limited due to the small number of samples. As with the simplicity routes, the routes suggested by participant 12 for everyday journeys display substantial differences to those of the other participants. They are longer, contain more decision points, turns and dwellings, fewer points of interest and lower percentages of parkland and vegetation. Comments made during the trials suggested that the participant was relying almost entirely on areas with which they had some familiarity, implying that these attribute variations were unlikely to have been conscious decisions. However, they may have played a part in deciding which routes the user chose to explore during their limited contact with

the test area, and can therefore not be ruled out.

(a)

(b)

(c)
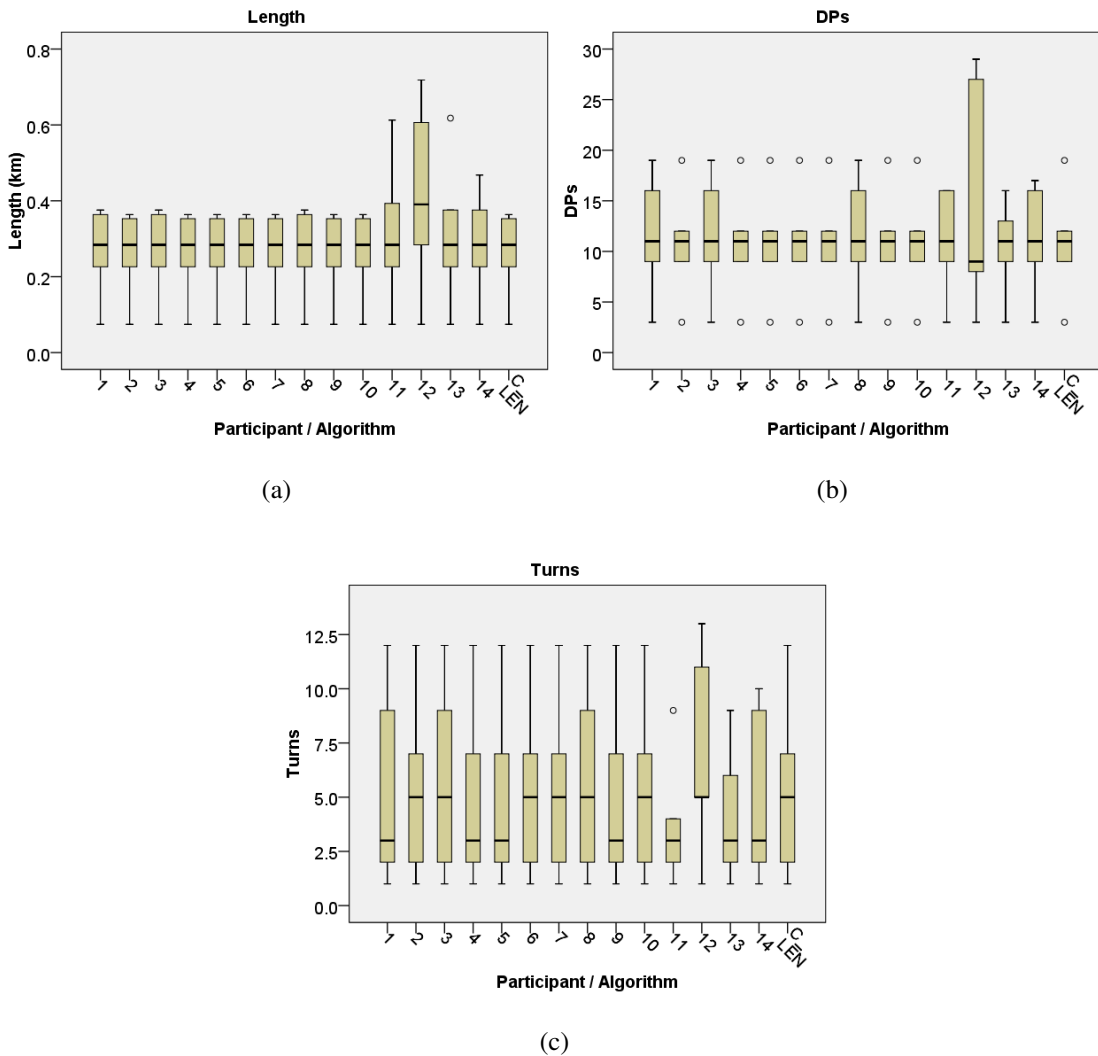
Figure 7.17: Individual attribute analysis of the participant suggested everyday routes for start-end pairs showing (a) length, (b) DPs and (c) turns. Boxes indicate 25% to 75% and whiskers indicate the $\pm$ 1.5*interquartile range.

Of the other 13 participants, participants 11, 13 and 14 show the largest variations. Participant 11 appears to have chosen routes which have a slight increase in length, in order to reduce the number of turns required, and increase the amount of vegetation encountered. Although this is contrary to the results of previous research and Hypothesis H11, the increase in length is small. It does however suggest that the participant was considering the three highest ranking attributes found in Experiment 6 (length, vegetation and turns). The lack of increase in points of interest (ranked equal with turns in Experiment 6) may indicate that this attribute is not actually of equal importance, or that the inclusion

of 'sights' may have been at the expense of one of the others. Although less distinct, a similar pattern is seen in the routes selected by participant 13.

Figure 7.18: Individual attribute analysis of the participant suggested everyday routes for start-end pairs showing (a) vegetation, (b) land use, (c) dwellings and (d) POIs. Boxes indicate 25% to 75% and whiskers indicate the $\pm$ 1.5*interquartile range.

In contrast, participant 14 appears to have selected at least one route which is slightly longer, and contains more decision points and turns, in order to include higher levels of vegetation and parkland. Although land use is ranked low according to Experiment 6, comments made by this participant suggest that this participant was including 'green spaces' in their everyday routes intentionally.

As would be expected given the number of identical routes suggested for this type of journey, the remaining participants show very similar characteristics. In addition the

similarity to the characteristics displayed by the shortest route algorithm would indicate that this would be likely to select routes appropriate for everyday journeys.

7.2.4 Leisure Routes

The results of the analysis of the routes suggested by 13 of the participants (one participant was excluded as discussed earlier) for leisure journeys are shown in Table 7.5. As with tourist travel, the routes chosen for leisure journeys (shown in Figure 7.19) have very little consistency. With the exception of the Zone 2 point pair, each participant suggested a different route between the start and end points for all pairs, and 48 route comparisons showed no similarity at all. There may be many explanations for these variations, from issues with understanding the task, to inconsistency of the attributes being considered by different participants.

Zone	Routes Suggested	Similarity	
		Mean	SD
1	13	13%	15%
2	12	17%	19%
3	13	17%	16%
4	13	16%	14%
5	13	17%	13%
Mean	12.8	16%	15%

Table 7.5: Analysis of the participant leisure route suggestions. Number of routes suggested, and mean and standard deviation similarity between all routes.

Figure 7.20a shows the attributes that were stated as being influential on route choice for the leisure travel. If green spaces are taken to mean parkland and vegetation as examples given by participants indicate, then all of the anticipated attractiveness attributes (vegetation, land use, dwellings and points of interest) and length were mentioned as reasons for route choice, although not in as high proportions as would be expected. Although six participants gave a list of points of interest they wished their subject to visit, only one participant mentioned all four attractiveness attributes, and none included length with this combination. In addition, facilities such as coffee shops and benches featured in the reasoning, as did familiarity (or lack of it).

The overall level of participant familiarity with the leisure routes, shown in Figure 7.20b and Figure 7.4b, is lower than for routes suggested in either the simplicity or everyday scenarios, giving a possible explanation of why the routes for the leisure scenario are so varied. Although a small number of participants were very familiar with the majority
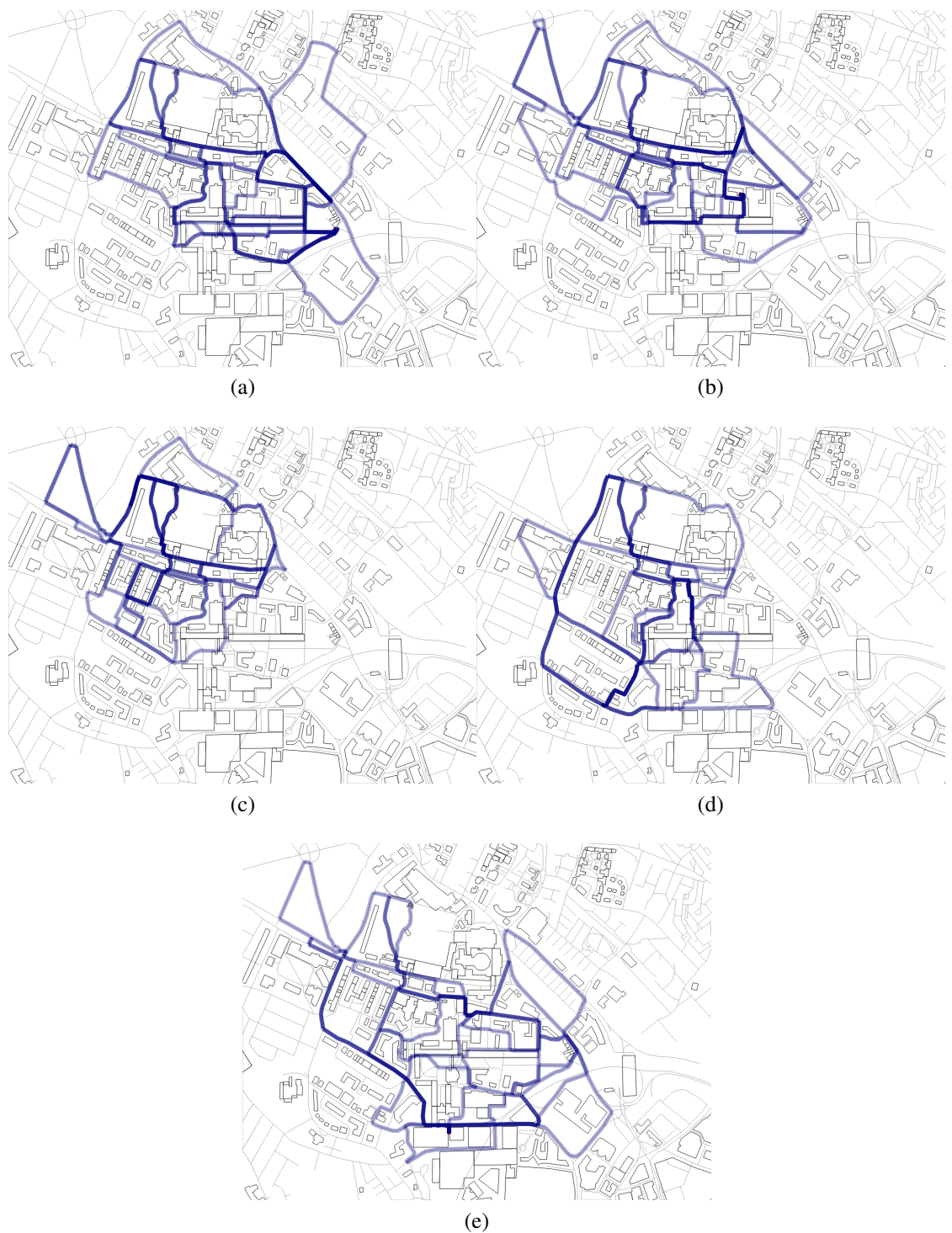
Figure 7.19: Participant suggested leisure routes for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5, where the more opaque the route section the more suggested routes it appears in.

of the routes they suggested (two participants were very familiar with four of their five routes), eight participants had little or no familiarity with at least one of their suggested routes. This lack of knowledge about the route that they recommend may indicate that participants are not suggesting the most appropriate routes in all cases.

(a) Frequency of attributes stated as influences on leisure route choice.

(b) Familiarity of suggested leisure routes for point pairs in each zone.
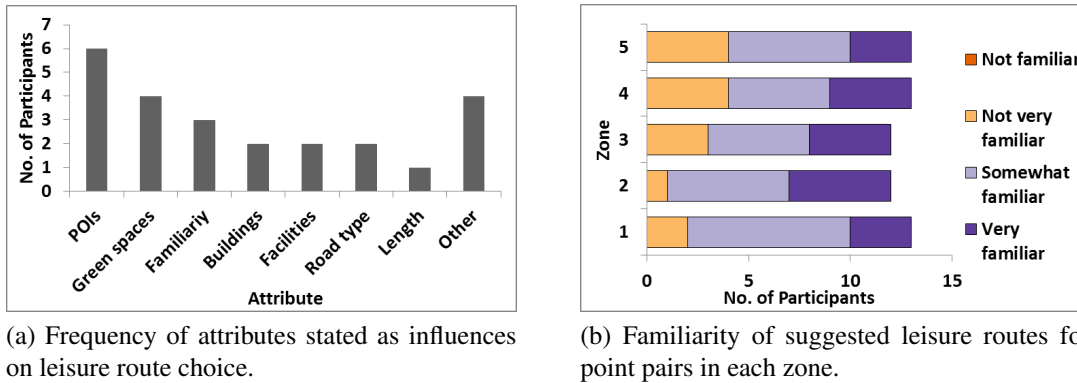
Figure 7.20: Analysis of the data from the leisure section of the participant questionnaire.

Again it is possible that the varying routes suggested by participants may actually have similar characteristics in terms of the expected attributes. An analysis of the user suggested leisure routes, in terms of the individual attributes, is shown in Figure 7.21 and Figure 7.22. To provide a point of reference the data for the shortest route algorithm (C_{LEN}) is again also shown, although the conclusions that can be drawn are limited due to the small number of samples.

As with the tourist journey type, the participants' leisure routes also show varying characteristics. Figure 7.22a again indicates that participants chose routes with different lengths, although none of those suggested reached the anticipated 2km threshold (see Section 6.4 for details of this). Despite this the increase compared to the routes selected by C_{LEN} appears to confirm Hypothesis H12 (Section 2.1) and the results of Experiment 5 (Section 3.8).

The increases also shown in decision points (Figure 7.21b) and turns (Figure 7.21c) suggest that these attributes have little or no detrimental effect on the selection of routes for leisure journeys as proposed in Hypothesis H6. Comparison with the values observed for tourist routes does suggest that these attributes may be slightly more influential for leisure routes, which may confirm that these should be included as very low ranking contributors as indicated by the results of Experiment 6 (Section 3.9).

A wide variation is again seen in the proportions of vegetation and land use shown in Figure 7.22a and Figure 7.22b, both between and within participants. This leads to contradictory conclusions. Substantial increases in vegetation are shown for participants
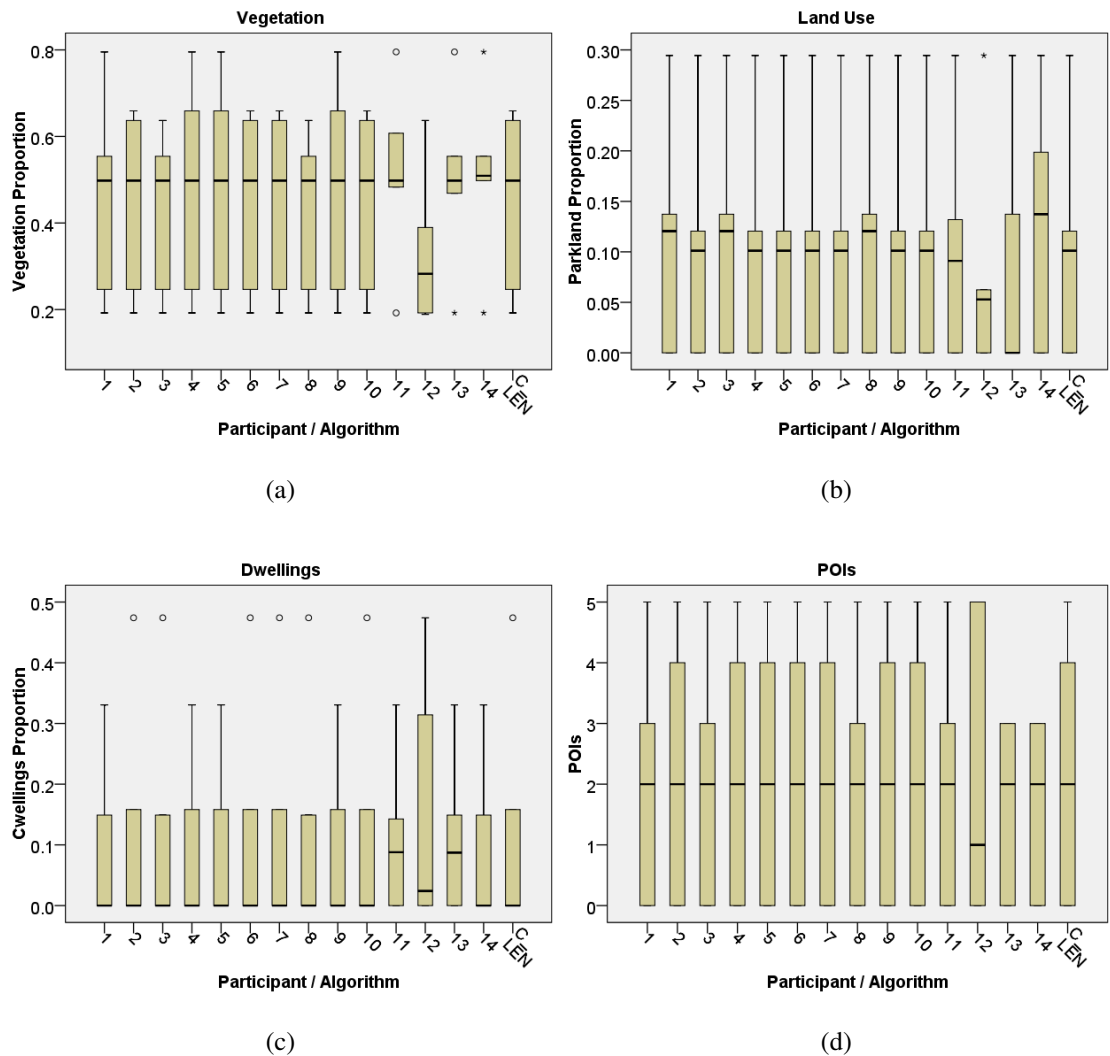
(a)

(b)

(c)
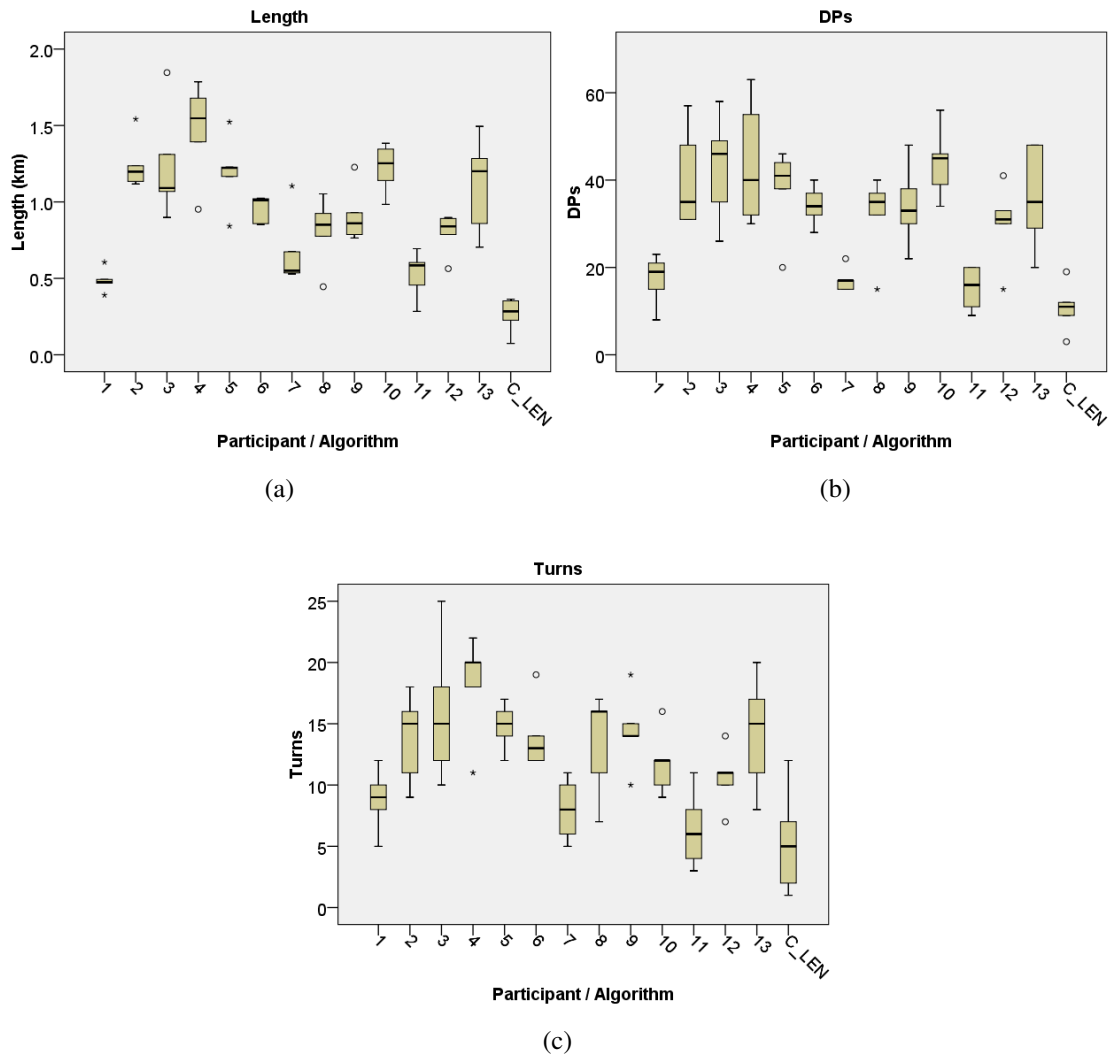
Figure 7.21: Individual attribute analysis of the participant suggested leisure routes for start-end pairs showing (a) length, (b) DPs and (c) turns. Boxes indicate 25% to 75% and whiskers indicate the $\pm$ 1.5*interquartile range.

2, 4 and 5, compared to the minimum length routes (C_{LEN}), meaning that it is likely that these participants are likely to be influenced by this attribute with choosing leisure routes (as indicated by Hypothesis H6 and Experiment 3, Section 3.6). In contrast the lower vegetation means for participants 9, 11, 12 and 13 suggest that they are unlikely to be considering vegetation when selecting leisure routes. The higher means shown by participants 2, 4, 5, 8 and 10 do suggest that they do consider parkland to be important for leisure journeys (as anticipated by Hypothesis H6, Section 3.6) and Experiment 3), whereas the low values shown by participants 7, 9, 10, 11 and 12 probably indicate that they do not. Overall, this analysis appears to contradict Hypothesis H9.
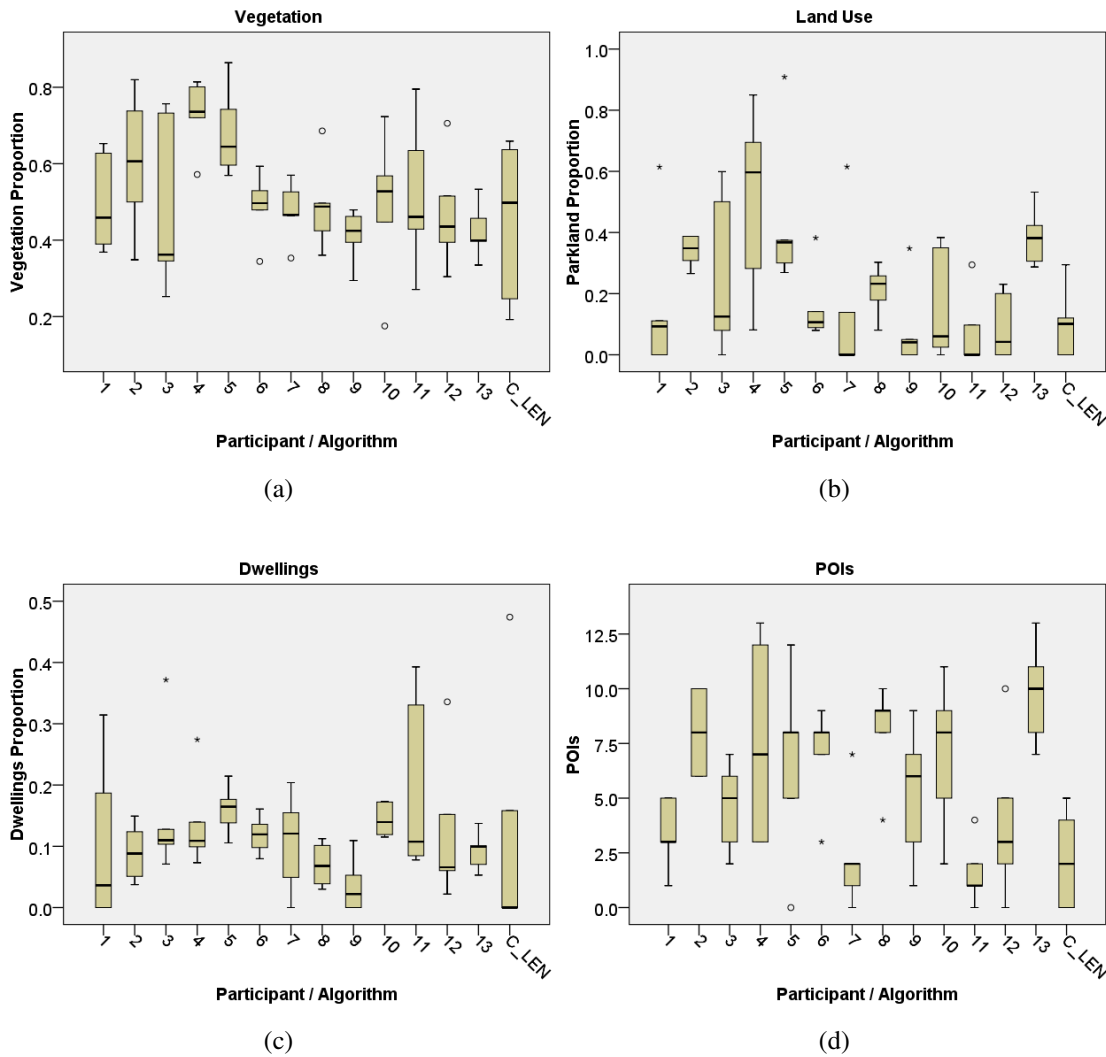
Figure 7.22: Individual attribute analysis of the participant suggested leisure routes for start-end pairs showing (a) vegetation, (b) land use, (c) dwellings and (d) POIs. Boxes indicate 25% to 75% and whiskers indicate the $\pm$ 1.5*interquartile range.

Despite Figure 7.22c again showing some variation, all of the participant routes do have higher mean proportions of dwellings compared to the minimum length routes selected by C_{LEN}. As with tourist routes, this indicates that housing is likely to be influential for leisure routes as suggested by Hypothesis H6 and the results of Experiment 3 (Section 3.6).

As with vegetation and land use, the importance of points of interest for leisure journeys (Figure 7.22d) appears to vary from participant to participant. Four participants (1, 7, 11 and 12) all have lower mean points of interest, of which two (participants 7 and 11) have means which are at least as low as those observed for the shortest paths selected by

C_{LEN}. However, participants 8 and 13 show relatively high numbers of points of interest on their selected routes. In all cases the mean number of 'sights' for leisure journeys are lower than those encountered on tourist routes suggested by the same participant. This is likely to indicate that any influence this attribute has on leisure routes will be less than on their tourist equivalents, as suggested by the lower ranking shown in the results of Experiment 6.

Overall, this analysis appears to show that different participants are influenced by different attributes when suggesting routes appropriate for leisure travel. Some choose routes with high vegetation, land use and points of interest, whereas others seem to stick to routes that are much shorter and contain only dwellings. This again indicates that the use of one general fixed algorithm may not be appropriate for all users.

7.3 Route Comparison

The participant routes that were gathered and analysed in the previous section form a suitable baseline for evaluating the algorithm-selected routes. Although they do not necessarily represent the best possible routes for each attribute category or journey type, each is considered appropriate for that type of journey by at least the participant or participants that suggested it. By comparing the routes selected by the algorithms described in Chapter 6 against these human suggestions, the appropriateness of the algorithm routes can be determined.

7.3.1 Simplicity Routes

Figure 7.23 shows routes suggested by the algorithm with cost function C_{SIMP}, and Table 7.6 gives an analysis of the similarity between the algorithm selected routes and human suggestions. Table 7.6 shows that the majority of the routes chosen by C_{SIMP} are exact matches to the most preferred routes (pairs 1, 2, 3 and 4); however, the route suggested for pair 5 has very little similarity to that chosen by the participants (see Figure 7.23e). The mean similarities are more varied however, with Zone 4 and Zone 5 algorithm routes having less than a 50% mean similarity when compared to the different participant routes. For the Zone 4 point pair this value is misleading, being mostly due to the effects of the substantially different routes suggested by participant 12. The weighted similarity, which weights the indivdual route similarities according to the number of participants suggesting them, offers a better way of representing these results. The weighted mean better denotes the true Zone 4 route similarity (where the algorithm matched the most preferred route),

and emphasises how low the mean is for Zone 5.

Figure 7.23: Algorithm and participant suggested simplicity routes for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5. Participant routes are shown in blue, the most preferred in purple, and the algorithm suggested route in red.

Compared to the most preferred route for Zone 5, the C_{SIMP} suggested route is longer (70%), but contains fewer decision points (13 vs 19) and fewer turns (6 vs 12). This unanticipated observation suggests that these attributes play less of a role in selecting simple routes that anticipated by Hypothesis H8 (Section 3.1).

Zone	Similarity				Weighted Similarity	
	Most Preferred	Maximum	Mean	SD	Mean	SD
1	100%	100%	100%	N/A	100%	N/A
2	100%	100%	100%	N/A	100%	N/A
3	100%	100%	66%	30%	72%	25%
4	100%	100%	36%	42%	64%	41%
5	3%	34%	12%	12%	9%	11%
Mean	81%	87%	63%	28%	69%	26%

Table 7.6: Comparison of the routes suggested by C_{SIMP} with those that were suggested by participants. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

As the shortest path is commonly used by existing pedestrian navigation systems, the route selections made by the shortest route algorithm (C_{LEN}) provide a suitable baseline to provide a point of reference or with which to judge the overall performance of the journey specific algorithms discussed in this research. Figure 7.12a strengthens the argument for the use of this metric for comparison to the 'simplest' routes, as although lower in frequency than expected, length is the highest ranking attribute (mentioned by 5 of the 14 participants).
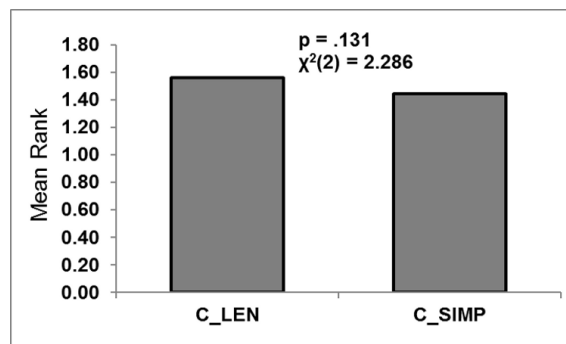

Figure 7.24: Route similarity Friedman Test results (rank and p values) for the simplicity and C_{LEN} cost functions.

Figure 7.24 shows the results of conducting a Friedman statistical analysis of the similarities between the routes suggested by C_{LEN} and C_{SIMP} and those suggested by the

participants. These results indicate that C_{LEN} performs slightly better than C_{SIMP} for the chosen pairs and environment, but that there is no statistical significance for this difference. The similarity between the shortest routes (as suggested by C_{LEN}) and participants' suggestions, shown in Table 7.7, gives exact matches to the most preferred participant route for all point pairs. Although the Zone 4 and Zone 5 mean similarities may seem low, the weighted mean similarities again give a better representation of the overall performance.

C_{LEN}					
Zone	**Similarity**			**Weighted Similarity**	
	Most Preferred	**Mean**	**SD**	**Mean**	**SD**
1	100%	100%	N/A	100%	N/A
2	100%	100%	N/A	100%	N/A
3	100%	60%	36%	69%	29%
4	100%	36%	42%	64%	41%
5	100%	50%	40%	72%	40%
Mean	100%	69%	39%	81%	37%

Table 7.7: Comparison of the routes suggested by C_{LEN} against those that were suggested by participants. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

Overall, this suggests that C_{LEN} selects the most appropriate 'simple' routes rather than the proposed C_{SIMP}. This is not entirely unexpected as much previous research [98, 182, amongst many others] has established the strong relationship between length and simplicity. It does however indicate that the simplicity attributes described in Section 3.5 have less influence over route choice than anticipated.

7.3.2 Everyday Routes

The everyday routes shown in Figure 7.25 and analysis in Table 7.8, indicate that C_{EVER} produces routes that are exact matches to those given by the majority of participants. Together they therefore illustrate that C_{EVER} gives appropriate routes for everyday travel, and length alone is sufficient to suggest routes for this type of journey (as $C_{EVER} = C_{LEN}$). Given that C_{LEN} also suggests the most appropriate 'simple' routes as indicated by the previous section, this also implies that the majority of participants are also choosing the simplest routes for everyday travel. Although not surprising as length was ranked as the most influential attribute in Experiment 6 (Section 3.9), it does indicate that the influence of vegetation, turns and points of interest and land use are lower than would be expected.

For vegetation this may be an artifact of the test area which contains a high proportion of vegetation even on short routes, and may need to be included in differing environments. However, the lack of corresponding comments made by participants when asked for their reasoning on route choice suggests that this inclusion is unlikely to be required.

The low unweighted mean shown for Zone 5 (39%) indicates that a small number of participants considered routes with substantial differences to be more appropriate. This reflects the differences illustrated in Section 7.2.3, which indicated that a few participants were willing to select slightly longer routes in order to increase levels of vegetation, parkland or points of interest. Exact matches for two of these alternatives were found using other cost functions from Section 6.3, but the lack of similarity to routes from other participants or zones make any relationships tenuous at best. In addition, the substantial differences in route selections made by participant 13 will have a large influence on this unweighted mean. The weighted mean (78%) also shown in Table 7.8 provides a better indication of the overall performance of the C_{EVER} algorithm.

For the second route suggested by half of the participants for the Zone 3 point pair, which has a slight increase length and corresponding rise in vegetation, several cost functions for Section 7.2.3 ($C2_{EVER}$, $C7_{EVER}$, $C9_{EVER}$, $C11_{EVER}$ and $C12_{EVER}$) found an exact match. Although this implies that the participants who selected this route may be considering a different set of attributes for route choice, the size of the variation and number of matching cost functions may also indicate a coincidence rather than a relationship.

Zone	Similarity			Weighted Similarity	
	Most Preferred	Mean	SD	Mean	SD
1	100%	100%	N/A	100%	N/A
2	100%	63%	53%	95%	20%
3	100%	76%	34%	76%	25%
4	100%	61%	40%	76%	29%
5	100%	39%	36%	78%	37%
Mean	100%	68%	41%	85%	28%

Table 7.8: Comparison of the routes suggested by C_{EVER} against those that were suggested by participants. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

Overall, these results show that C_{EVER} selects routes which are appropriate for everyday travel.

(a)

(b)

(c)

(d)

(e)

Figure 7.25: Algorithm and participant suggested everyday routes for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5. Participant routes are shown in blue, the most preferred in purple, and the algorithm suggested route in red.
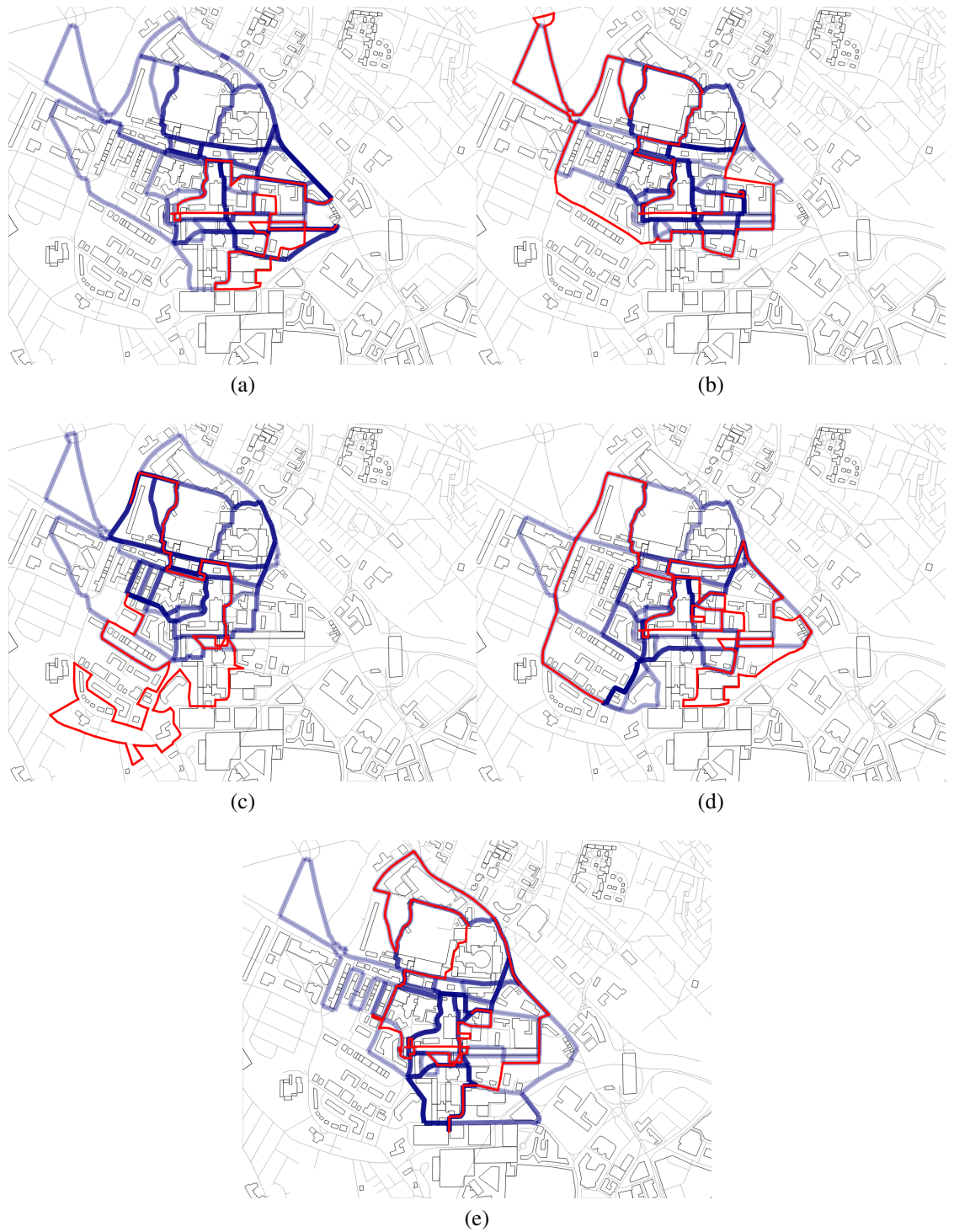
7.3.3   Tourist Routes

Figure 7.26: Routes suggested by C_{TOUR} for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5.

The routes suggested by C_{TOUR} are shown in Figure 7.26, and a comparison between the algorithm and participant suggested routes (Table 7.9) indicates that route similarity is low. Although a maximum similarity of 41% is encouraging given the variation in human suggestions and the thousands of possible routes, this is still lower than anticipated. It would be expected that if even a single participant was choosing routes by considering the suggested attributes (vegetation, land use dwellings, points of interest and length) in the same way as C_{TOUR}, then the maximum similarity would be much higher.

Zone	Similarity		
	Best	**Mean**	**SD**
1	30%	15%	8%
2	41%	13%	11%
3	18%	9%	6%
4	31%	10%	8%
5	27%	13%	8%
Mean	29%	12%	8%

Table 7.9: Comparison of the routes suggested by C_{TOUR} against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

There may be many reasons for this lack of mean similarity, such as differing routes having similar characteristics, but examining the attribute characteristics of the routes may offer some explanations. Figure 7.27 shows an analysis of the characteristics of the participant suggestions compared to those of C_{LEN} and C_{TOUR}. It indicates that C_{TOUR} selects routes that are substantially longer than those chosen by the participants, containing more decision points (DPs) and turns. For these three attributes, the participant routes actually have more in common with C_{LEN}. C_{TOUR} also suggests higher proportions of vegetation for the majority of routes, where again the participants' suggestions are closer to those selected by C_{LEN}. However, for land use and dwellings, the attribute values of the participant routes are more similar to C_{TOUR}, although C_{TOUR} has a slightly higher proportion of dwellings on its routes.

Far more points of interest are encountered by C_{TOUR} route selections than by the participant suggestions, although these in turn contain substantially more than the routes from C_{LEN}. Overall, the participants seem to have considered length to be far more influential than is accounted for in C_{TOUR}, whereas C_{TOUR} places more influence on the higher ranking attributes from Table 6.1. This again increases the likelihood that humans select tourist routes which are good but not necessarily the best, with shorter routes than expected being preferred.
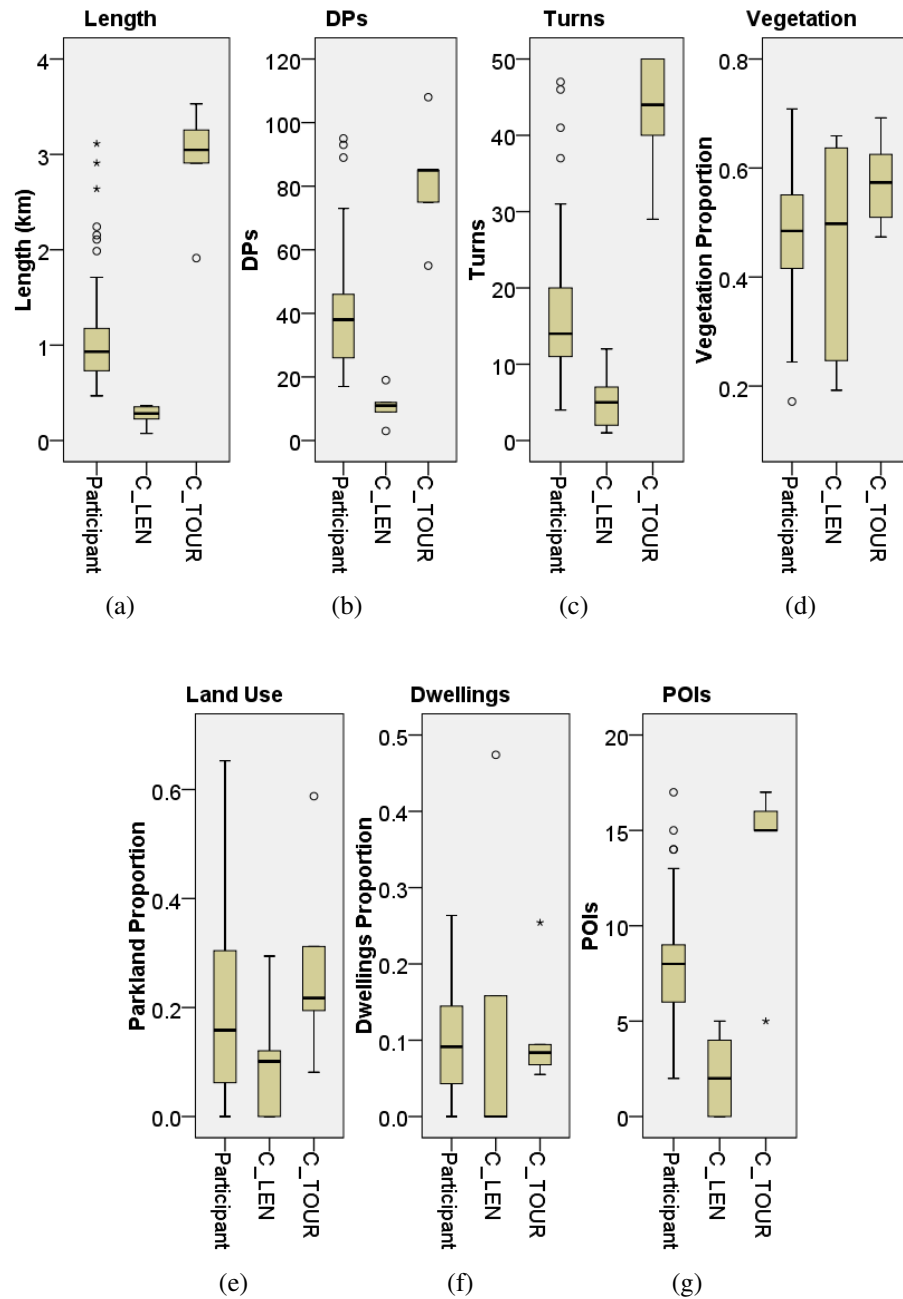
Figure 7.27: Attribute characteristics of participant selected routes compared to those of C_{LEN} and C_{TOUR}. Shown are (a) length, (b) DPs, (c) turns, (d) vegetation, (e) land use, (f) dwellings and (g) POIs.

The attribute variations observed in Section 7.2.1 suggest that different participants may use different attributes to chose tourist routes. This may also indicate that one of the alternative cost functions from Section 6.4 could provide more appropriate tourist routes. Figure 7.28 shows the Friedman analysis of the similarities between the participant

suggested routes and those suggested by each of the tourist cost algorithms. The arrows overlaid on Figure 7.28 show that $C1_{TOUR}$ selects routes with the highest similarity to the participants suggestions, and performs significantly better than any other cost function, excluding $C8_{TOUR}$.
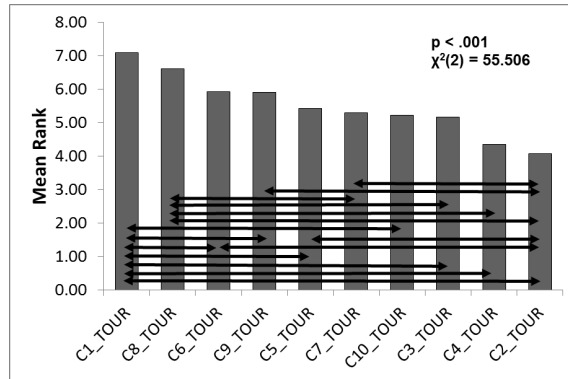
Figure 7.28: Friedman Test Results (rank and p values) for the tourist cost functions - pairwise (Wilcoxons) statistical significance is indicated by the arrows overlaid on the plot.

Table 7.10 indicates the similarity between the routes chosen by $C1_{TOUR}$ and those suggested by the participants, and although the results are better than C_{TOUR}, the maximum similarity is reduced to only 40% and the mean similarity is still only 16%. This suggests that although the human participants may be placing more importance on the POIs on a route than C_{TOUR}, there must still be something else determining that a route is selected as suitable for tourist journeys.

Zone	Similarity		
	Best	**Mean**	**SD**
1	29%	13%	8%
2	35%	21%	8%
3	29%	17%	9%
4	36%	10%	9%
5	40%	21%	11%
Mean	34%	16%	9%

Table 7.10: Comparison of the routes suggested by $C1_{TOUR}$ against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

An observation made in Section 7.2.1 and the characteristics shown in Figure 7.27a may provide one explanation for the lack of similarity. Figure 7.9a indicates that although the length of the routes suggested by the participants varies both within participant and

across participant, all but six of the routes are shorter than the anticipated 2km. This is somewhat surprising, given that no limit was stated on the time or distance for tourist routes during the relevant trials. In addition, some of the algorithm selected routes are substantially longer than this 2km threshold as set by $Length_{LIMIT}$ in the C_{LENCAP} component of C_{TOUR}.

In addition to suggesting a possible explanation for the low mean similarities shown between participant and algorithm selected routes, this observation also offers a simple method to potentially increase similarity and therefore the appropriateness of the routes. $Length_{LIMIT}$ in C_{LENCAP} presents an easily adjustable value, which can be varied without directly affecting the weight placed on distance within the relevant cost functions. However, despite not directly affecting distance weighting, this reduction in length is likely to have an impact on the other attributes considered influential for tourist journeys. Chapter 5 and Chapter 6 have indicated that shorter routes tend to contain lower proportions of the attractiveness attributes (vegetation, parkland, dwellings and points of interest), but given the participant route characteristics shown in Figure 7.27 the resulting effect is more likely to be advantageous than detrimental. The mean distance traversed by participant suggested routes is 1km, indicating that this would provide a better value for $Length_{LIMIT}$ than the 2km used in C_{TOUR}. Substituting this value, gives the similarity analysis shown in Table 7.11.

| | **Similarity** | | |
Zone	**Maximum**	**Mean**	**SD**
1	51%	22%	12%
2	22%	11%	5%
3	45%	14%	10%
4	39%	11%	10%
5	24%	11%	7%
Mean	36%	14%	9%

Table 7.11: Comparison of the routes suggested by C_{TOUR} with $Length_{LIMIT} = 1km$ against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.
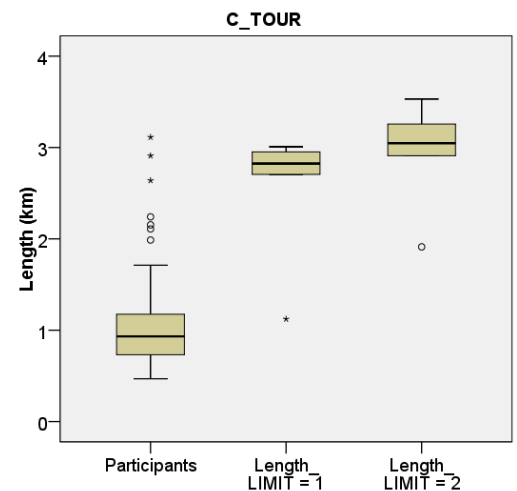

Figure 7.29: Comparison of the lengths of the routes suggested by C_{TOUR} with $Length_{LIMIT} = 1km$ and $Length_{LIMIT} = 2km$ against those that were suggested by participants for each point pair.

Although the maximum similarity shown in Table 7.11 is increased to 51%, the mean

similarity is actually lower. Substantials drops are also shown in the maximum similarity for two zones, 16% for the Zone 5 point pair and 13% for the Zone 2 point pair. In addition Figure 7.29 shows that the mean length of the algorithm selected routes is still 2.8km (compared to 3km where $Length_{LIMIT} = 2km$), and up to 2km longer than the participant mean length for each zone. This indicates that the influence of C_{LENCAP} is too low to substantially affect the length of the routes selected with $Length_{LIMIT} = 1km$, and that reducing this value still further will be unlikely to produce substantially better results.
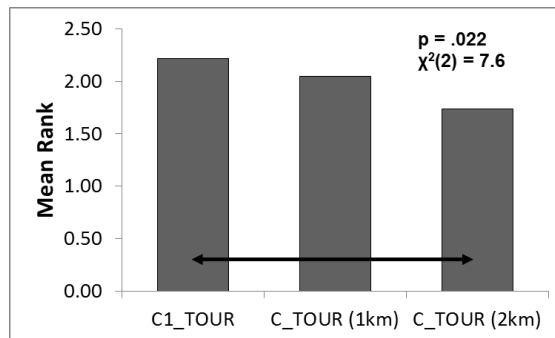
Figure 7.30: Friedman Test Results (rank and p values) for C_{TOUR} with $Length_{LIMIT} = 1km$, C_{TOUR} with $Length_{LIMIT} = 2km$ and $C1_{TOUR}$ - pairwise (Wilcoxons) statistical significance is indicated by the arrows overlaid on the plot.

There is also no statistical significance between C_{TOUR} with $Length_{LIMIT} = 1km$ and C_{TOUR} with $Length_{LIMIT} = 2km$, or C_{TOUR} with $Length_{LIMIT} = 1km$ and $C1_{TOUR}$, as shown by the mean ranks given and lack of arrows in Figure 7.30. Overall, this suggests that $C1_{TOUR}$ suggests the most appropriate routes of all of the cost functions tested, but that with a mean similarity of 16% and substantially longer routes, these may still not be suitable for tourist journeys. Figure 7.31 shows the routes chosen by $C1_{TOUR}$.

(a) (b)

(c) (d)

(e)
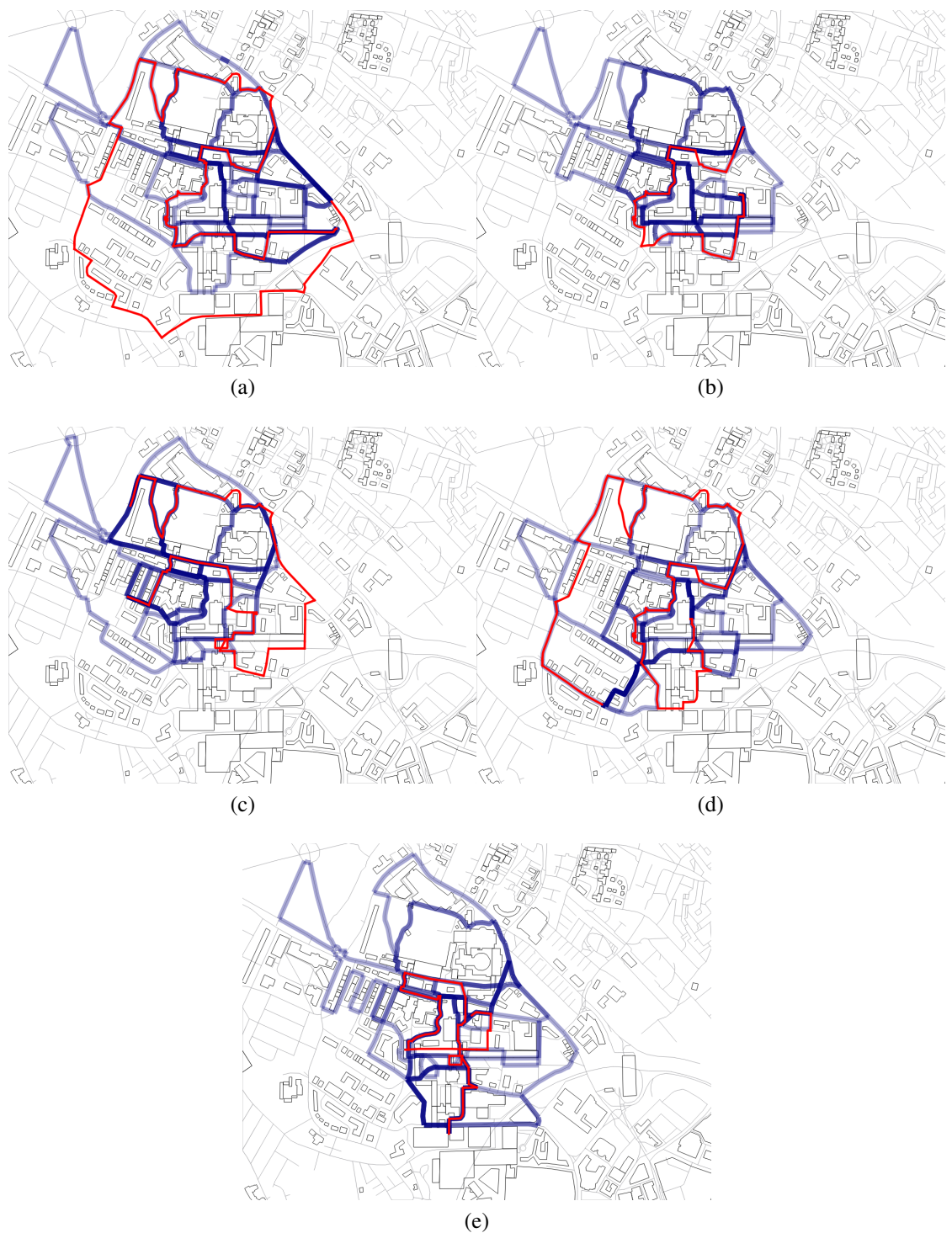
Figure 7.31: Routes suggested by $C1_{TOUR}$ for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5.

7.3.4 Leisure Routes

Figure 7.32 shows the routes chosen by C_{LEIS} for each start-end test pair, and the analysis in Table 7.12 indicates that although C_{LEIS} produces routes which are good in terms of

Figure 7.32: Routes suggested by C_{LEIS} for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5.

their attribute trade-off (as discussed in Section 6.5), it performs poorly in terms of matching the routes suggested by participants. Although the mean similarities were expected to

be poor, given the lack of consistency between the participant routes, a maximum similarity of 37% is again encouraging.

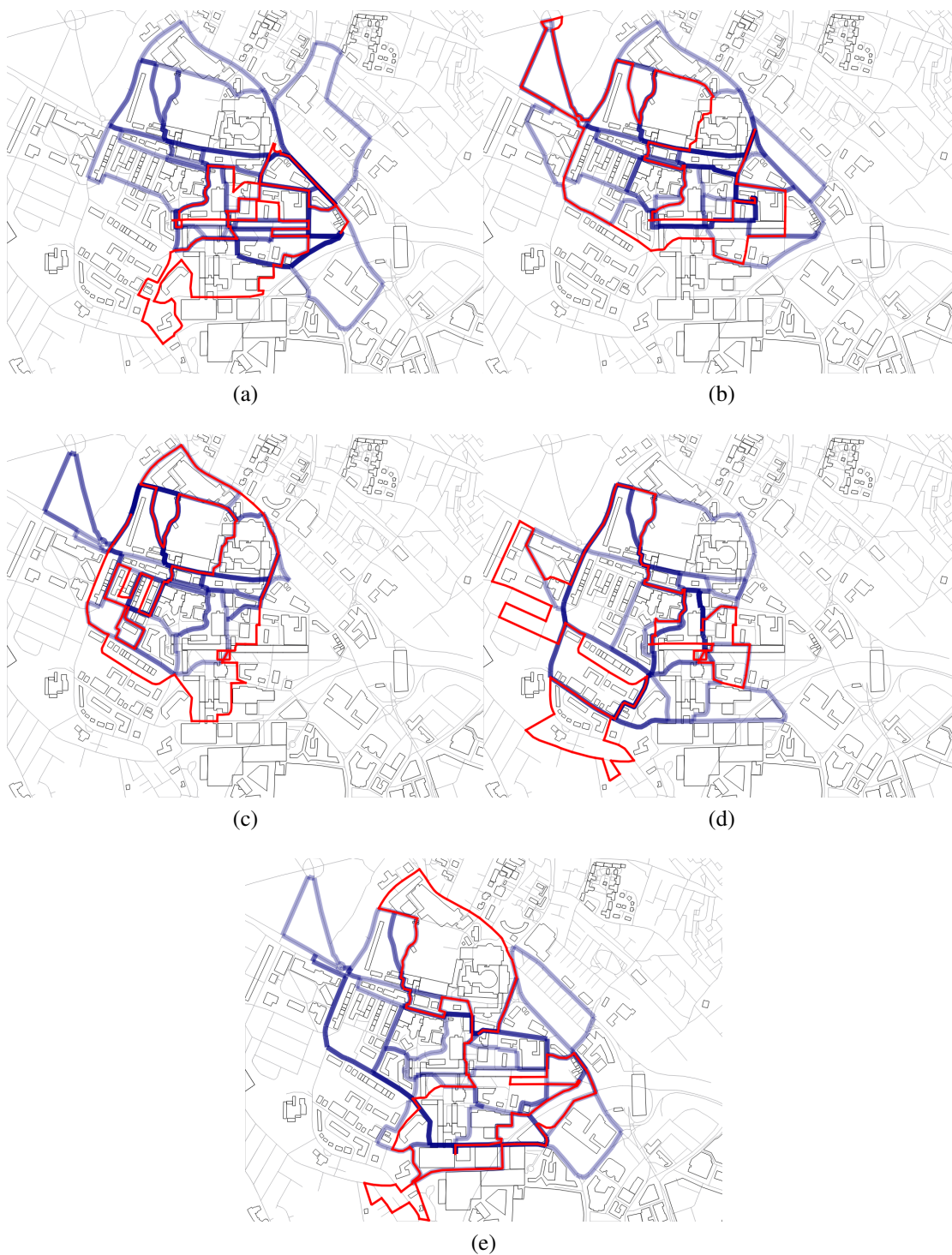| Zone	Similarity		
	Maximum	**Mean**	**SD**
1	19%	12%	6%
2	37%	13%	10%
3	19%	12%	4%
4	20%	11%	7%
5	22%	10%	6%
Mean	24%	12%	6%

Table 7.12: Comparison of the routes suggested by C_{LEIS} against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

As with tourist routes there may be many reasons for this lack of similarity, but it may be possible that the algorithm is choosing different routes with similar characteristics. To examine this possibility, Figure 7.33 shows the attribute characteristics of the participant suggested routes compared to those selected by C_{LEN} and C_{LEIS}. They indicate that C_{LEIS} again selects routes that are substantially longer than those chosen by the participants (Figure 7.33a), containing more decision points (DPs shown in Figure 7.33b) and turns (Figure 7.33c), with the participant routes having more in common with C_{LEN}. C_{LEIS} also suggests higher proportions of vegetation for the majority of routes (Figure 7.33d), with many of the participants' suggestions being closer in value to those selected by C_{LEN}. However, for land use (Figure 7.33e) and dwellings (Figure 7.33f), the attribute values of the participant routes are far more similar to C_{LEIS}, although C_{LEIS} has slightly higher proportions of both. For points of interest (Figure 7.33g) far more are encountered by C_{LEIS} route selections than by the participant suggestions, although the number included by participants is substantially higher than those with C_{LEN}. These observations suggest that participants seem to have considered length to be far more influential than anticipated by C_{LEIS}, as was seen with C_{TOUR}. Given that C_{LEIS} places more importance on the higher ranking attributes from Experiment 6 such as vegetation and points of interest, this may account for some of the inconsistencies observed between human and algorithm suggested routes. This finding again increases the likelihood that humans select leisure routes which are good but not necessarily the best, with far shorter routes than expected being preferred.
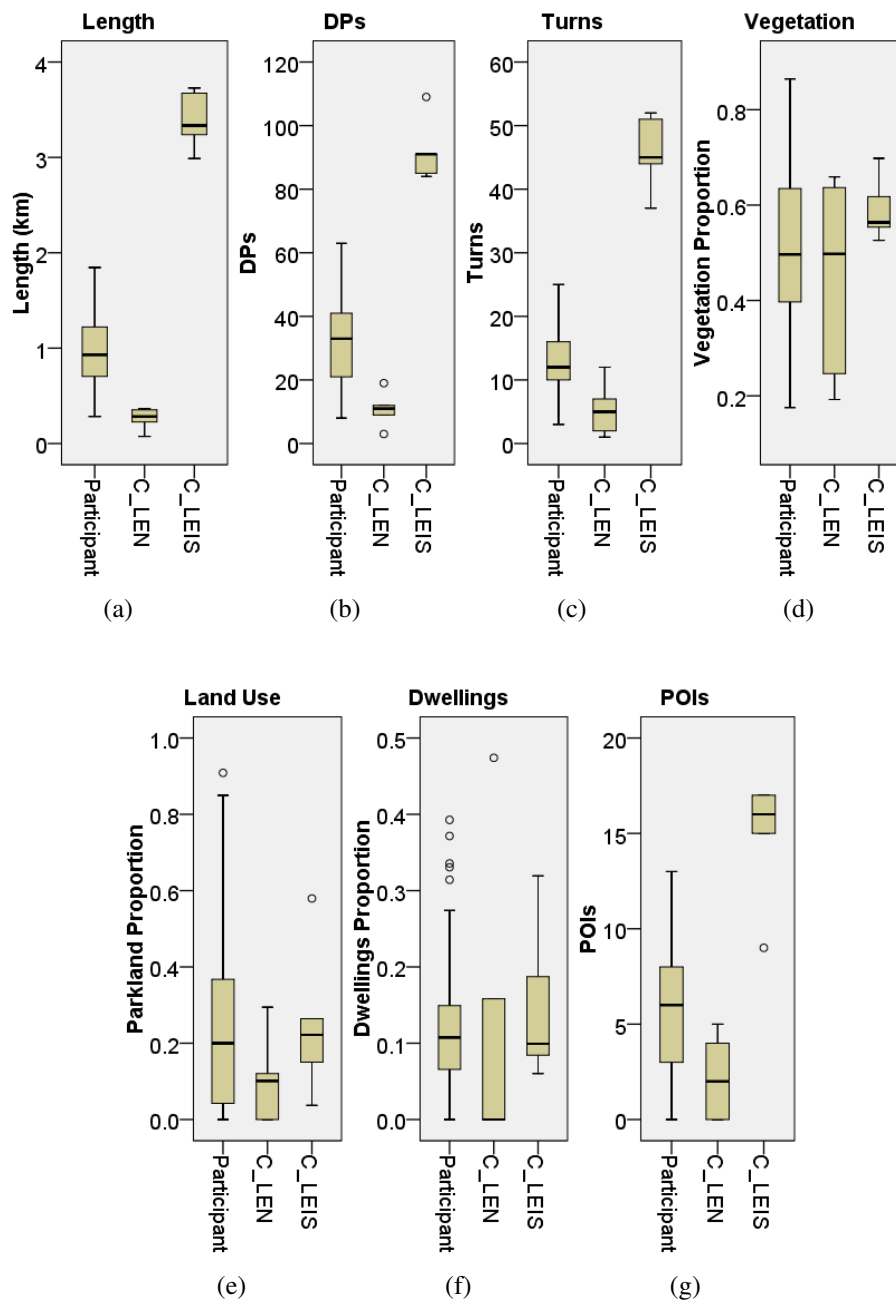
Figure 7.33: Attribute characteristics of participant selected routes compared to those of C_{LEN} and C_{LEIS}. Shown are (a) length, (b) DPs, (c) turns, (d) vegetation, (e) land use, (f) dwellings and (g) POIs.

The attribute variations observed in Section 7.2.4 suggest that different participants use different attributes to choose leisure routes, which may indicate that one of the alternative cost functions from Section 6.5 could provide more appropriate routes. Figure 7.34 shows the Friedman analysis of the similarities between the participant suggested routes
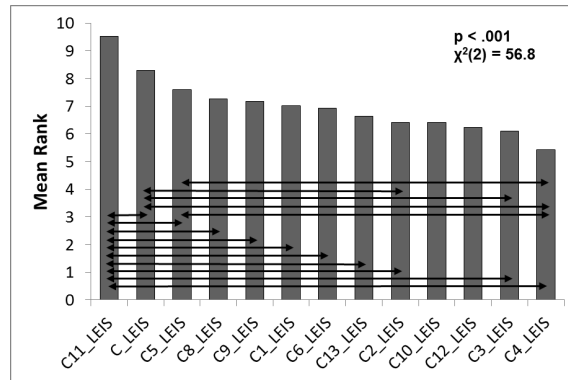
Figure 7.34: Friedman Test Results (rank and p values) for the leisure cost functions - pairwise (Wilcoxons) statistical significance is indicated by the arrows overlaid on the plot.

and those suggested by each of the tourist cost algorithms. The arrows overlaid on Figure 7.34 show that $C11_{LEIS}$ selects routes with the highest similarity to the participants' suggestions, and performs significantly better than any other cost function.

Zone	Similarity		
	Maximum	**Mean**	**SD**
1	41%	22%	14%
2	54%	21%	16%
3	83%	29%	20%
4	31%	11%	8%
5	21%	11%	6%
Mean	46%	19%	13%

Table 7.13: Comparison of the routes suggested by $C11_{LEIS}$ against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

Although $C11_{LEIS}$ performs best overall, and despite their low ranking positions in Figure 7.34, $C10_{LEIS}$ and $C12_{LEIS}$ both provide exact matches for two of the participant suggested routes (one for Zone 2, and the other for Zone 3). Despite these matches the mean similarities are lower than $C11_{LEIS}$, suggesting that they are likely to represent oddities of the environment and cost functions rather than a detectable relationship.

Figure 7.21a indicates that although the length of the routes suggested by the participants varies both within participant and across participant, all of the routes are shorter than the anticipated 2km. As with tourist routes, $Length_{LIMIT}$ in C_{LENCAP} presents an easily adjustable value, which can be varied without directly affecting the weight placed on distance within the relevant cost functions. The mean distance traversed by partic-

Zone	Similarity		
	Maximum	**Mean**	**SD**
1	7%	1%	2%
2	100%	18%	27%
3	77%	22%	21%
4	46%	16%	14%
5	67%	16%	21%
Mean	59%	14%	17%

(a)

Zone	Similarity		
	Maximum	**Mean**	**SD**
1	7%	1%	2%
2	51%	11%	16%
3	100%	28%	29%
4	46%	16%	14%
5	67%	16%	21%
Mean	54%	14%	16%

(b)

Table 7.14: Comparison of the routes suggested by (a) $C10_{LEIS}$ and (b) $C12_{LEIS}$ against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

ipant suggested routes is again 1km, suggesting that this would provide a better value for $Length_{LIMIT}$ than the 2km used in C_{LEIS}. Substituting this value gives the similarity analysis shown in Table 7.15.

Zone	Similarity		
	Maximum	**Mean**	**SD**
1	30%	16%	8%
2	25%	10%	7%
3	43%	15%	11%
4	27%	10%	7%
5	21%	10%	5%
Mean	29%	12%	8%

Table 7.15: Comparison of the routes suggested by C_{LEIS} with $Length_{LIMIT} = 1km$ against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.
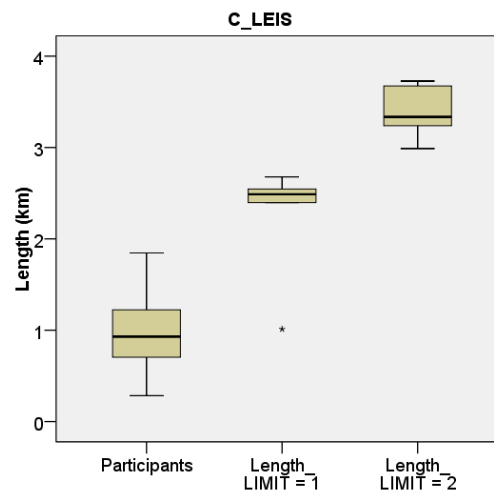
Figure 7.35: Comparison of the lengths of the routes suggested by C_{LEIS} with $Length_{LIMIT} = 1km$ and $Length_{LIMIT} = 2km$ against those that were suggested by participants for each point pair.

Although the maximum similarity shown in Table 7.15 is increased to 43%, and the mean similarity to 12%, these values are still low. A substantial drop of 12% is also shown in the maximum similarity for the Zone 2 point pair. As with the tourist routes, the mean length of the algorithm selected routes is also still much higher than that of the participant

suggestions (mean of 2.5km compared to the 1km expected), as shown in Figure 7.35. This again indicates that the influence of C_{LENCAP} is too low to substantially affect the length of the routes selected with $Length_{LIMIT} = 1km$, and that reducing this value still further will be unlikely to produce substantially better results.

In contrast Table 7.16 indicates that both maximum and mean similarities between $C11_{LEIS}$ and participant routes, are reduced by lowering $Length_{LIMIT}$ to 1km. This is contrary to the anticipated effect, but given that the mean lengths of the routes in Zone 2 and 3 are actually increased by this change in threshold, then it is not all that surprising. The reason for this increase is the low influence of C_{LENCAP} in $C11_{LEIS}$ as seen earlier, which is compounded by the drop in $Length_{LIMIT}$. The decrease indicates that $C11_{LEIS}$ with $Length_{LIMIT} = 2km$ actually provides more appropriate routes than $C11_{LEIS}$ with $Length_{LIMIT} = 1km$.

Zone	Similarity		
	Maximum	**Mean**	**SD**
1	36%	20%	11%
2	47%	15%	13%
3	43%	18%	13%
4	28%	16%	5%
5	21%	11%	6%
Mean	35%	16%	10%

Table 7.16: Comparison of the routes suggested by $C11_{LEIS}$ with $Length_{LIMIT} = 1km$ against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.
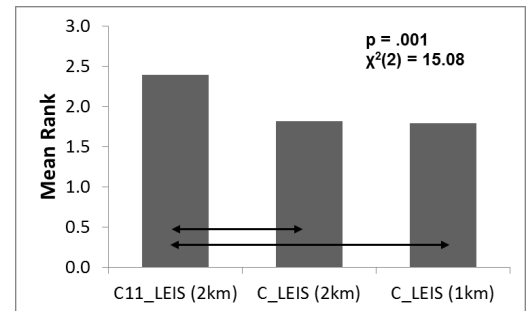

Figure 7.36: Friedman Test Results (rank and p values) for C_{LEIS} with $Length_{LIMIT} = 1km$, C_{LEIS} with $Length_{LIMIT} = 2km$ and $C11_{LEIS}$ with $Length_{LIMIT} = 2km$ - pairwise (Wilcoxons) statistical significance is indicated by the arrows overlaid on the plot.

Figure 7.36 shows no significance in the slight increase of C_{LEIS} with $Length_{LIMIT} = 1km$ over C_{LEIS} with $Length_{LIMIT} = 2km$, and that both are still significantly outperformed by $C11_{LEIS}$ with $Length_{LIMIT} = 2km$. Overall, this indicates that $C11_{LEIS}$ with $Length_{LIMIT} = 2km$ suggests the most appropriate routes of all of the cost functions tested, but that with a mean similarity of 19% and substantially longer routes than those chosen by participants, these may still not be suitable for leisure journeys. Figure 7.37 shows the routes chosen by $C11_{LEIS}$ with $Length_{LIMIT} = 2km$.

Figure 7.37: Routes suggested by $C11_{LEIS}$ with $Length_{LIMIT} = 2km$ for start-end pairs in Zone (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5.

7.3.5 Further Analysis

In addition to (and offering a possible explanation for) the low similarities seen in Section 7.3.3 and Section 7.3.4, two important findings were observed, both of which may have a
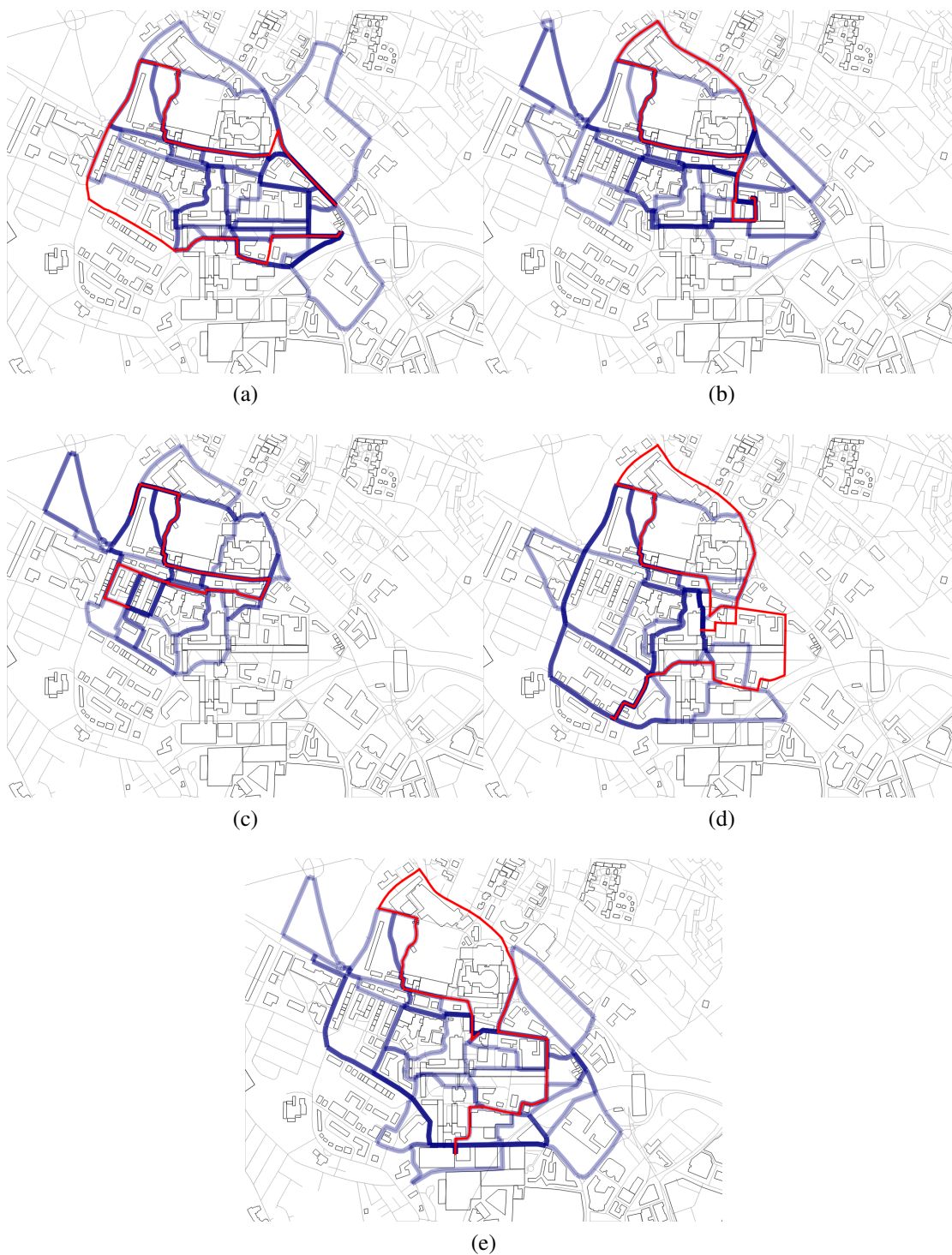
substantial impact on the choice of roues for leisure and tourist journeys:

1. None of the developed cost functions for leisure or tourist travel suggest routes which are consistently as short as those selected by the participants that took part in the trials described at the beginning of this chapter.

2. It is possible that participants are selecting routes which are 'good enough' rather than the best possible option.

In addressing these two points it was hoped that the route similarities exhibited by the algorithm routes could be increased, and a better understanding of the appropriateness of these routes could be established. This section examines a possible solution for the route length problem described in point 1, and investigates whether the 'good enough' theory postulated in point 2 can be justified.

7.3.5.1 Length

Section 7.3.3 and Section 7.3.4 indicate that $C1_{TOUR}$ and $C11_{LEIS}$ produce the most appropriate routes of the developed cost functions for tourist and leisure travel. However, not only do both of these cost functions select routes that are either unlikely to be suitable for pedestrian travel or not comparable to participant suggestions due to their length, but Section 7.3.4 also indicates that it is doubtful that reducing $Length_{LIMIT}$ for $C11_{LEIS}$ will improve this situation. An alternative solution to selecting shorter routes by promoting routes of a certain length as in C_{LENCAP} is to simply reject routes which are too long. A major advantage to this approach is that it can be applied to any cost function, not only those that contain C_{LENCAP}, making it suitable for reducing route length in $C1_{TOUR}$ as well as $C11_{LEIS}$. Selecting a rejection threshold of 2km (as would be appropriate for a 30 minute walking time), and applying it to $C1_{TOUR}$ as well as $C11_{LEIS}$, produces routes with the spread of lengths shown in Figure 7.38.

Figure 7.38a shows that this approach makes the routes selected by $C1_{TOUR}$ far more suitable for pedestrian journeys, with a maximum length of 1.9km (compared to 3.2km without route rejection) and mean of 1.7km (compared to 2km). Although it is less clear from Figure 7.38b, the maximum route length for route $C11_{LEIS}$ is also slightly reduced (1.6km compared to 1.7km) as is the mean (1.3km compared to 1.4km).

The comparisons of the routes produced by $C1_{TOUR}$ and $C11_{LEIS}$ with route rejection against those selected by participants are shown in Table 7.17. Although Table 7.17a indicates that the mean similarity of the routes suggested by $C1_{TOUR}$ does not increase with route rejection applied (when compared to the values in Table 7.10), it also doesn't
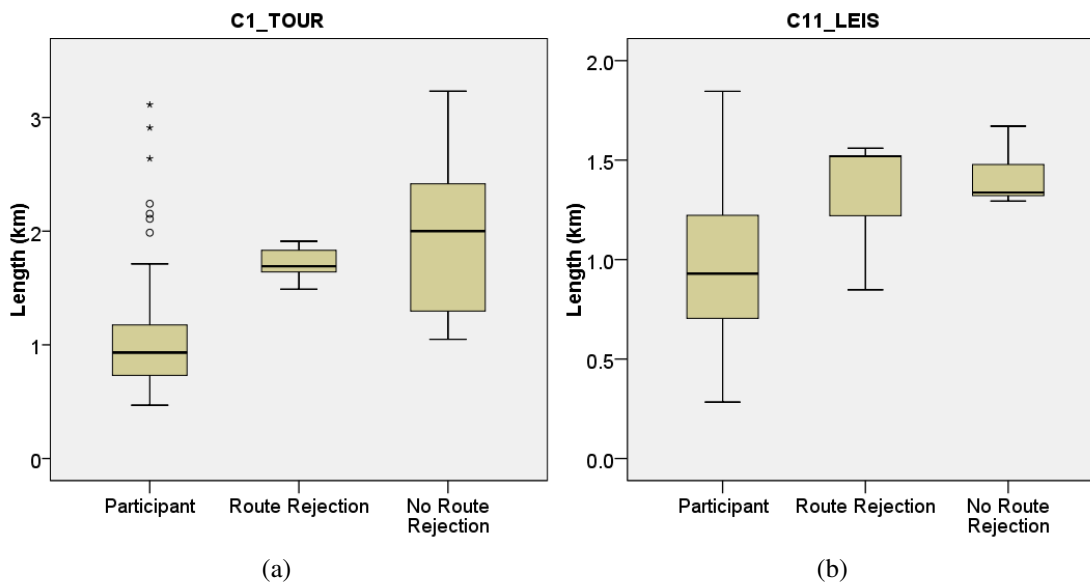
Figure 7.38: Comparison of the lengths of the routes suggested by (a) $C1_{TOUR}$ and (b) $C11_{LEIS}$ with and without route rejection, against those that were suggested by participants for each point pair.

Zone	Similarity				Zone	Similarity		
	Maximum	Mean	SD			Maximum	Mean	SD
1	46%	23%	11%		1	55%	22%	17%
2	31%	15%	9%		2	54%	21%	16%
3	36%	15%	11%		3	19%	10%	5%
4	35%	12%	9%		4	24%	13%	4%
5	29%	17%	8%		5	46%	20%	13%
Mean	36%	16%	10%		Mean	39%	17%	11%
	(a)					(b)		

Table 7.17: Comparison of the routes suggested by (a) $C1_{TOUR}$ and (b) $C11_{LEIS}$ with route rejection against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

show a drop as would be expected if the approach was detrimental to the selection of tourist routes. Additionally, the maximum similarity increases from 40% to 46%.

Increases are also observed in the maximum and mean similarities between the routes chosen by $C11_{LEIS}$ ($Length_{LIMIT} = 1km$) with and without route rejection, when compared to the participant suggestions as shown in Table 7.18b. In this case the maximum similarity increases from 35% to 39% (from Table 7.16), and the mean similarity from

16% to 17%. However, the mean and maximum similarities are again lower than $C11_{LEIS}$ with $Length_{LIMIT} = 2km$, as shown in Table 7.13. The majority of this decrease is due to the removal of the near-miss found by $C11_{LEIS}$ with $Length_{LIMIT} = 2km$ in Zone 3. Despite this a substantial increase in the mean similarity of the Zone 5 routes is indicates, and also in the maximum similarities in Zone 1 and Zone 5.

Taking the restriction of route length further, although Figure 7.9a and Figure 7.21a show a small amount of within participant variation in route length, in general this is much smaller than the between participant variations. This indicates that the length of the routes suggested by individual participants is fairly consistent for both tourist and leisure journeys (with a small number of exceptions), and suggests that adjusting route length for each participant will produce closer matches to their routes.

To achieve this, $Length_{LIMIT}$ can be set for the individual participant by calculating the mean length of the routes suggested by them alone. This gives a route cost which is adapted to the specific needs of the user (or participant in this instance) without the need for increasing the complexity of the algorithm or the cost functions themselves. Using this value to also control the rejection threshold by setting it to twice the value of $Length_{LIMIT}$ (which is slightly larger than the range of lengths of routes suggested by the individual participants) will also give more control over route length. Analyses of the similarities resulting from varying $Length_{LIMIT}$ and the rejection threshold for $C1_{TOUR}$ and $C11_{LEIS}$ in this way are shown in in Table 7.18.

Zone	Similarity				Zone	Similarity		
	Maximum	**Mean**	**SD**			**Maximum**	**Mean**	**SD**
1	42%	26%	10%		1	34%	19%	10%
2	31%	17%	10%		2	54%	22%	16%
3	36%	19%	9%		3	45%	19%	13%
4	60%	17%	15%		4	29%	15%	7%
5	29%	16%	9%		5	46%	16%	14%
Mean	40%	19%	10%		Mean	41%	18%	12%
(a)					(b)			

Table 7.18: Comparison of the routes suggested by (a) $C1_{TOUR}$ and (b) $C11_{LEIS}$ with variable $Length_{LIMIT}$ and route rejection, against those that were suggested by participants for each point pair. Mean and standard deviation similarity are between the algorithm route and participants' suggested routes.

Table 7.18a again shows increases in both maximum and mean similarities between $C1_{TOUR}$ with variable $Length_{LIMIT}$ and participant suggestions, when compared to $C1_{TOUR}$ with fixed $Length_{LIMIT}$ (where both have route rejection). The overall maximum similar-

ity of 60% is much higher than expected and although the mean similarity is substantially lower at 19%, it is still higher than the 16% seen previously. Figure 7.39a also indicates that the routes chosen by $C1_{TOUR}$ with variable $Length_{LIMIT}$ are more suitable in terms of length, having a similar distribution to those of the participant suggestions, albeit slightly longer.
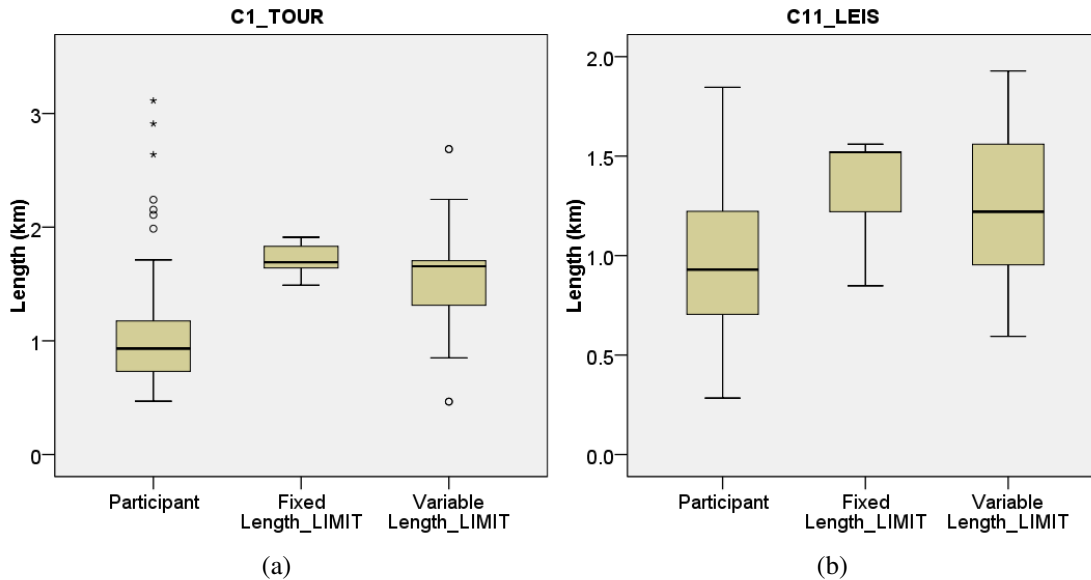

Figure 7.39: Comparison of the lengths of the routes suggested by (a) $C1_{TOUR}$ and (b) $C11_{LEIS}$ with fixed and variable $Length_{LIMIT}$, against those that were suggested by participants for each point pair.

Varying $Length_{LIMIT}$ also gives an increase in the mean similarity for $C11_{LEIS}$, as shown in Table 7.18b. In addition, although the overall maximum similarity is reduced from 55% to 54%, the maximum similarity across the zones (given as the mean of the maximums in Table 7.18b) shows a slight increase of 1%. Although this still represents a drop on similarity to $C11_{LEIS}$ with a fixed $Length_{LIMIT} = 2km$ this is again due to the elimination of the near-miss in Zone 3, and a better spread of maximum similarity is observed across the remaining zone. As with $C1_{TOUR}$, Figure 7.39b also indicates that lengths of the routes chosen by $C11_{LEIS}$ with variable $Length_{LIMIT}$ have a similar distribution to those of the participant suggestions, but are generally slightly longer.

In conclusion, this section indicates that rejecting routes that are considerably longer than the participant suggestions, and varying $Length_{LIMIT}$, allows $C1_{TOUR}$ and $C11_{LEIS}$ to produce more similar routes to human selections. This in turn implies that the algorithm suggestions are more appropriate for tourist and leisure journeys than the algorithms which do not include these approaches. The following section will examine the attribute

characteristics of these routes, and further investigate the theory that human participants are willing to accept routes that are 'good enough', to determine whether they routes selected in this way are really appropriate for tourist and leisure journeys.

7.3.5.2 Acceptance of 'Good' Routes

Section 7.2.1, Section 7.3.3 and Section 7.3.4 all discuss the possibility of participants being willing to accept 'good' leisure and tourist routes, instead of searching for the 'best' route in each case. In practical terms for the context of the present research, this means sacrificing high values of the top ranking attributes as found by the experiments in Chapter 3, in order to reduce length or increase lower ranking ones. Section 7.3.5.1 has already shown that humans are willing to accept shorter routes in exchange for a loss in attractiveness attributes, and Figure 7.40 and Figure 7.41 again show indications of this effect.

Figure 7.40g and Figure 7.41 suggest that points of interest are the main attribute to suffer from this effect, even on routes of similar lengths. For leisure journeys this is somewhat understandable, as points of interest rank only second in terms of importance, and levels of vegetation, parkland and dwellings are slightly higher for participant routes than those suggested by $C11_{LEIS}$. This may just indicate that the weights chosen to be the best in Section 6.5 apply more influence to points of interest than is actually true of the human route selection process for leisure journeys.

However, for tourist routes it is much harder to explain. Previous research [205, 206, amongst many others], Hypothesis H10 and the results of Experiment 6 (Section 3.9) all indicate that points of interest have the most influence on tourist journeys, but Figure 7.10d and Figure 7.40g appear to contradict this. It could be argued that participants are only selecting routes which pass the 'sights' which are mentioned on their questionnaires, rather than the list compiled from all participant suggestions, but the number of routes which pass no or few points of interest indicated in Figure 7.8 seem to disprove this theory. Vegetation does show a slight increase in proportion over the algorithm selected routes, which may indicate that participants were willing to accept more of this attribute in exchange for fewer points of interest.

To examine the hypothesis that humans select 'good enough' routes, the difference between the route cost for participant routes and $C11_{LEIS}$ and between participant and $C1_{TOUR}$ routes (both with with variable $Length_{LIMIT}$) were examined as shown in Figure 7.42. If the variation between participant and algorithm were purely due to the hypothesised acceptance of 'good' routes, then the difference in route cost would be expected to be small, indicating that the routes are almost equivalent, or consistent across the routes for each participant. Figure 7.42 indicates that neither of these is consistently true. In
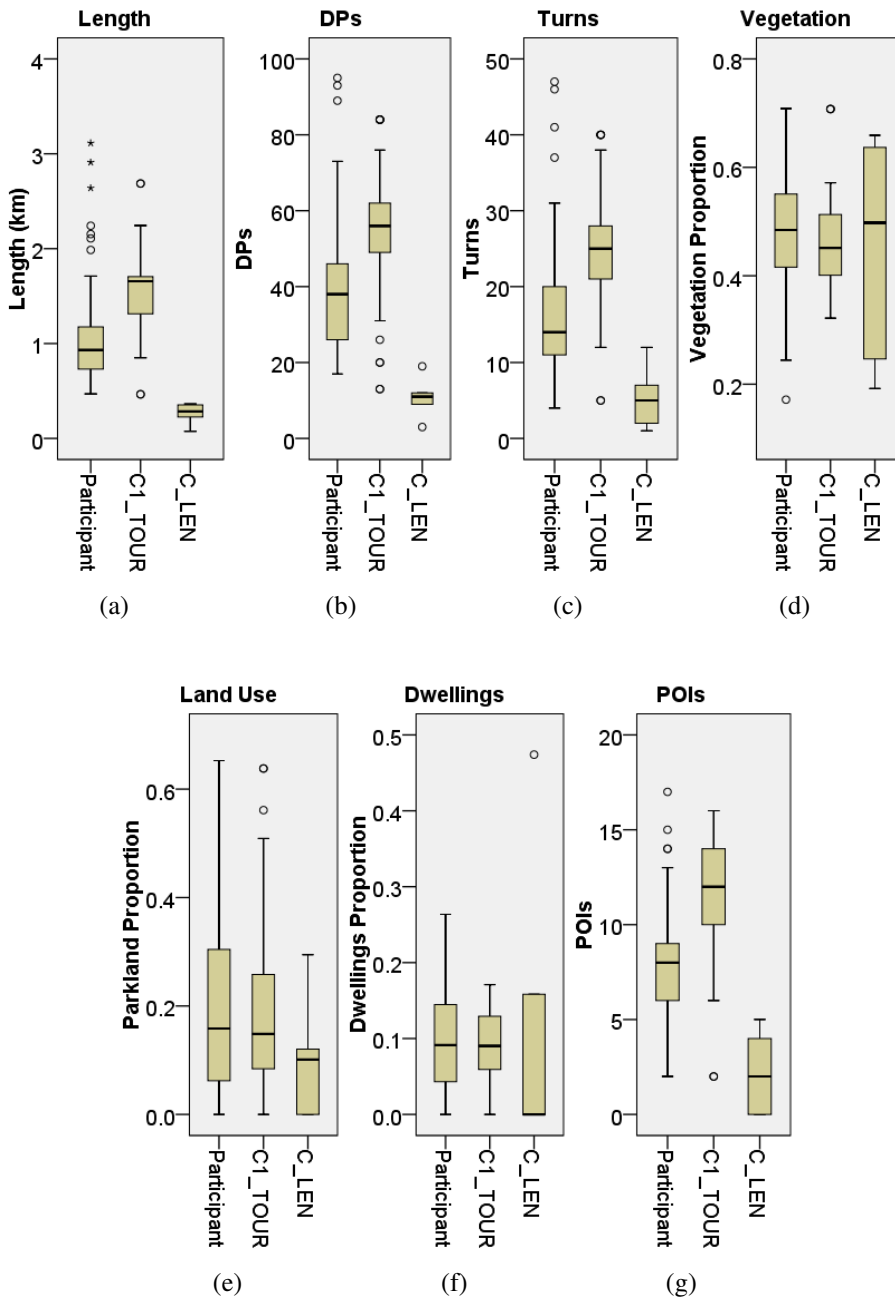
Figure 7.40: Attribute characteristics of participant selected routes compared to those of C_{LEN} and $C1_{TOUR}$. Shown are (a) length, (b) DPs, (c) turns, (d) vegetation, (e) land use, (f) dwellings and (g) POIs.

particular, the cost difference for $C1_{TOUR}$ appears to vary widely both within participant route choices and between participants.

Figure 7.43 shows that 6% of the routes suggested by $C11_{LEIS}$, and 10% of the routes suggested by $C1_{TOUR}$, have a larger route cost than those given by participants. This
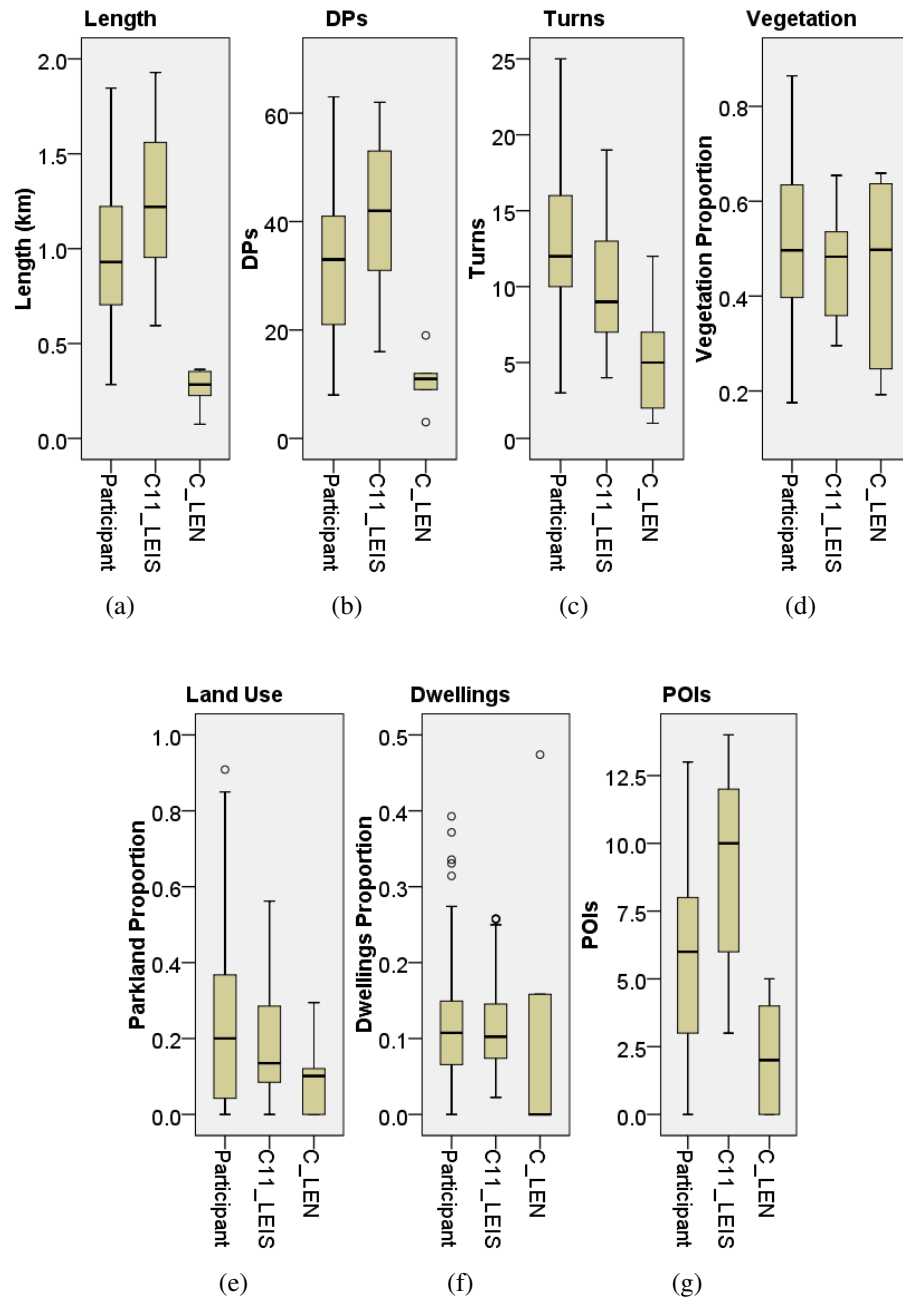
Figure 7.41: Attribute characteristics of participant selected routes compared to those of C_{LEN} and $C11_{LEIS}$. Shown are (a) length, (b) DPs, (c) turns, (d) vegetation, (e) land use, (f) dwellings and (g) POIs.

illustrates the heuristic limitations of the algorithm, and suggests that it may have a greater influence on variation than anticipated. With variations in cost differences differences of up to 140%, there is also very little consistency of route cost across each participant. However, excluding the routes with negative route difference, of the remaining routes
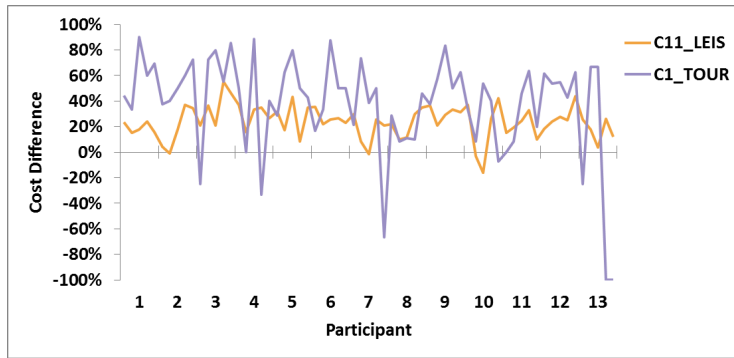
Figure 7.42: Percentage route cost difference for the participant routes when compared to those produced by $C11_{LEIS}$ and $C1_{TOUR}$ with variable $Length_{LIMIT}$ (ordered by participant).

46% of those for leisure and 41% of those for tourist have a route cost increase of less than 5%. This indicates that some willingness to accept 'good' routes is likely to play a part when selecting tourist or routes, although the definition of 'good enough' may not be purely based on this measure of route cost.
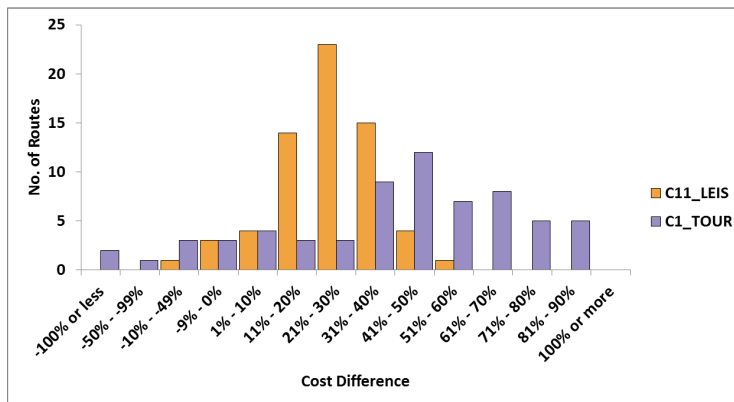

Figure 7.43: Frequency of route cost difference for the participant routes when compared to those produced by $C11_{LEIS}$ and $C1_{TOUR}$ with variable $Length_{LIMIT}$.

Overall, Figure 7.40 and Figure 7.41 do suggest that the routes produced by $C1_{TOUR}$ and $C11_{LEIS}$ do have many characteristics which are similar to the participant route choices. Furthermore in the case of $C1_{TOUR}$, these routes are actually likely to be more appropriate for tourist journey, as they include a higher proportion of points of interest for only small length increases. With some evidence that human participants are likely to choose routes which are 'good' but not necessarily the 'best', then these variations are likely to be acceptable for the majority of users. This would imply that $C1_{TOUR}$ and $C11_{LEIS}$ may in fact suggest appropriate routes for tourist and leisure travel.

There are however some important limitations to this outcome. Firstly, only a small

number of participant suggested routes were examined here, and a much larger test set would be required to confirm these findings. Secondly only a single environment was used to generate these routes, and artifacts specific to the test area chosen may be influential on route selection. An example of this might be the high levels of vegetation found to be present on the shortest routes for each point pair. Thirdly that comparison with participant suggested routes is the most suitable method for evaluating route choice, and that the the routes provided by this particular group of participants do actually constitute those that they thought were best for each type of journey. This may not be the case for all participants, particularly those with low levels of familiarity of the test area, and experts on an area may provide better examples. Finally, and most importantly, that the level of 'good enough' is sufficient to be acceptable to the majority of users. Although the variations between attribute distributions are small in most cases, they may be considered important to some users.

7.4   Conclusions

The route suggestion algorithms discussed in this section have varying levels of success for the four journey types that were tested. C_{EVER} performed best, selecting routes which exactly matched those suggested by the majority of participants for everyday journeys. This is somewhat unsurprising as C_{EVER} considers only route length, which is commonly accepted as the most important attribute for everyday travel. C_{SIMP} also performed well, suggesting exact matches for four of the five start-end pairs' most preferred routes. However even for the simplest route category, C_{LEN} found routes which were exact matches for the most preferred by participants, indicating there is no evidence that the multi-attribute approach suggests more appropriate routes than the commonly used shortest path algorithm.

In contrast, none of the developed cost functions for tourist or leisure routes produced an exact match to participant suggestions. Statistical analysis of the similarities between algorithm and participant selected routes, indicate that C_{TOUR} and C_{LEIS} are outperformed by the alternative $C1_{TOUR}$ and $C11_{LEIS}$. In the case of C_{TOUR} this is understandable as $C1_{TOUR}$ was rejected purely on its poor performance in terms of dwellings, but both instances indicate that the participants were using attributes for route choice that were different to those judged to produce the 'best' routes in Chapter 6. However the lack of similarity between these higher ranking cost function routes and participant suggestions is still low, although this lack of similarity between algorithm and participant is not all that surprising given that there is very little consistency between the participants'

routes for each start-end pair. A more in depth analysis indicated that both algorithms were selecting routes substantially longer than those chosen by participants, and an initial adaptation which reduced the level $Length_{LIMIT}$ suggested that this change alone would not be enough to either significantly improve the performance of C_{TOUR} and C_{LEIS}, or have a large enough impact on the length of the routes selected by $C11_{LEIS}$.

Further analysis of the route characteristics led to the development of a route rejection approach, which the reduced the length of routes produced by the more appropriate $C1_{TOUR}$ and $C11_{LEIS}$ to be more consistent with those of the participant suggestions. Taking this approach further, $Length_{LIMIT}$ and the route rejection threshold were tailored to the mean route length of the individual participant, and $C11_{LEIS}$ and $C1_{TOUR}$ were found to give routes which were again more similar to participant suggested routes.

Investigation into the concept that participants are willing to choose routes which are 'good' but not necessarily the best did find some evidence that this was actually the case. In addition comparison of the characteristics of participant and algorithm selected routes indicate that they are similar, and that this willingness to accept routes that are 'good enough' would be likely to make the variations seen permissible to most users. This therefore implies that the routes suggested by $C11_{LEIS}$ and $C1_{TOUR}$ with variable $Length_{LIMIT}$ and route rejection, may be appropriate for tourist and leisure journeys.

There are however a number of limitations to these conclusions. The maps used for evaluation purposes here may have influenced the routes selected as some attributes (such as length and turns) are more obvious than others (for example vegetation or dwellings), and therefore placed limitations on the validity of this approach. Similar issues exist with other areas of the study, with the satellite images used only being accurate when they were taken, and the instructions issued possibly influencing the information given in response (for example the tourist journey instructions asked for lists of points of interest). In addition only a small number of participants (14) were attracted to the study over a number of weeks, and only five start-end pairs and a single geographic area were included in the dataset, limiting the route suggestions available for an evaluation baseline. Despite this the results of the study do show variation in routes, and were considered suffucient in the present context.

Not all participants were familiar with the test points chosen, or all accessible areas of campus, indicating that their knowledge of the selected environment was limited. Despite this all participants also indicated that they were familiar with the majority of locations chosen, or at least with other points and areas nearby, and their level of knowledge was considered sufficient for evaluation purposes.

Limitations were also present in the method used to measure algorithm performance.

Similarity calculated by route overlap is only one of many metrics which could have been used to evaluate algorithm performance, and indeed route similarity overall. Although this may not be the most suitable option available, and with no accepted approach to evaluating routes in this way, it was considered a sufficient measure of performance in this context.

Finally the variations observed in the attribute characteristics of the algorithm selected routes may be more influential than anticipated, and may make these routes inapproriate in some cases. The concept that 'good' rather than 'the best' routes were being selected by participants was investigated by examining route cost differences, and although this did suggest that some reduction was acceptable in the highest ranking attributes was likely to be acceptable to users, this may not be the most suitable method for evaluating whether or not routes are actually appropriate. In addition overall trends in attribute characteristics were examined to determine if these routes were 'good enough', and individual routes may have attribute values that do not meet this criteria.

Chapter 8

Conclusions and Future Work

People plan their routes through new environments every day, but what factors influence these wayfinding decisions? Although previous researchers have studied how humans navigate, what affects their success and ability to do this and the attributes which encourage people to walk, until now few studies have investigated which factors are important when selecting a route for pedestrian travel beyond the shortest path approach.

With a greater freedom of movement than offered by roads for vehicular transport, and different requirements, an alternative approach was needed to find an answer for automatically selecting routes suitable for journeys on foot. Combining previous research on preference for pedestrian routes [83], studies indicating that different types of journey require different characteristics [6] and new empirical findings, a method for suggesting more appropriate routes was found.

Although previous research produced a number of route recommendation systems, the majority of these were restricted to a single route type or user group [36, 42, 95, 145, 152, amongst many others]. In addition many systems used complex approaches such as genetic algorithms or fuzzy logic [5, 33], or required information which is not readily available [4, for example]. By taking a simple approach based on Dijkstra's algorithm, a series of algorithms have been developed to select appropriate routes for different types of journey, using only data that can be gathered from maps or satellite images.

This chapter will examine in more detail the main achievements and limitations of the research within this thesis. It will also discuss future work, and ways in which the research could be extended.

8.1 Achievements and Limitations

The aim of this research was to develop an approach to route suggestion that could recommend routes according to the type of journey a person is making. It was designed to be simple, to make it computationally efficient and suitable for use in power-conscious environments. It was also based on only attributes which could gathered easily from maps or satellite images.

Three major contributions were made by this research:

1. This research has established previously unknown rank orders for the influence of seven environment and route attributes (length, turns, decision points, vegetation, land use, dwellings and points of interest), on 'attractive' route selection and for leisure and tourist journeys. In addition, this research has extended the previously known rank order of these attributes for everyday journeys. These ranks were found empirically by running a set of six experiments with 450 participants, which compared more attributes simultaneously than earlier studies.

2. This research has also developed new multi-attribute algorithms, based on Dijkstra's algorithm, which suggest routes that people would find attractive, and for tourist and leisure journeys. These algorithms were formulated using the developed environment model and rank-ordering, taken directly from experiments. By avoiding the complexity of previous approaches, these algorithms could be widely used across a variety of environments, and easily extended for different groups of users.

3. Two of the tourist and leisure algorithms that were developed have been shown to select routes that were similar to those suggested by participants in a user experiment for tourist and leisure journeys, respectively.

The above contributions were made by undertaking research in four areas. These are summarised in the following sections.

8.1.1 Ranking attributes according to their influence

By conducting a large scale user study, this research has established the rank of seven environment and route attributes (length, turns, decision points, vegetation, land use, dwellings and points of interest), in terms of influence, for route selection in two different attribute categories (simplicity and attractiveness) and three journey types (everyday, leisure and tourist). Although previous research has established a rank for the influence of some attributes on everyday route selection [83], and studies have examined

which attributes affect route choice and scene preference for leisure and tourist journeys [78, 106, 120, 134, 180, and many more], this study has compared more attributes simultaneously than in any earlier research. In addition, despite being common in fields such as travel choice [141], bicycle route selection [107] and walkability evaluation [119], the stated preference approach taken here is not typically applied to pedestrian route choice, and the use of static virtual environments within this approach is even rarer.

Three experiments used a stated preference approach to determine which of the selected attributes were influential for each attribute category and journey type, with simplicity and attractiveness having approximately the same results as seen previously [120, 180, 191, 210, 222, amongst others], but more attributes being established as important for everyday, leisure and tourist routes. These experiments also established the ecological validity of the computer-generated experimental approach taken, by replicating the findings of real-world research. Three further experiments then established previously unknown orders of influence or ranks. These orders were determined with some level of confidence, by relying on statistical significance within the participants' stated preference. For attractiveness and for leisure and tourist journeys the ranks are new, and for everyday journeys the number of attributes known to be influential from previous research [83] were increased.

Despite combining the results taken from 450 participants, and being based on statistical significance, there are still limitations to the ranks found by this study. Only a small number of the total possible attributes which can affect route preference were tested, with attributes chosen according to how well they could be represented as well as how important they were thought to be. Indeed, of the 15 attributes shown in Table 2.1 (Section 2.1), which is by no means exhaustive, only seven were chosen for investigation. Temporal, seasonal or weather attributes were rejected, along with many of the factors affecting aesthetics, as they cannot be determined directly from a map. However, wherever possible, the attributes considered most important by the findings of previous studies [83, 120, 121, 142, for example] were retained. Although this may mean that a highly influential factor of the route selection process was not present in the study, the data available does represent a typical real world route choice scenario. When visiting an unknown area, a person is likely to plan their route using the maps, satellite images and street views provided by online and other sources [30], indicating that the resulting routes will be dependent on the data available within these resources. Issues surrounding temporal and seasonal attributes were also minimised by providing the same description of a 'warm sunny day' to participants who asked for context in both this and the evaluation study. Although this study is not an exhaustive examination of all of the factors

contributing to route preference, it does suggest a basis for how people choose routes.

Although the virtual stated preference approach taken by this study offered many desirable characteristics, it also has limitations. The representations used to determine influence were static virtual environments, whereas real environments are dynamic and may be affected by other factors. There is no direct evidence available for how preferences in virtual and real environments correlate, and users here were presented with choices that may not be available in real-world situations. Despite these drawbacks the results indicated that that the preferences given by participants for this study are comparable with those found in real world approaches (where available).

8.1.2 Developing a representation of the environment

Using the results and relationships found by the user experiments which established how different levels of attributes affect route simplicity and attractiveness (described in Chapter 3), combined with findings from previous research [191, 203, 210, 222], this research has defined and built a suitable environment model. That model, which is based on a graph annotated with the attributes shown to be influential, is simple enough to be applicable to many geographical areas, but detailed enough to allow route selection. After comparing different digital map formats, OpenStreetMap was chosen as the most appropriate to form the basis of this representation, and the model was designed to be created automatically from OpenStreetMap data if the data is complete, or manually annotated if not.

Annotated graphs are a popular environment model for route selection [13, 31, 42, 57, 93, 95, 147, 195, 228, and many others], but the novelty of the model described by this thesis comes from the sheer number of attributes incorporated into it, and the individual approaches taken to determine the values to be recorded. Many of the models used previously contain only one or two attributes, and although earlier work has established approaches for defining a small number of attributes, such a comprehensive investigation into suitable ways of representing different attributes has not previously been attempted.

Despite its simplicity and suitability for route selection purposes, this model also has its limitations. The open-source nature of OpenStreetMap data suggests there may be inaccuracies in the map, and although errors were not a serious problem in this instance, even for the small test area chosen the data was incomplete. Manual annotation for the test area used in this research took less than three days, but in principle the required OpenStreetMap tags could be added by crowd sourcing, or automated satellite image analysis [111, 149, for example] used to establish attribute values.

Although the relationships underlying this model were based on empirical evidence, with many subjective attributes that limit the definition of clear thresholds, manually annotating the environment model was susceptible to possible limitations. It was likely to be affected by the views of the person doing it, and different individuals may have mapped the environment in different ways. In addition, the university campus was considered to be a static environment throughout this research, and although the attribute values were accurate at the time of mapping, and they may have changed or become obscured over time. An example of this are the recent changes due to the construction of a multi-storey car park, which has not only removed vegetation from an area of campus, but has also temporarily blocked access in the vicinity. As only a small area was chosen for testing in this research, the diversity of attributes and values will also have been restricted. There were no easy solutions to these issues, but participants were advised to consider their routes before this construction work began.

8.1.3   Calculating transformation functions

Combining the environment model and relationships taken from the empirical study, this research has constructed simple algorithms, based on Dijkstra's algorithm and a label correcting extension to this, to suggest suitable routes which attempt to maximise or minimise each individual attribute. These algorithms do not offer a solution for any journey type on their own, but they do give an indication of the performance of such a simple approach.

Using the same model and influence ranks taken from the study, this research has also constructed simple algorithms, again based on Dijkstra's algorithm, to suggest routes suitable for each attribute category and journey type. Each of these require different multi-attribute weighted cost functions, and time consuming algorithms to generate the associated weights, but all resulting algorithms apply the same simple approach to minimise route cost. By avoiding the complexity of previous approaches, these algorithms can be widely used in power conscious environments (i.e., in mobile apps), and easily extended for different groups of users. In addition, the runtime for each cost function indicates that all can produce routes for 174 test point pairs in 0.016 seconds or less. Although this may vary depending on the points and environment chosen, it suggests that the algorithms could be used in real-time to select appropriate routes for each journey type.

Restricting the algorithms to be minimum cost approaches did produce simple solutions; however, it also introduces its own problems. Despite the area chosen for route

selection being small, the algorithms created had heuristic limitations due to the need for the route selection problem to be tractable, and the results for leisure and tourist journeys suggest that users may have to settle for routes that are longer than might be considered suitable for pedestrian travel. As no tests were run on larger environment models, there is also no evidence that the algorithms will scale well. With only one area being used to generate and evaluate the performance of these algorithms, biases may have also been introduced by artifacts of this particular test area. If this proves to be the case then the weights generated may not be applicable to other areas, or alternative cost functions may provide better solutions. In addition, the route selection process is dependent on the quality and accuracy of the environment model and the attributes chosen to be included, and may only be applicable to the geographical area used for testing. Further investigation will be required to determine if these limitations are problematic, and to find appropriate solutions to these issues.

The limitations on this work are not restricted to the resulting cost functions, but also the process used to find them. The algorithm used to determine the 'best' weights uses metrics (such as similarity) that may not be the most appropriate, and applies no weighting to the attributes according to preference rank. However, due to a lack of existing guidance about how routes should be evaluated, these metrics were thought to be a sensible approach given the data used by the model. It is also possible that the weight combinations have been allowed to converge on a local maximum rather than the best possible solution, although it was hoped that the use of ranges rather single values would avoid this.

8.1.4   Empirically evaluating the routes suggested by the algorithms

In order to collate data to test the performance of the journey type algorithms, this research conducted a user study to collect participant route suggestions for each attribute category and journey type, and indications of their familiarity with the test area. As far as the author is aware this approach has not been taken before, but was devised to allow participants freedom to select any route, rather than offer an opinion on the route chosen by each algorithm.

Comparison between the algorithm route suggestions and those of the participants indicated that the simplicity and everyday algorithm routes are suitable (and the most preferred for the majority of test points), but that the shortest length approach gives the same or sometimes better results. Most importantly, although the routes produced for the leisure and tourist journeys are not even near matches for participant suggestions, this

research has constructed algorithms which select routes that are likely to be appropriate for the majority of users.

The maps used for evaluation purposes here may have influenced the routes selected as some attributes (such as length and turns) are more obvious than others (for example vegetation or dwellings), and therefore placed limitations on the validity of this approach. Similar issues exist with other areas of the study, with the satellite images used only being accurate when they were taken, and the instructions issued possibly influencing the information given in response (for example the tourist journey instructions asked for lists of points of interest). In addition only a small number of participants (14) were attracted to the study, and only start-end pairs (5) and a single geographic area were included in the dataset, limiting the route suggestions available for an evaluation baseline. As there is no currently accepted methodology for evaluating routes in this way, it is unclear whether a different approach, such as asking participants to offer their opinion on the routes selected by the algorithm, would have been more appropriate or given different results. The use of 2D paper maps to represent a complicated 3D environment did lead to a number of errors in route suggestions, which had to be corrected, and the resulting route may not truly represent the opinions of the participant.

Not all participants were familiar with the test points chosen, or all accessible areas of campus, indicating that their knowledge of the selected environment was limited. This is likely to be the case for many individuals passing through any geographical location, and all participants were willing to offer route suggestions even where they had no experience of either the start or end point that was presented. All participants also indicated that they were familiar with the majority of locations chosen, or at least with other points and areas nearby, and their level of knowledge was considered sufficient for evaluation purposes.

Limitations were also present in the method used to measure algorithm performance. Similarity calculated by route overlap is only one of many metrics which could have been used to evaluate algorithm performance, and indeed route similarity overall. Unlike previous route comparison approaches [57,93,187], this research aimed to evaluate routes in terms of overall suitability rather than the performance of a single attribute. Similarity by route overlap was chosen as it was considered a sufficient measure of this suitability, and with no accepted approach to evaluating routes in this way, it is more appropriate than the metrics suggested previously [57, 93, 187].

Finally the variations observed in the attribute characteristics of the algorithm selected routes may be more influential than anticipated, and may make these routes inappropriate in some cases. The concept that 'good' rather than 'the best' routes were being selected by participants was investigated by examining route cost differences, and although this did

suggest that some reduction in the highest ranking attributes was likely to be acceptable to users, this may not be the most suitable method for evaluating whether or not routes are actually appropriate. In addition overall trends in attribute characteristics were examined to determine if these routes were 'good enough', and individual routes may have attribute values that do not meet this criteria.

8.2 Future Work

Although this research has made substantial progress in the development of a route suggestion system, there are still many areas where it could be extended and improved. Some of these improvements could be used to address the current limitations, and others to improve the routes suggested. In general, the main considerations for future work can be divided into three categories; extending the environment, comprehensive route evaluation and adapting the algorithm.

8.2.1 Extending the Environment

The most obvious extension to the scope of this research would be the use of other geographical areas. Firstly this could be used to increase environment diversity. By selecting very different urban areas, both from the initial test area and from each other, the applicability of the environment model and performance of the algorithms could be examined to establish if they generalise across many environments. Secondly, increasing the size of the areas chosen will help to determine the scalability of the system. Initially the current test area could be expanded to cover the whole of Leeds city centre, rather than just the University of Leeds campus. As the environment model and algorithms are designed to be applicable in many locations, from urban to semi-rural, then this approach could also be utilised for any village, town or city area, assuming that the scale is still suitable for pedestrian travel. Increasing the areas tested would also allow other groups to participate in creating model and test datasets, but would require that OpenStreetMap data was correctly and completely tagged. As discussed previously this could be done by crowd sourcing, either by direct recruitment of volunteers or by encouraging those that already create and update OpenStreetMap, to include the required tags.

Crowd sourcing also offers access to other types of information that could be incorporated into the route selection process, which is not currently available. The existing system is restricted to the use of attributes which can be determined directly from maps, but given a suitable interface, communities, industry and local government could be en-

couraged to annotate area maps with local knowledge. Temporal attributes, difficult areas in adverse weather and the importance (or popularity) of points of interest (amongst others) could all be determined in this way, and many geographical areas could be populated simultaneously. Alternatively, much of this information could be mined from social media sources. The resulting annotated map could then be easily converted into an environment model similar to that described by this research.

Landmarks and familiarity are two further attributes which have been shown to be important, not only in route selection, but also in forming a cognitive map of an area [6, 142, 150, 199]. In order to make this system applicable to a wide range of users, including both of these attributes as a second layer of preference would be desirable, but neither is as straightforward as the attributes discussed so far. Although landmarks could be added to a map and therefore the environment model, issues with defining suitable visual cues and when they are required would need to be overcome. In contrast, despite the ease with which familiarity could be added to the existing algorithms, it is dependent on individual use and would require data associated with known routes to be stored and accessed from a database. Solving these problems and allowing familiarity and landmarks to be included could open this algorithm up to users who have issues forming cognitive maps, and is a vast and exciting area for future work.

8.2.2 Comprehensive Route Evaluation

One of the most surprising findings of this research is the size of the variations observed in participant-suggested leisure and tourist routes. Without a baseline to evaluate against, or a large dataset with which to determine relationships, creating an accurate route suggestion engine for these journey types is near impossible. Crowd sourcing again seems the most obvious solution to collating a route suggestion dataset, but this comes with its own difficulties. Crowd sourced experiments are essentially unsupervised, with participants being unable to ask for clarification or more information directly during each task. This lack of direct contact may also be detrimental to the experimenter as, without trawling through hours of audio recordings, the underlying motivations for route choice may be lost. In addition, participants' level of familiarity with the the environments being tested would have to be established, which also requires careful consideration during the design of each experiment. With a large number of routes being suggested, possibly across many environments, a suitable method for recording these, and automatically transferring the resulting data into the evaluation system, must be found.

Adapting the existing study to be one which can be completed online is one approach

to crowd sourcing. Initial questions could be used to establish familiarity, either with the Likert scale method used in this research, or by asking participants to identify relatively well-known (but unlabelled) points of interest on a map of the test area. With a suitable interface, routes could then be suggested by highlighting the required links on a map. This approach could be used to reduce the errors arising from the use of a 2D map, firstly by only allowing selection of valid links, but also by allowing access to other information sources such as 3D representations or other views. Although this approach is fairly straightforward, it may provide a comparatively small number of routes per participant compared to other acquisition techniques.

Another approach to gathering route suggestions would be through GPS tracking using devices such as mobile phones. In this case familiarity could be assumed, as the participants is actually walking the route that they are suggesting. Furthermore, routes could be generated automatically by mobile service providers to produce a large dataset, assuming legal issues associated with data protection could be overcome. However, current GPS devices are prone to inaccuracies in urban environments, data collection may take days or even weeks, and recruitment of participants may be difficult if they are uncomfortable with being remotely tracked. Assuming that enough suitable participants could be recruited, then this large dataset could be examined for consistency or patterns to aid with the construction and evaluation of future route suggestion algorithms.

8.2.3 Adapting the Algorithm

In addition to the expansion of the test area, and inclusion of supplementary attributes, more research is required into the route suggestion engine itself. All of the multi-attribute algorithms (with the exception of simplicity) suffer from the heuristic limitations of the approach used, and developing algorithms which guarantee to give the best solutions in polynomial time is essential. One possible solution to this may be to adjust the label correcting algorithm variant (described in Section 5.3.2.1) to be closer to the Bellman-Ford [16] approach, although this may produce algorithms that are no longer tractable. Alternatively, examination of minimum spanning tree algorithms such as the 'soft heap' approach [39] may provide more computationally efficient solutions. One of the drawbacks of the environment representation used in this research is that not only does it produce a relatively sparse undirected graph, but that it also contains links with negative weights that can only be generated at runtime (making it difficult to apply Johnson's algorithm to remove these weights). The search for an efficient algorithm to select the minimum cost routes through graphs of this type is still ongoing, and future developments

may offer suitable solutions.

In addition to increasing the efficiency and performance of the route selection algorithms, if the variation across the leisure and tourist routes remains even with large route datasets, then an amount of randomisation may also be required to successfully produce appropriate solutions. An example of this may be randomly selecting from the ten best routes suggested by the respective algorithm.

Alternatively, personalising the system to the user may improve the suitability of the routes suggested by the leisure and tourist algorithms. Section 7.3.5 indicated that allowing the user to select route length may improve the performance of the system for these types of journey, without greatly increasing the complexity of the approach. Extending this to allow users to select which points of interest they wish to include on their routes, making the subset allowed for route choice specific to the individual, may also prove fruitful. In addition, developing a simple method for incorporating the users prior knowledge of the area, or familiarity with routes through the environment, may result in substantial improvements in the suitability of the routes suggested. Each of these adaptations could be performed with relatively little increase in system complexity, and each deserves further investigation.

Appendix A

Influence of Environment and Route Attributes on Route Preference: User Study Images

This appendix shows the basic layouts that were combined to form the screens displayed in the user study described in Chapter 3.

	Level 1	**Level 2**	**Level 3**

Decision Points

(a) 1 Decision Point (no turn)

(b) 3 Decision Points (no turns)

(c) 5 Decision Points (no turns)

Turns

(d) 1 Turn (at decision points)

(e) 3 Turns (at decision points)

(f) 5 Turns (at decision points)

Landmarks

(g) 1 Landmark

(h) 3 Landmarks

(i) 5 Landmarks

Longest Leg First

(j) 1 Unit LLF

(k) 3 Units LLF

(l) 5 Units LLF
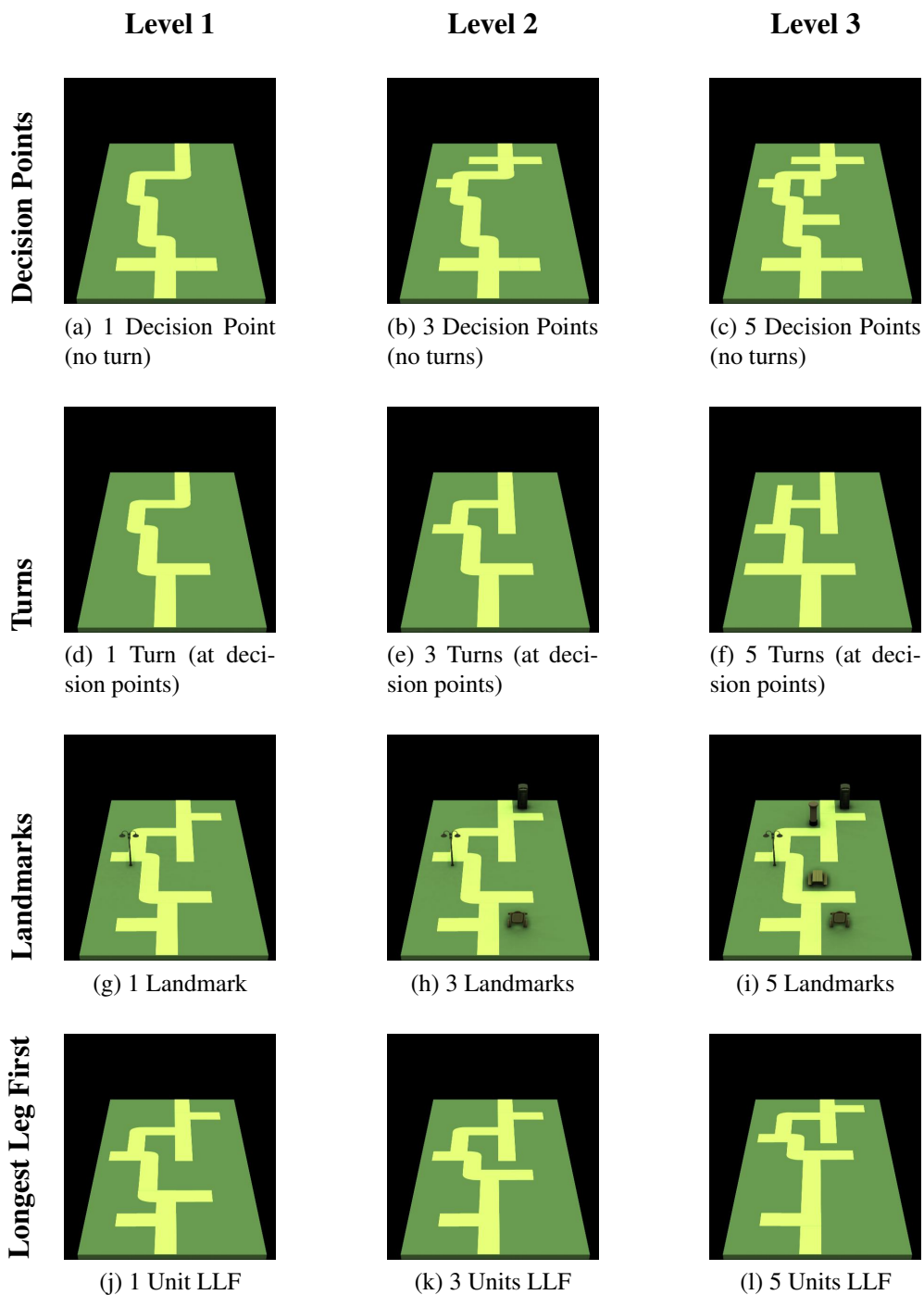
Figure A.1: Simplicity artificial environments displayed for the user study described in Chapter 3. Each image shows a single attribute level, and the features used to represent it.

Level 1	Level 2	Level 3

Vegetation

(a) 1 Unit of Vegetation

(b) 3 Units of Vegetation

(c) 5 Units of Vegetation

Land Use

(d) Urban

(e) Farmland

(f) Parkland

Dwellings

(g) Multi Occupancy Dwellings

(h) Single Occupancy Dwellings

(i) Historic Dwellings

Points of Interest

(j) 1 Point of Interest

(k) 3 Points of Interest

(l) 5 Points of Interest
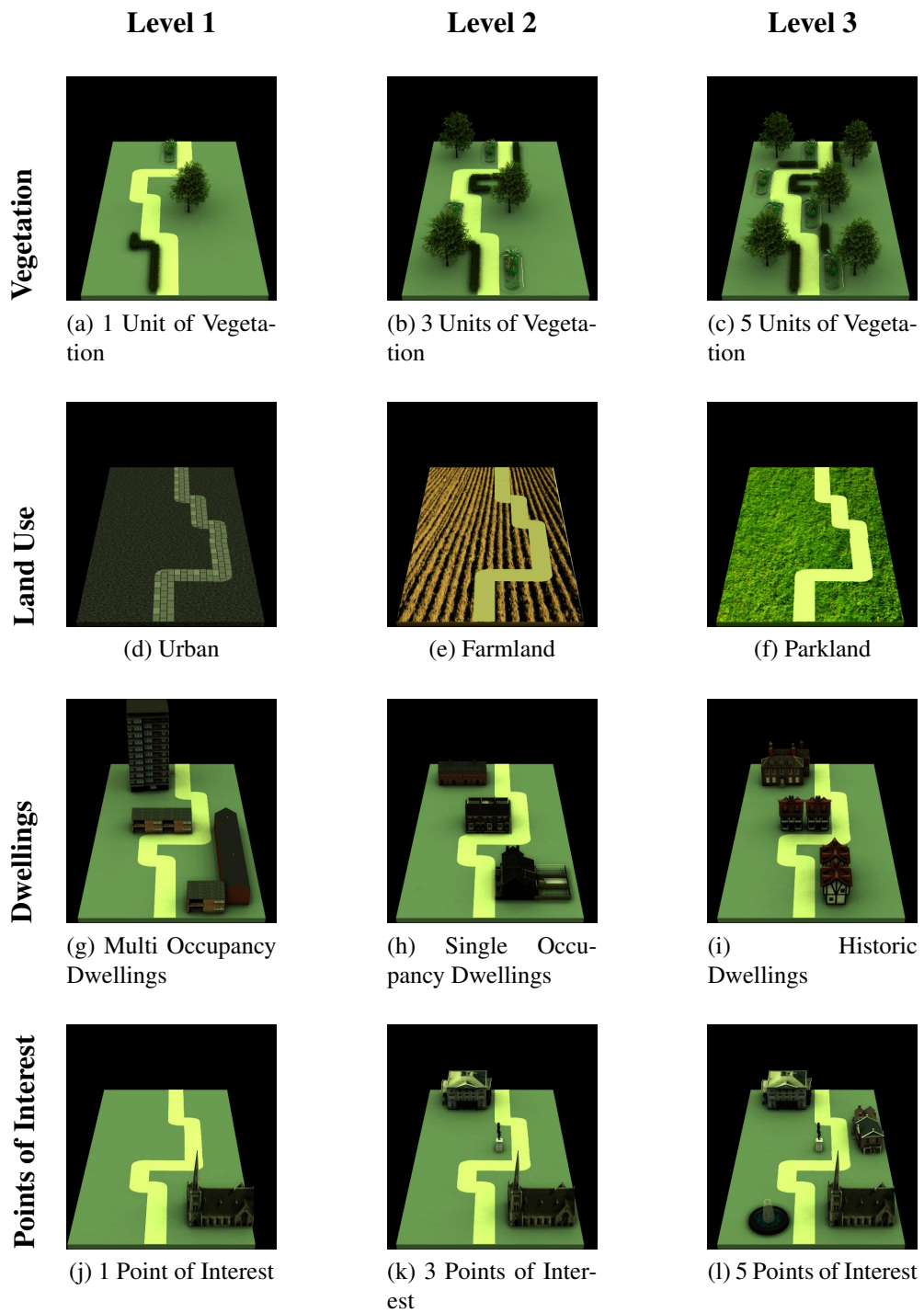
Figure A.2: Attractiveness artificial environments displayed for the user study described in Chapter 3. Each image shows a single attribute level, and the features used to represent it.

Appendix B

Evaluating the Routes: User Study Materials

This appendix shows the materials used in the user study described in Chapter 7.

B.1   Printed Materials

This section shows the printed materials that were used for reference by the participants.

Participant Information Sheet

Research Project: A Cognitive Solution for Real-World Pedestrian Wayfinding Problems

Experiment 2: Investigate the routes suggested for different wayfinding scenarios.

This sheet will hopefully provide you with enough information about the study to allow you to make an informed decision about participation. However, if you have any questions or would like to discuss anything with me please let me know.

<u>What is the project's purpose?</u>

I am a PhD Student in the School of Computing at the University of Leeds. My research aims to create a system which can suggest routes through an unknown or partially known environment, for a variety of different purposes (ie a stroll for leisure or daily journey to work). It seeks to not only provide routes in these situations, but also encourages the user to learn some information about their location and the environment around them.

<u>Do I have to take part?</u>

This research is subject to ethical guidelines set out by the University. These guidelines include principles such as obtaining your informed consent, notifying you of your right to withdraw at any time, and protection of your anonymity. You should not take part in this study if you suffer from visual impairment, as the format of the data presented here is not appropriate for your needs.

<u>How do I give my consent to participate in this research?</u>

You will be asked to sign a consent form before the study begins. In addition, consent will be assumed if you submit your route suggestions at the end of this study. You may leave the room before the study begins or fail to submit any suggestions if you do not wish to participate in this research.

<u>What will happen to me if I take part?</u>

The experiment will consist of a recorded test lasting approximately and hour. During this you will be provided with a number of scenarios consisting of a start and end point, shown on a map, and a common wayfinding problem. For each scenario, you will be asked to think of a route between the two given points, according to the specified problem. Materials will be provided for you to produce instructions on how to follow this route, and you will have as much time as required. If you cannot suggest a suitable route, then please indicate this by saying or writing "don't know", and the experimenter will move on to the next scenario.

The experiment procedure is as follows:

- Experimenter will explain how the experiment will work, and ask for your consent to participate.
- A scenario will be provided for you to practice on.
- 19 scenarios will then be given to you and materials provided with which you can produce your route instructions.
- The experiment is then complete and you will be invited to submit your route instructions.

What if I decide to withdraw before the end of the study?

If you wish to withdraw before the study is complete, please indicate this to the experimenter and the study will be stopped. You will not be asked to submit your route instructions, and any recordings made during your participation in the study will be deleted. As all submissions are anonymous, it is not possible to withdraw your consent after your route instructions have been submitted.

What will happen to the route instructions after I submit it?

The data you have provided will be added to that of the other participants and analysed at a later date.

What are the possible disadvantages and risks of taking part?

There are no risks from taking part in this research other than those from reading from the screen, as you would normally during a lecture.

What are the possible benefits of taking part?

Each participant will be reimbursed for their time according to ethical guidelines, and you may benefit from the experience of taking part in a formal user study (e.g., to help design such studies of your own).

Will my taking part in this project be kept confidential?

The information that we collect from you during the course of the research will be anonymous and you will not be able to be identified in any reports or publications

What will happen to the results of the research project?

The research may be reported at academic conferences and in academic journals, but no-one should be able to identify you and at no point will your identity be divulged.

Contact Information

Lead investigator: Sarah Cook   (sc10sc@leeds.ac.uk)

Address: School of Computing

                        University of Leeds

                        Leeds, LS2 9JT

Participant Instructions

During this experiment you will be given four scenarios, and five pairs of start and end points shown on a map of campus. The four scenarios are as follows:

1. Simplicity – "Imagine that someone has approached you and asked you for the simplest route from point A to point B. Please describe the route you would suggest."
2. Everyday – "Imagine that you were giving instructions to someone who needed to travel from point A to point B every day, for example as part of their daily commute to work. Please describe the route you would suggest."
3. Leisure – "Imagine that you were giving instructions to someone who had some free time and wanted to go for a stroll from point A to point B. Please describe the route you would suggest. (You can assume that they have approximately 30 minutes free)"
4. Tourist – "Imagine that you were giving instructions to a tourist who wanted to go on a tour of the campus sights, starting at point A and ending at point B. Please describe the route you would suggest."

For each scenario and pair of points, you will be asked to supply instructions which would enable this person to travel from start to end along your suggested route correctly.

To enable you to do this you will be supplied with the following materials:

- A large campus map with the start and end points marked.
- Photos of the start and end points.
- Paper copies of the campus map on which you can mark your route.
- Paper for either written instructions or to sketch your route.
- Writing materials.

Feel free to write, draw or speak the instructions, or any combination of these, but please provide enough information for the route to be clear.

If you can't provide a route, then please say or write "Don't know".

You will also be asked to answer three questions about your route, and one on each scenario, provided on a separate sheet.

There are no right or wrong routes, it is personal preference, and you may ask any questions that arise as the experiment proceeds.
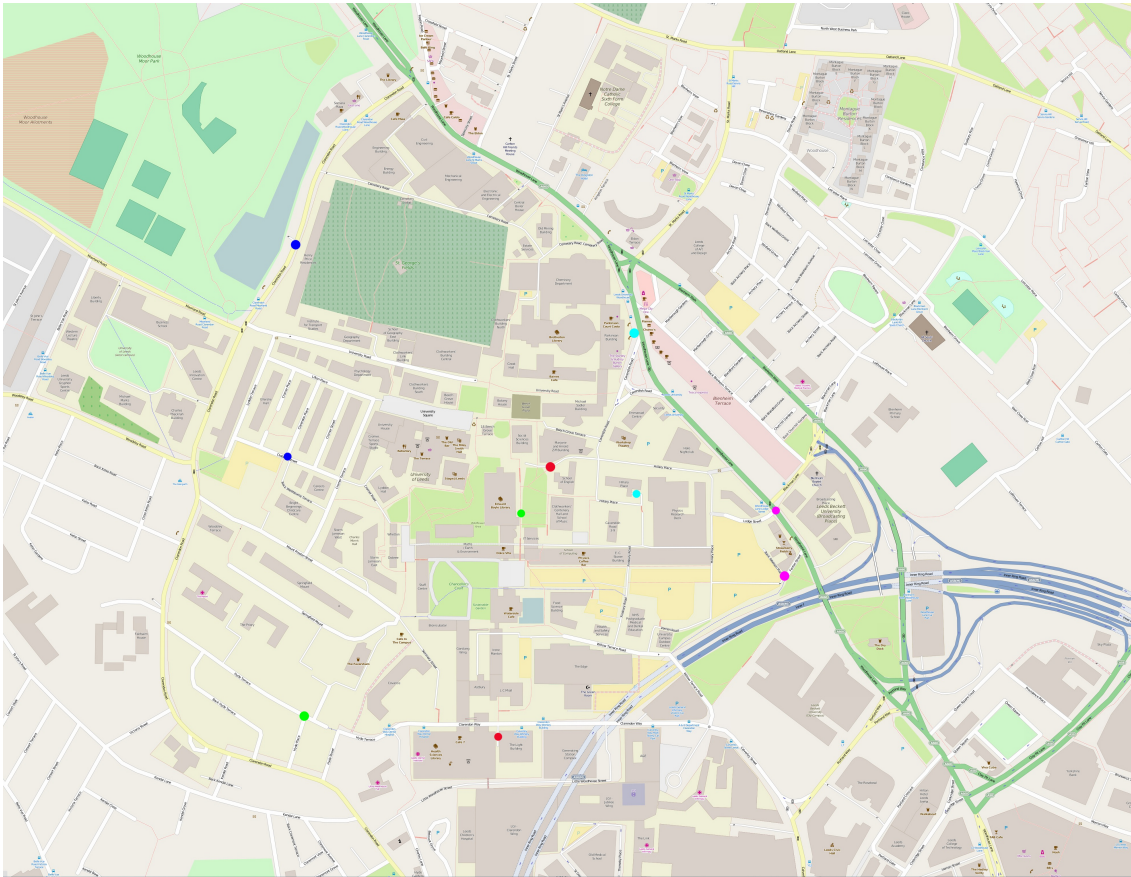
Figure B.1: Image used for the A2 laminated map for the user study described in Chapter 7.

B.2 Displayed Images

This section shows the displayed images that were used for reference by the participants.

Simplicity


A

Henry Price
Residences

B
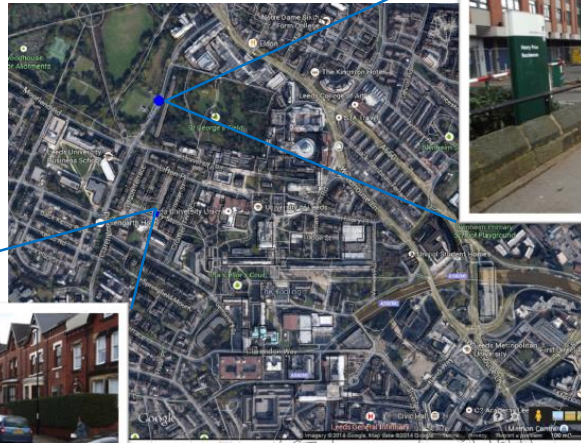
Outside Bright Beginnings Nursery

1

Simplicity

A

Parkinson
Steps

B
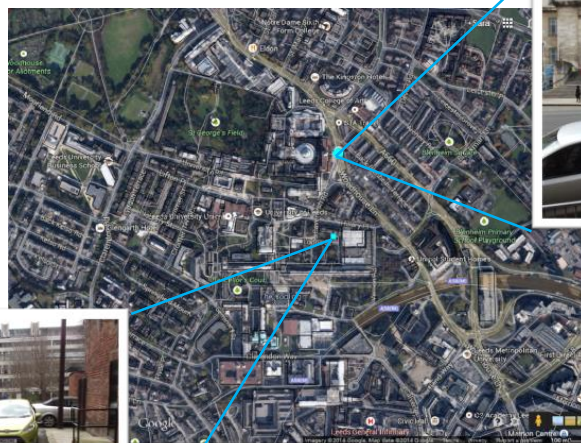
Opposite Physics Deck (Hillary Place)

2

Simplicity

B
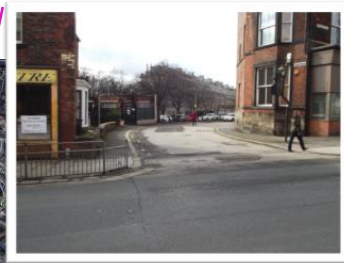
Print Shop
(Woodhouse
Lane)

A

Fenton Street Entrance

3

Simplicity

A

Corner of
Ziff Building

B
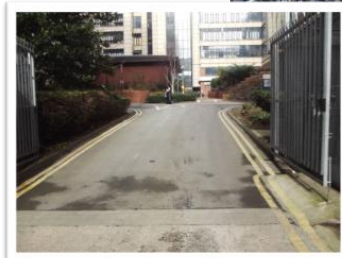
Between Worsley and Light Buildings

4

Simplicity

A

Behind the Faversham

B

Between Edward Boyle and Maths

5

Everyday

A

Behind the Faversham

B
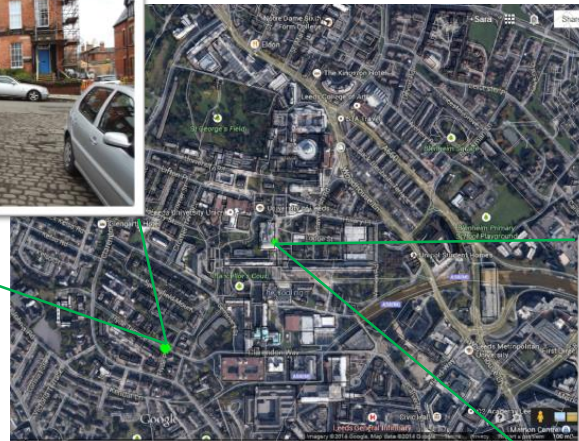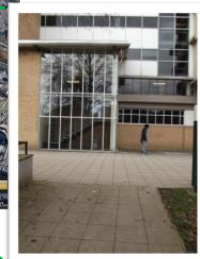
Between Edward Boyle and Maths

6

Everyday

A

Henry Price Residences

B
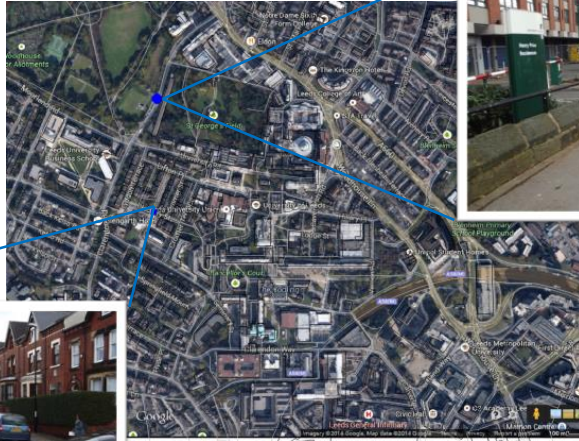
Outside Bright Beginnings Nursery

7

Everyday

A

Parkinson Steps

B
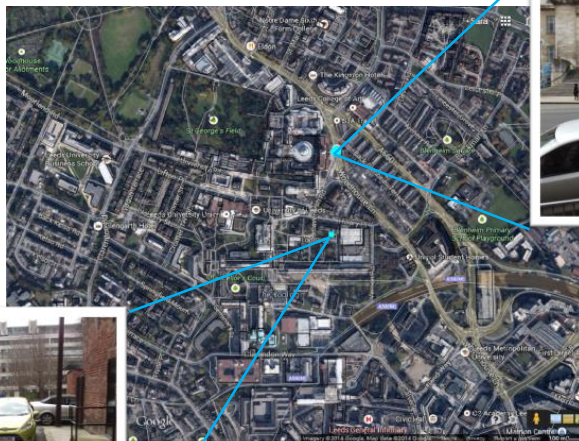
Opposite Physics Deck (Hillary Place)

8

Everyday

B

Print Shop
(Woodhouse
Lane)

A
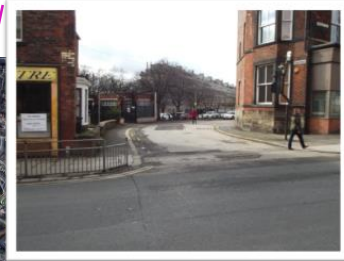
Fenton Street Entrance

9

Everyday

A

Corner of
Ziff Building

B
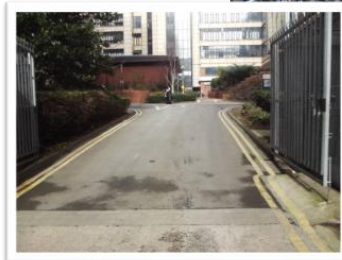
Between Worsley and Light Buildings

10

Leisure

A

Parkinson Steps

B
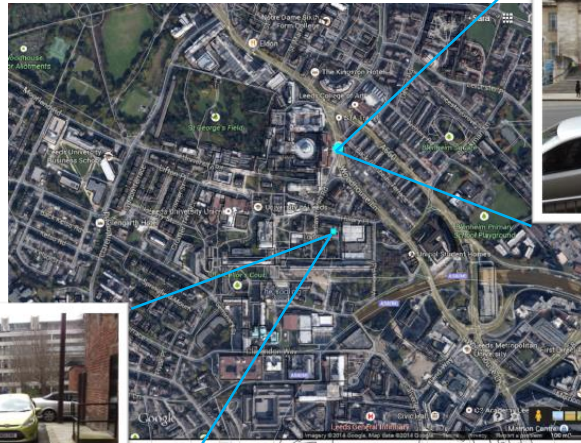
Opposite Physics Deck (Hillary Place)

11

Leisure

B

Print Shop (Woodhouse Lane)

A
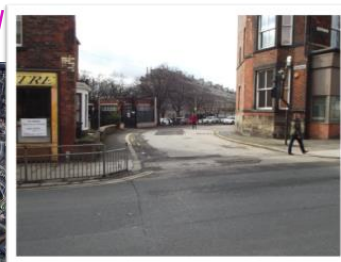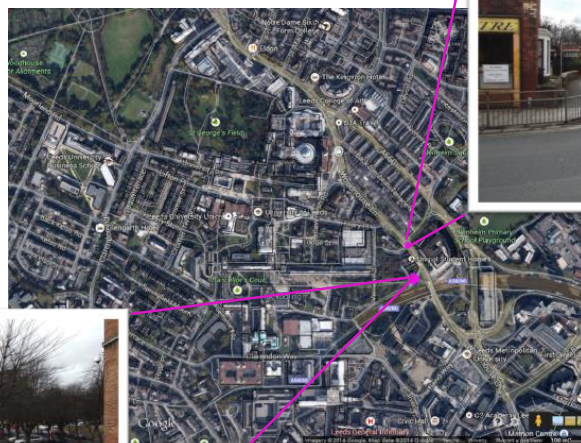
Fenton Street Entrance

12

Leisure

A

Corner of
Ziff Building

B
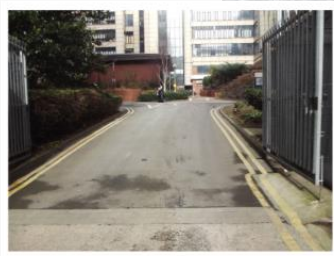
Between Worsley and Light Buildings

13

Leisure

A

Behind the Faversham

B
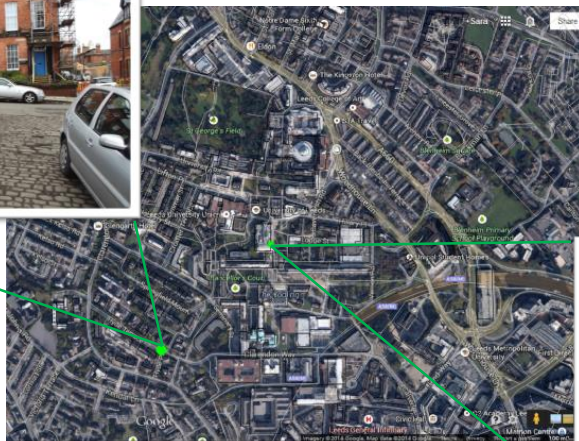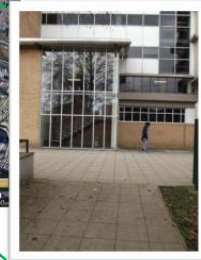
Between Edward Boyle and Maths

14

Leisure

A

Henry Price
Residences

B
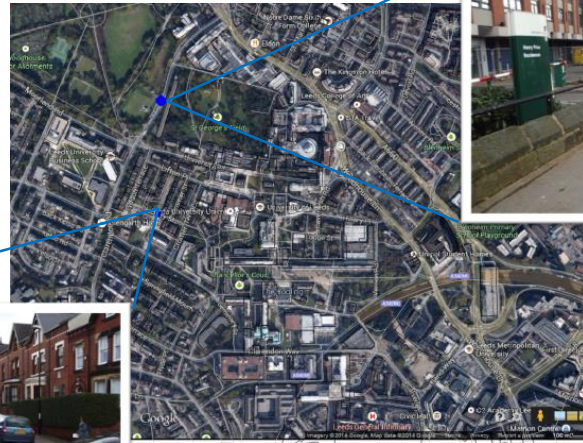
Outside Bright Beginnings Nursery

15

Tourist

B

Print Shop
(Woodhouse
Lane)

A
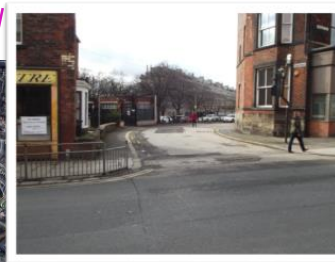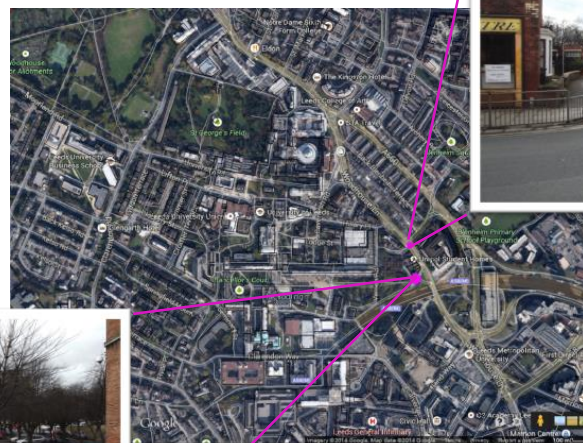
Fenton Street Entrance

16

Tourist

A

Corner of
Ziff Building

B
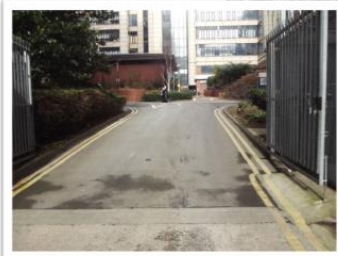
Between Worsley and Light Buildings

17

Tourist

A

Behind the Faversham

B
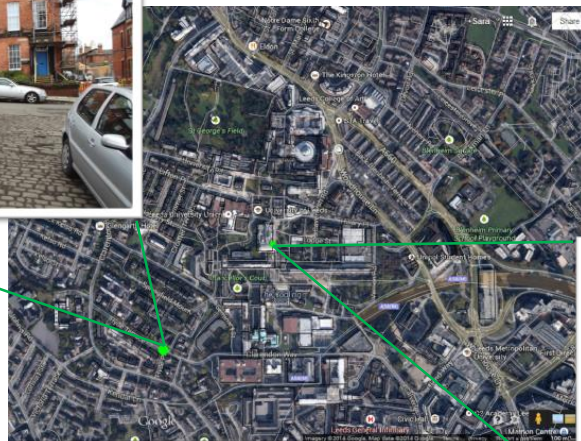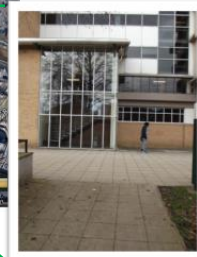
Between Edward Boyle and Maths

18

Tourist

A

Henry Price
Residences

B
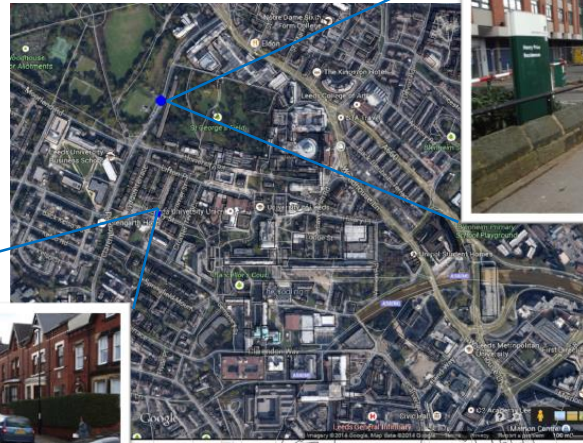
Outside Bright Beginnings Nursery

19

Tourist

A

Parkinson
Steps

B
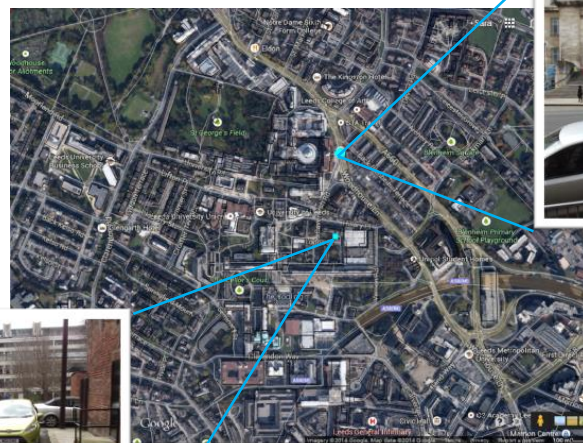
Opposite Physics Deck (Hillary Place)

20

B.3 Participant Response Materials

This section shows the printed materials that were used by the participants to provide their responses.

N
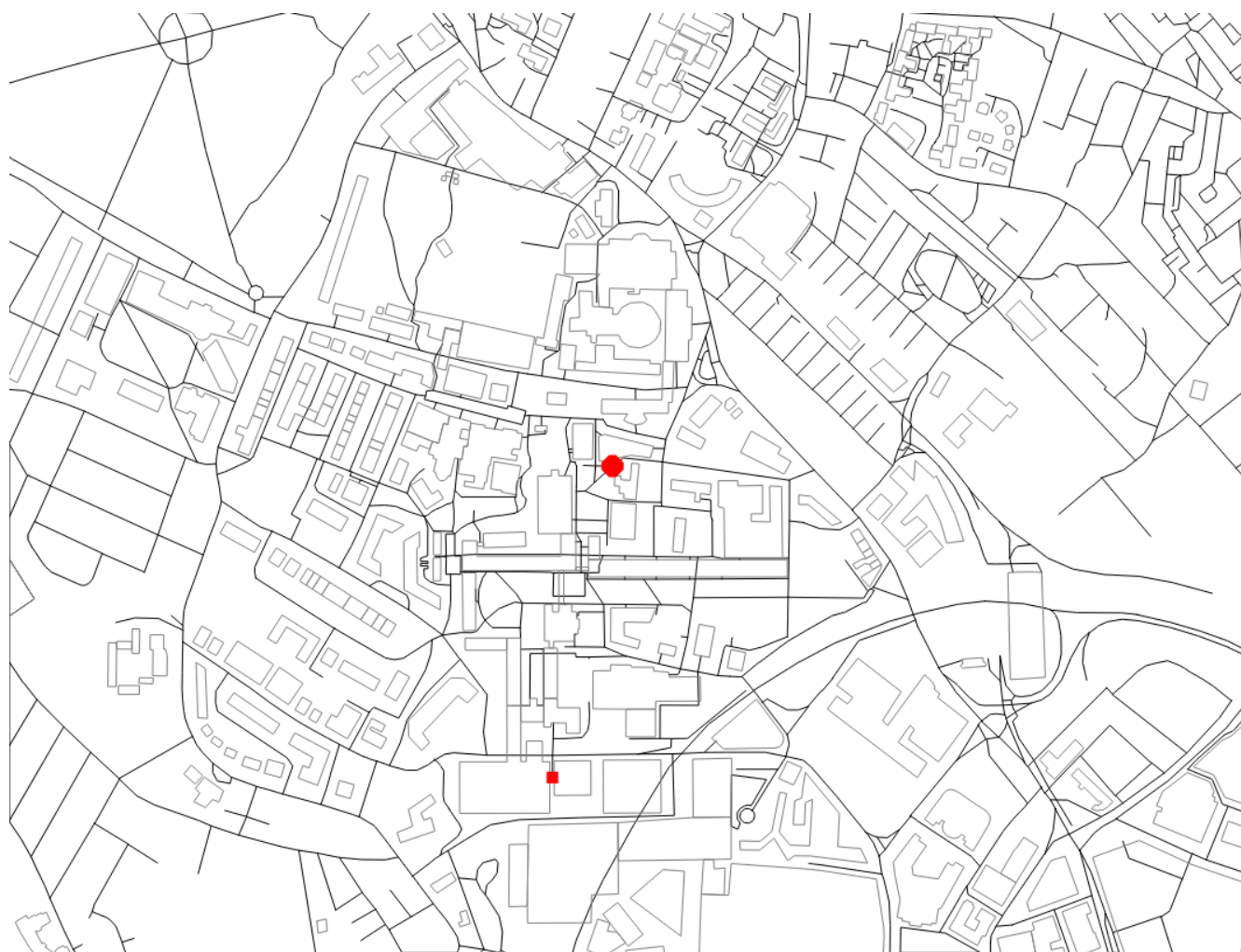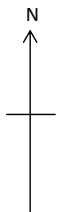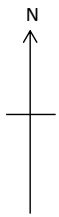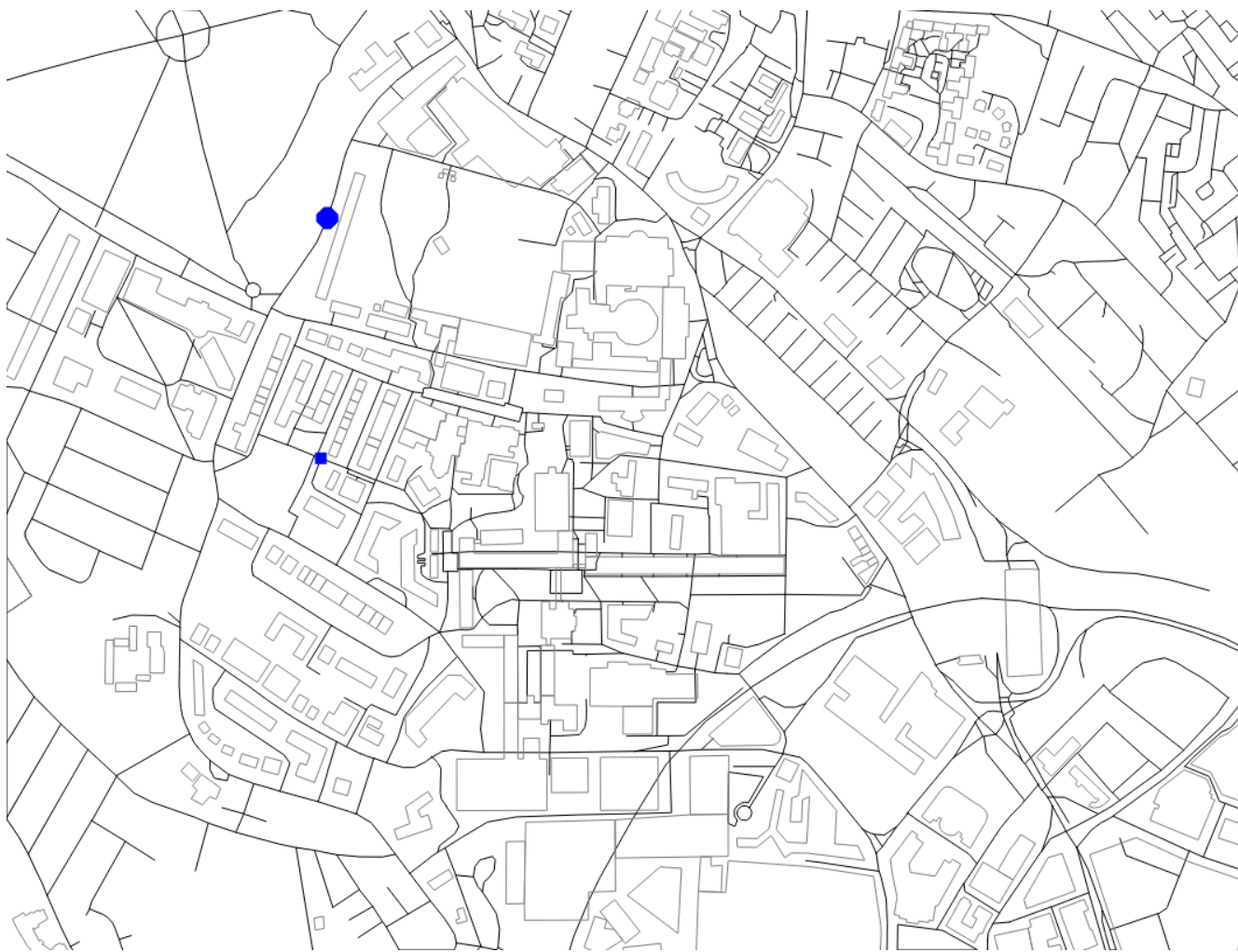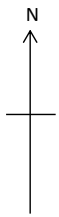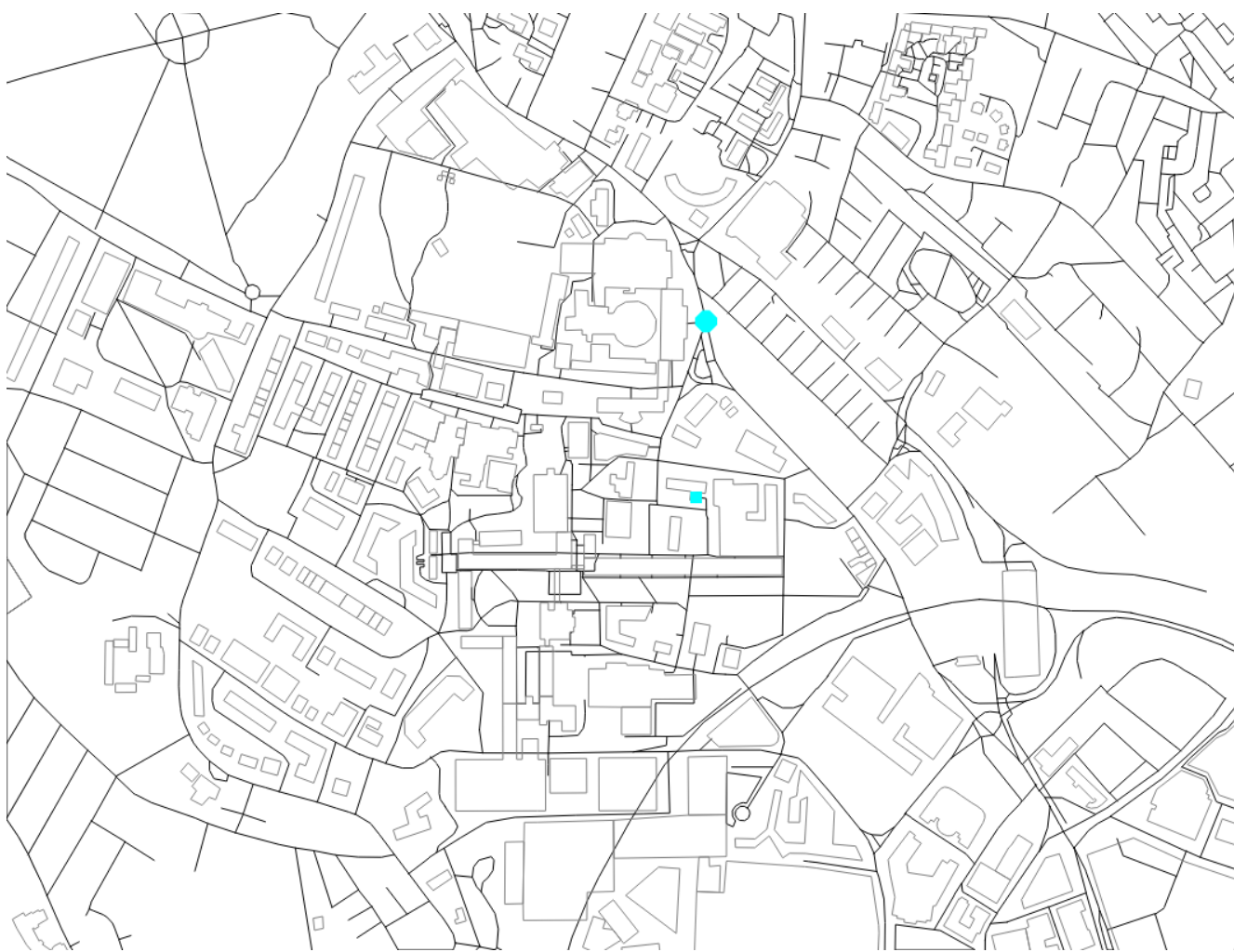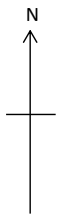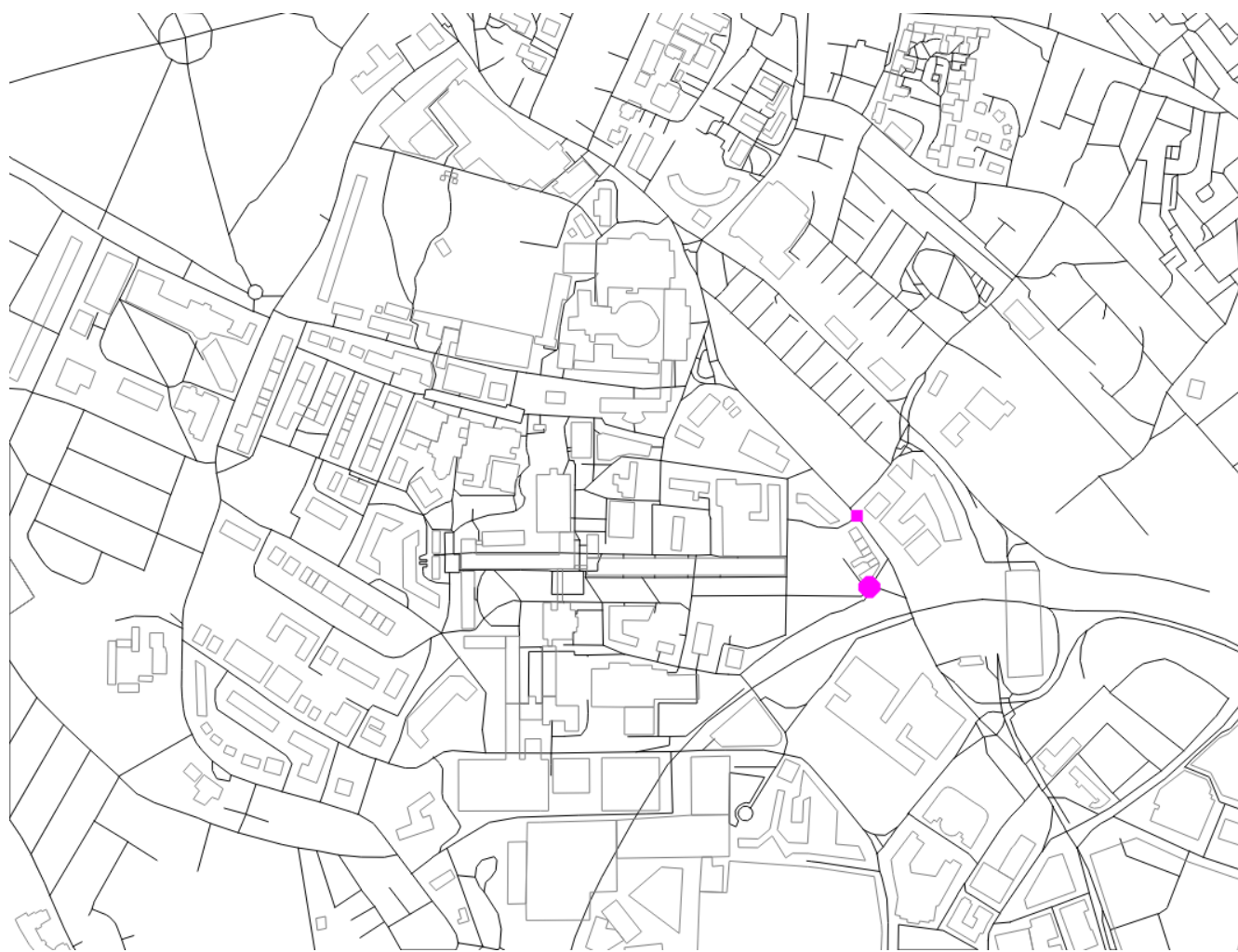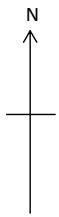
Age ☐   **Gender**   **M** ☐   **F** ☐

1. **How familiar are you with the :**

   Start point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

   End point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

   Route

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

2. **How familiar are you with the :**

   Start point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

   End point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

   Route

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

3. **How familiar are you with the :**

    <u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

    <u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

    <u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

4. **How familiar are you with the :**

    <u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

    <u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

    <u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

5. **How familiar are you with the :**

Start point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

End point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

Route

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

For this scenario, please give your reasons for choosing the routes:

6. **How familiar are you with the :**

    <u>**Start point**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

    <u>**End point**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

    <u>**Route**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

7. **How familiar are you with the :**

    <u>**Start point**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

    <u>**End point**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

    <u>**Route**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

8. **How familiar are you with the :**

   <u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

   <u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

   <u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

9. **How familiar are you with the :**

   <u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

   <u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

   <u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

10. How familiar are you with the :

Start point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

End point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

Route

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

For this scenario, please give your reasons for choosing the routes:

11. How familiar are you with the :

<u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

<u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

<u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

12. How familiar are you with the :

<u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

<u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

<u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

13. How familiar are you with the :

    <u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

    <u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

    <u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

14. How familiar are you with the :

    <u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

    <u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

    <u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

15. How familiar are you with the :

<u>**Start point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

<u>**End point**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

<u>**Route**</u>

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

For this scenario, please give your reasons for choosing the routes:

16. How familiar are you with the :

Start point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

End point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

Route

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

17. How familiar are you with the :

Start point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

End point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

Route

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

18. How familiar are you with the :

Start point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

End point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

Route

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

19. How familiar are you with the :

Start point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

End point

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

Route

Very familiar	Somewhat familiar	Not very familiar	Not familiar
☐	☐	☐	☐

20. How familiar are you with the :

<u>**Start point**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

<u>**End point**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

<u>**Route**</u>

Very familiar	**Somewhat familiar**	**Not very familiar**	**Not familiar**
☐	☐	☐	☐

For this scenario, please give your reasons for choosing the routes:

Bibliography

[1] OpenStreetMap contributors. OpenStreetMap. `http://www.openstreetmap.org/#map=17/53.80564/-1.55347`, 2014.

[2] G. Abowd, C. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. Cyberguide: A mobile context-aware tour guide. *Wireless Networks*, 3(5):421–433, 1997. 10.1023/A:1019194325861.

[3] T. Achterberg, T. Koch, and A. Martin. Branching rules revisited. *Operations Research Letters*, 33(1):42–54, January 2005.

[4] Y. Akasaka and T. Onisawa. Construction of pedestrian navigation system and its evaluation. In *IEEE International Conference on Fuzzy Systems*, volume 3, pages 1525–1530. IEEE, 2004.

[5] Y. Akasaka and T. Onisawa. Individualized Route Selection for Pedestrian Using Fuzzy Measures and Integrals. In *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, pages 4778–4783. IEEE, 2006.

[6] G.L. Allen. Cognitive abilities in the service of wayfinding: A functional approach. *RTPG*, 51(4):555–561, 1999.

[7] D.L. Applegate, editor. *The traveling salesman problem: a computational study*. Princeton series in applied mathematics. Princeton University Press, Princeton, 2006. OCLC: ocm83853510.

[8] M. Arikawa, S. Konomi, and K. Ohnishi. Navitime: Supporting pedestrian navigation in the real world. *IEEE Pervasive Computing*, 6:21–29, 2007.

[9] F. Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405, September 1991.

[10] Autodesk Inc. Autodesk®3ds max®2012. `http://www.autodesk.co.uk/adsk/servlet/pc/index?siteID=452932&id=14596087`, 2011.

[11] S. Bafna. Space Syntax: A Brief Introduction to Its Logic and Analytical Techniques. *Environment & Behavior*, 35(1):17–29, January 2003.

[12] J.N. Bailenson, M.S. Shum, and D.H. Uttal. The initial segment strategy: A heuristic for route selection. *Memory & Cognition*, 28(2):306–318, 2000.

[13] J. Baus, C. Kray, and A. Krger. Visualisation of route descriptions in a resource-adaptive navigation aid. *Cognitive Processing*, 2(23):323345, 2001.

[14] Bayon, D. Best satnav for walking. `http://www.alphr.com/features/381412/best-satnav-for-walking`, 2013.

[15] A.K. Beeharee and A. Steed. A natural wayfinding exploiting photos in pedestrian navigation systems. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, page 81, Helsinki, Finland, 2006. ACM Press.

[16] R. Bellman. On a routing problem. Technical report, DTIC Document, 1956.

[17] M. Bellmore and G.L. Nemhauser. The Traveling Salesman Problem: A Survey. *Operations Research*, 16(3):538–558, June 1968.

[18] M. Ben-Akiva, MJ Bergman, Andrew J. Daly, and Rohit Ramaswamy. Modeling inter-urban route choice behaviour. In *Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Press, Utrecht*, page 299330, 1984.

[19] G. Benelli, A. Bianchi, P. Marti, E. Not, and D. Sennati. HIPS: hyper-interaction within physical space. In *Multimedia Computing and Systems, 1999. IEEE International Conference on*, volume 2, pages 1075–1078, Florence , Italy, 1999. IEEE Comput. Soc.

[20] J.L. Bentley. Experiments on Traveling Salesman Heuristics. In *Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '90, pages 91–99, Philadelphia, PA, USA, 1990. Society for Industrial and Applied Mathematics.

[21] Berg Insight AB,. Mobile Navigation Services and Devices. Technical Report 7th Edition, Berg Insight, 2014.

[22] D.P. Bertsekas. A simple and fast label correcting algorithm for shortest paths. *Networks*, 23(8):703–709, December 1993.

[23] D.P. Bertsekas, F. Guerriero, and R. Musmanno. Parallel asynchronous label-correcting methods for shortest paths. *Journal of Optimization Theory and Applications*, 88(2):297–320, February 1996.

[24] I. Bishop and E. Lange. Visualization classified. *Visualization in Landscape and Environmental Planning. Technology and applications. London, New York, Taylor & Francis*, pages 23–34, 2005.

[25] I. D. Bishop and B. Rohrmann. Subjective responses to simulated and real environments: a comparison. *Landscape and Urban Planning*, 65(4):261 – 277, 2003.

[26] I.D. Bishop, J.R. Wherrett, and D.R. Miller. Assessment of path choices on a country walk using a virtual environment. *Landscape and Urban Planning*, 52(4):225 – 237, 2001.

[27] F.F. Boctor, G. Laporte, and J. Renaud. Heuristics for the traveling purchaser problem. *Computers & Operations Research*, 30(4):491–504, April 2003.

[28] B. Bontoux and D. Feillet. Ant colony optimization for the traveling purchaser problem. *Computers & Operations Research*, 35(2):628–637, February 2008.

[29] P.C. Boxall, W.L. Adamowicz, J. Swait, M. Williams, and J. Louviere. A comparison of stated preference methods for environmental valuation. *Ecological Economics*, 18(3):243 – 253, 1996.

[30] B. Brown and M. Chalmers. Tourism and mobile technology. In *Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work*, pages 335–354, Norwell, MA, USA, 2003. Kluwer Academic Publishers.

[31] D. Caduff and S. Timpf. The Landmark Spider: Representing Landmark Knowledge for Wayfinding Tasks. In Hegarty and R. Lowe, editors, *Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance*, pages 30–35. AAAI Press, Stanford, CA, USA, 2005.

[32] E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta. A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks. In *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, page 697711. Pergamon, 1996.

[33] B. Chakraborty and T. Hashimoto. A framework for user aware route selection in pedestrian navigation system. In *Aware Computing (ISAC), 2010 2nd International Symposium on*, pages 150–153, Tainan, 2010. IEEE.

[34] B. Chandra, H.J. Karloff, and C.A. Tovey. New Results on the Old k-Opt Algorithm for the TSP. In *SODA*, pages 150–159, 1994.

[35] Y.-J. Chang, Y.-Y. Chu, C.-N. Chen, and T.-Y. Wang. Mobile computing for indoor wayfinding based on bluetooth sensors for individuals with cognitive impairments. In *Mobile computing for indoor wayfinding based on bluetooth sensors for individuals with cognitive impairments*, pages 623–627. IEEE, 2008.

[36] Y.-J. Chang, S.-M. Peng, T.-Y. Wang, S.-F. Chen, Y.-R. Chen, and H.-C. Chen. Autonomous indoor wayfinding for individuals with cognitive impairments. *Journal of NeuroEngineering and Rehabilitation*, 7(1):45, 2010.

[37] Y.-J. Chang, S.-K. Tsai, and T.-Y. Wang. A context aware handheld wayfinding system for individuals with cognitive impairments. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 27–34, Halifax, Nova Scotia, Canada, 2008. ACM Press.

[38] Y.-J. Chang and T.-Y. Wang. Comparing picture and video prompting in autonomous indoor wayfinding for individuals with cognitive impairments. *Personal and Ubiquitous Computing*, 14:737–747, 2010.

[39] B. Chazelle. A minimum spanning tree algorithm with inverse-Ackermann type complexity. *Journal of the ACM*, 47(6):1028–1047, 2000.

[40] T. Chen. Using a weight-assessing model to identify route choice criteria and information effects. *Transportation Research Part A: Policy and Practice*, 35:197–224, 2001.

[41] B.V. Cherkassky, A.V. Goldberg, and T. Radzik. Shortest paths algorithms: Theory and experimental evaluation. *Mathematical Programming*, 73(2):129–174, May 1996.

[42] K. Cheverst, N. Davies, K. Mitchell, and A. Friday. Experiences of developing and deploying a context-aware tourist guide. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 20–31, Boston, Massachusetts, United States, 2000. ACM Press.

[43] N. Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Technical Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, 1976.

[44] N. Christofides, V. Campos, A. Corbern, and E. Mota. An algorithm for the Rural Postman problem on a directed graph. In Giorgio Gallo and Claudio Sandi, editors, *Netflow at Pisa*, pages 155–166. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986.

[45] G. Clarke and J.W. Wright. Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. *Operations Research*, 12(4):568–581, August 1964.

[46] A. Corbern and J.M. Sanchis. A polyhedral approach to the rural postman problem. *European Journal of Operational Research*, 79(1):95–114, November 1994.

[47] T.H. Cormen, editor. *Introduction to algorithms*. MIT Press, Cambridge, Mass, 3rd ed edition, 2009. OCLC: ocn311310321.

[48] G.A. Croes. A Method for Solving Traveling-Salesman Problems. *Operations Research*, 6(6):791–812, December 1958.

[49] N. Dalton, J. Peponis, and R. Dalton. To tame a TIGER one has to know its nature:extending weighted angular integration analysis to the descriptionof GIS road-centerline data for large scale urban analysis. In *4th International Space Syntax Symposium*, London, UK, June 2003.

[50] A.-M. Davies, R. Laing, and S. Scott. Combining visualisation and choice experiments in built environment research. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pages 785 – 790, 2002.

[51] C. Davies and D. Peebles. Spaces or Scenes: Map-based Orientation in Urban Environments. *Spatial Cognition & Computation*, 10(2-3):135–156, 2010.

[52] R. Dial, F. Glover, D. Karney, and D. Klingman. Shortest path forest with topological ordering: An algorithm description in SDL. *Transportation Research Part B: Methodological*, 14(4):343–347, December 1980.

[53] H.K. Dieter, D. Fox, O. Etzioni, G. Borriello, and L. Arnstein. An Overview of the Assisted Cognition Project. In *AAAI-2002 Workshop on Automation as Caregiver: The Role of Intelligent Technology in Elder Care*, pages 60–65, 2002.

[54] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

[55] E. Dodsworth and A. Nicholson. Academic Uses of Google Earth and Google Maps in a Library Setting. *Information Technology and Libraries*, 31(2), 2012.

[56] M. Dorigo and L.M. Gambardella. Ant colonies for the travelling salesman problem. *Biosystems*, 43(2):73–81, July 1997.

[57] M. Duckham and L. Kulik. Simplest Paths: Automated Route Selection for Navigation. In G. Goos, J. Hartmanis, J. Leeuwen, W. Kuhn, M.F. Worboys, and S. Timpf, editors, *Spatial Information Theory. Foundations of Geographic Information Science*, volume 2825, pages 169–185. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[58] R.A. Dwyer. A faster divide-and-conquer algorithm for constructing delaunay triangulations. *Algorithmica*, 2(1-4):137–151, November 1987.

[59] H.A. Eiselt, M. Gendreau, and G. Laporte. Arc Routing Problems, Part I: The Chinese Postman Problem. *Operations Research*, 43(2):231–242, April 1995.

[60] H.A. Eiselt, M. Gendreau, and G. Laporte. Arc Routing Problems, Part II: The Rural Postman Problem. *Operations Research*, 43(3):399–414, June 1995.

[61] S. Even and G. Even. *Graph Algorithms*. Cambridge University Press, 2 edition, 2011.

[62] U.S. Department of Transportation Federal Highway Administration. National household travel survey. `http://nhts.ornl.gov/`, 2009.

[63] P. Fernndez de Crdoba, L.M. Garca Raffi, and J.M. Sanchis. A heuristic algorithm based on Monte Carlo methods for the Rural Postman Problem. *Computers & Operations Research*, 25(12):1097–1106, December 1998.

[64] A. Fernndez-Montes, J.A. lvarez, J.A. Ortega, N.M. Madrid, and R. Seepold. An Orientation Service for Dependent People Based on an Open Service Architecture. In Andreas Holzinger, editor, *HCI and Usability for Medicine and Health Care*, volume 4799, pages 155–164. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[65] S. Fickas, M. Sohlberg, and P.-F. Hung. Route-following assistance for travelers with cognitive impairments: A comparison of four prompt modes. *International Journal of Human-Computer Studies*, 66(12):876–888, 2008.

[66] J. Fink and A. Kobsa. User modeling for personalized city tours. *Artificial Intelligence Review*, 18(1):33–74, 2002. 10.1023/A:1016383418977.

[67] A.J. Flanagin and M.J. Metzger. The credibility of volunteered geographic information. *GeoJournal*, 72(3-4):137–148, 2008.

[68] R.W. Floyd. Algorithm 97: Shortest path. *Communications of the ACM*, 5(6):345, June 1962.

[69] K.D. Forbus, J.V. Mahoney, and K. Dill. How qualitative spatial reasoning can improve strategy game AIs. *IEEE Intelligent Systems*, 17(4):25–30, July 2002.

[70] S. Fortune. A sweepline algorithm for Voronoi diagrams. *Algorithmica*, 2(1-4):153–174, November 1987.

[71] S. Fortune. Voronoi diagrams and Delaunay triangulations. *Computing in Euclidean geometry*, 1:193–233, 1992.

[72] G.N. Frederickson. Approximation Algorithms for Some Postman Problems. *Journal of the ACM*, 26(3):538–554, July 1979.

[73] C. Freksa. Spatial Aspects of Task-Specific Wayfinding Maps. In *Proceedings of visual and spatial reasoning in design.* University of Sydney: Key Centre of Design Computing and Cognition., 1999.

[74] L. Fu, D. Sun, and L.R. Rilett. Heuristic shortest path algorithms for transportation applications: State of the art. *Computers & Operations Research*, 33(11):3324–3343, November 2006.

[75] Garmin. Garmin | nvi 3790t, 2011.

[76] G.D. Garrod and K.G. Willis. The non-use benefits of enhancing forest biodiversity: A contingent ranking study. *Ecological Economics*, 21(1):45–61, 1997.

[77] G. Ghiani and G. Laporte. A branch-and-cut algorithm for the Undirected Rural Postman Problem. *Mathematical Programming*, 87(3):467–481, May 2000.

[78] B. Giles-Corti. Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. *Preventive Medicine*, 35(6):601–611, 2002.

[79] F. Glover. Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, 13(5):533–549, January 1986.

[80] F. Glover, R. Glover, and D. Klingman. Computational study of an improved shortest path algorithm. *Networks*, 14(1):25–36, 1984.

[81] A.V. Goldberg and C. Harrelson. Computing the Shortest Path: A Search Meets Graph Theory. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 156–165, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

[82] B. Golden, L. Levy, and R. Dahl. Two generalizations of the traveling salesman problem. *Omega*, 9(4):439–441, January 1981.

[83] Reginald Golledge. Path selection and route preference in human navigation: A progress report. In A. Frank and W. Kuhn, editors, *Spatial Information Theory A Theoretical Basis for GIS*, volume 988 of *Lecture Notes in Computer Science*, pages 207–222. Springer Berlin / Heidelberg, 1995. 10.1007/3-540-60392-1_14.

[84] Google. Walk this way - official google mobile blog. `http://googlemobile.blogspot.com/2010/09/walk-this-way.html`, 2010.

[85] Google. Google maps. `https://www.google.co.uk/maps/@53.805906,-1.550087,2190m/data=!3m1!1e3`, 2014.

[86] Google. Google Maps/Google Earth APIs Terms of Service | Google Maps APIs, 2016.

[87] S. Gopal, R.L. Klatzky, and T.R. Smith. Navigator: A psychologically based model of environmental learning through navigation. *Journal of Environmental Psychology*, 9(4):309 – 331, 1989.

[88] R.J. Gould. Updating the hamiltonian problem: A survey. *Journal of Graph Theory*, 15(2):121–157, June 1991.

[89] I. Gowda, D. Kirkpatrick, D. Lee, and A. Naamad. Dynamic Voronoi diagrams. *IEEE Transactions on Information Theory*, 29(5):724–731, September 1983.

[90] G.W. Groves and J.H. Van Vuuren. Efficient heuristics for the Rural Postman Problem. *ORiON*, 21(1), June 2005.

[91] K.Z. Haigh and J.R. Shewchuk. Geometric similarity metrics for case-based reasoning. In *Proc. of AAAI-94, Workshop on Case-Based Reasoning*, page 182187, 1994.

[92] N. Hanley, S. Mourato, and R.E. Wright. Choice Modelling Approaches: A Superior Alternative for Environmental Valuation? *Journal of Economic Surveys*, 15(3):435–462, 2002.

[93] S. Haque, L. Kulik, and A. Klippel. Algorithms for reliable navigation and wayfinding. In *Spatial Cognition V Reasoning, Action, Interaction*, volume 4387, pages 308–326. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[94] P. Hart, N. Nilsson, and B. Raphael. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

[95] A. Helal, S.E. Moore, and B. Ramachandran. Drishti: an integrated navigation system for visually impaired and disabled. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, pages 149–156, Zurich , Switzerland, 2001. IEEE Comput. Soc.

[96] A. Hertz, G. Laporte, and P.N. Hugo. Improvement Procedures for the Undirected Rural Postman Problem. *INFORMS Journal on Computing*, 11(1):53–62, February 1999.

[97] Alain Hertz. Recent Trends in Arc Routing. In M.C. Golumbic and I.B.-A. Hartman, editors, *Graph Theory, Combinatorics and Algorithms*, volume 34, pages 215–236. Springer-Verlag, New York, 2005.

[98] C. Heye and S. Timpf. Factors influencing the physical complexity of routes in public transportation networks. In *Electronic proceedings of the 10th international conference on travel behaviour research*, 2003.

[99] B. Hillier and A. Penn. Dense civilisations: the shape of cities in the 21st century. *Applied Energy*, 43(1-3):41–66, 1992.

[100] B. Hillier, A. Penn, J. Hanson, T. Grajewski, and J. Xu. Natural movement: or, configuration and attraction in urban pedestrian movement. *Environ. Plann. B*, 20(1):29–66, 1993.

[101] H. Hochmair. Decision support for bicycle route planning in urban environments. In *Proceedings of the 7th AGILE Conference on Geographic Information Science*, pages 697–706. Crete University Press, Heraklion, Greece, 2004.

[102] H.H. Hochmair and V. Karlsson. Investigation of preference between the least-angle strategy and the initial segment strategy for route selection in unknown environments. In D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan,

D. Terzopoulos, D. Tygar, M.Y. Vardi, G Weikum, C Freksa, M. Knauff, B. Krieg-Brckner, B. Nebel, and T. Barkowsky, editors, *Spatial Cognition IV. Reasoning, Action, Interaction*, volume 3343, pages 79–97. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[103] H. Hoffman. Virtual Reality: A New Tool for Interdisciplinary Psychology Research. *CyberPsychology & Behavior*, 1(2):195–200, 1998.

[104] J.A. Hoogeveen. Analysis of Christofides' heuristic: Some paths are more difficult than cycles. *Operations Research Letters*, 10(5):291–295, July 1991.

[105] S.-Y. Hu, S.-C. Chang, and J.-R. Jiang. Voronoi State Management for Peer-to-Peer Massively Multiplayer Online Games. In *2008 5th IEEE Consumer Communications and Networking Conference*, pages 1134–1138. IEEE, 2008.

[106] N. Humpel, N. Owen, D. Iverson, E. Leslie, and A. Bauman. Perceived environment attributes, residential location, and walking for particular purposes. *American Journal of Preventive Medicine*, 26(2):119–125, 2004.

[107] J.D. Hunt and J.E. Abraham. Influences on bicycle use. *Transportation*, 34(4):453–470, 2007.

[108] T. Ishikawa and K.A. Kastens. Why some students have trouble with maps and other spatial representations. *Journal of Geoscience Education*, 53(2):184–197, 2005.

[109] R. Jacob, M. Marathe, and K. Nagel. A computational study of routing algorithms for realistic transportation networks. *Journal of Experimental Algorithmics*, 4:6–es, December 1999.

[110] P. Jansen-Osmann. Using desktop virtual environments to investigate the role of landmarks. *Computers in Human Behavior*, 18(4):427 – 436, 2002.

[111] X. Jin and C.H. Davis. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Advances in Signal Processing*, 2005(14):2196–2206, 2005.

[112] D.B. Johnson. Efficient Algorithms for Shortest Paths in Sparse Networks. *Journal of the ACM*, 24(1):1–13, 1977.

[113] D.S. Johnson. Local optimization and the Traveling Salesman Problem. In M.S. Paterson, editor, *Automata, Languages and Programming*, volume 443, pages 446–461. Springer-Verlag, Berlin/Heidelberg, 1990.

[114] D.S. Johnson and L.A. McGeoch. The traveling salesman problem: A case study in local optimization. *Local search in combinatorial optimization*, 1:215–310, 1997.

[115] D.S. Johnson and L.A. McGeoch. Experimental Analysis of Heuristics for the STSP. In D.-Z. Du, P.M. Pardalos, G. Gutin, and A.P. Punnen, editors, *The Traveling Salesman Problem and Its Variations*, volume 12, pages 369–443. Springer US, Boston, MA, 2007.

[116] R. Kaplan, S. Kaplan, and T. Brown. Environmental preference: A comparison of four domains of predictors. *Environment and Behavior*, 21(5):509–530, 1989.

[117] D. Karger, R. Motwani, and G.D.S. Ramkumar. On approximating the longest path in a graph. In G. Goos, J. Hartmanis, Frank Dehne, Jrg-Rdiger Sack, Nicola Santoro, and Sue Whitesides, editors, *Algorithms and Data Structures*, volume 709, pages 421–432. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.

[118] D. Karger, R. Motwani, and G.D.S. Ramkumar. On approximating the longest path in a graph. *Algorithmica*, 18(1):82–98, May 1997.

[119] C.E. Kelly, M.R. Tight, F.C. Hodgson, and M.W. Page. A comparison of three methods for assessing the walkability of the pedestrian environment. *Journal of Transport Geography*, 19(6):1500 – 1508, 2011. Special section on Alternative Travel futures.

[120] R.L. Kent. Attributes, features and reasons for enjoyment of scenic routes: a comparison of experts, residents, and citizens. *Landscape Research*, 18(2):92–102, 1993.

[121] R.L. Kent. Determining scenic quality along highways: a cognitive approach. *Landscape and Urban Planning*, 27(1):29–45, 1993.

[122] A. Klippel, K.-F. Richter, T. Barkowsky, and C. Freksa. The Cognitive Reality of Schematic Maps. In L. Meng, T. Reichenbacher, and A. Zipf, editors, *Map-based Mobile Services*, pages 55–71. Springer-Verlag, Berlin/Heidelberg, 2005.

[123] G.A. Klunder and H.N. Post. The shortest path problem on large-scale real-road networks. *Networks*, 48(4):182–194, December 2006.

[124] M. Konecny. Review: Cartography: challenges and potential in the virtual geographic environments era. *Annals of GIS*, 17(3):135–146, 2011.

[125] C. Kray and C. Elting. Presenting route instructions on mobile devices. In *Proceedings of IUI 03*, page 117124, Miami, Florida, USA, 2003. ACM Press.

[126] A. Krger, A. Butz, C. Mller, C. Stahl, R. Wasinger, K.-E. Steinberg, and A. Dirschl. The connected user interface: realizing a personal situated navigation service. In *Proceedings of the 9th international conference on Intelligent user interfaces*, pages 161–168. ACM Press, 2004.

[127] B. Kuipers. Modeling spatial knowledge. *Cognitive Science*, 2(2):129–153, 1978.

[128] B. Kuipers, D.G. Tecuci, and B.J. Stankiewicz. The skeleton in the cognitive map: A computational and empirical exploration. *Environment & Behavior*, 35(1):81–106, 2003.

[129] R. Laing, D. Miller, A.-M. Davies, and S. Scott. Urban green space: the incorporation of environmental values in a decision support system. *ITCon*, 11:177–196, 2006.

[130] E. Lange. The limits of realism: perceptions of virtual landscapes. *Landscape and Urban Planning*, 54(1-4):163–182, 2001.

[131] G. Laporte. The traveling salesman problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, 59(2):231–247, June 1992.

[132] G. Laporte, J. Riera-Ledesma, and J.-J. Salazar-Gonzlez. A Branch-and-Cut Algorithm for the Undirected Traveling Purchaser Problem. *Operations Research*, 51(6):940–951, December 2003.

[133] J.J. LaViola, Jr. A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32(1):47–56, 2000.

[134] C. Lee and A. Vernez Moudon. Correlates of walking for transportation or recreation purposes. *Journal of Physical Activity and Health*, 3(1), 2006.

[135] J.K. Lenstra and A.R. Kan. Some Simple Applications of the Travelling Salesman Problem. *Operational Research Quarterly (1970-1977)*, 26(4):717, November 1975.

[136] A.N. Letchford. *Polyhedral Results for Some Constrained Arc-Routing Problems*. PhD thesis, Department of Management Science, Lancaster University, England, 1996.

[137] S. Lin and B.W. Kernighan. An Effective Heuristic Algorithm for the Traveling-Salesman Problem. *Operations Research*, 21(2):498–516, April 1973.

[138] J.T. Linderoth and M.W.P. Savelsbergh. A Computational Study of Search Strategies for Mixed Integer Programming. *INFORMS Journal on Computing*, 11(2):173–187, May 1999.

[139] A.L. Liu, H. Hile, G. Borriello, P.A. Brown, M. Harniss, H. Kautz, and K. Johnson. Customizing Directions in an Automated Wayfinding System for Individuals with Cognitive Impairment. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '09, pages 27–34, New York, NY, USA, 2009. ACM.

[140] J. Loh, T. Schietecat, T.F. Kwok, L. Lindeboom, and P. Joore. Technology applied to address difficulties of Alzheimer patients and their partners. In *Proceedings of the conference on Dutch directions in HCI*, page 18, Amsterdam, Holland, 2004. ACM Press.

[141] J.J. Louviere. Conjoint analysis modelling of stated preferences: A review of theory, methods, recent developments and external validity. *Journal of Transport Economics and Policy*, 22(1):pp. 93–119, 1988.

[142] K. Lynch. *The Image of the City (Harvard-MIT Joint Center for Urban Studies Series)*. The MIT Press, 1960. Published: Paperback.

[143] J.N. Macgregor and T. Ormerod. Human performance on the traveling salesman problem. *Perception & Psychophysics*, 58(4):527–539, 1996.

[144] D. Maddison and S. Mourato. Valuing different road options for Stonehenge. *Conservation and Management of Archaeological Sites*, 4(4):203–212, 2001.

[145] R. Malaka and E. Zipf. Deep map - challenging IT research in the framework of a tourist information system. In *Information and Communication Technologies in Tourism 2000*, page 1527. Springer, 2000.

[146] R. Matai, S. Singh, and M. Lal. Traveling Salesman Problem: an Overview of Applications, Formulations, and Solution Approaches. In Donald Davendra, editor, *Traveling Salesman Problem, Theory and Applications*. InTech, November 2010.

[147] L. McGinty and B. Smyth. Personalised route planning: A case-based approach. In Enrico Blanzieri and Luigi Portinale, editors, *Advances in Case-Based Reasoning*, volume 1898, pages 431–443. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[148] C. McNamara and M. Neale. What to look for to find the best handheld gps to meet your needs and interests? Learn lessons from our review of handheld gpss such as: which features really matter, and will spending more give you any practical benefit. `http://www.outdoorgearlab.com/Handheld-Gps-Reviews/Buying-Advice`, 2013.

[149] V. Mesev. Identification and characterisation of urban building patterns using IKONOS imagery and point-based postal data. *Computers, Environment and Urban Systems*, 29(5):541–557, 2005.

[150] P.-E. Michon and M. Denis. When and why are visual landmarks used in giving directions? In Daniel R. Montello, editor, *Spatial Information Theory*, volume 2205, pages 292–305. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[151] C. Millington, C.W. Thompson, D. Rowe, P. Aspinall, C. Fitzsimons, N. Nelson, and N. Mutrie. Development of the scottish walkability assessment tool (SWAT). *Health & Place*, 15(2):474 – 481, 2009.

[152] A. Millonig and K. Schechtner. Developing landmark-based pedestrian-navigation systems. *IEEE Transactions on Intelligent Transportation Systems*, 8:43–49, 2007.

[153] H. Millward, J. Spinney, and D. Scott. Active-transport walking behavior: destinations, durations, distances. *Journal of Transport Geography*, 28:101–110, April 2013.

[154] MiTAC International Corporation. MiTAC International Corp., to provide the cloud computing solutions,GPS navigation service and mobile solutions. `http://www.mitac.com/`, 2015.

[155] L. Naismith, M. Sharples, and J. Ting. Evaluation of CAERUS: a context aware mobile guide. In *Proceedings of mLearn 2005 - Mobile*, page 112115, 2005.

[156] J.L. Nasar. Urban design aesthetics: The evaluative qualities of building exteriors. *Environment and Behavior*, 26(3):377–401, 1994.

[157] F. Neuhaus. UrbanDiary-A Tracking Project Capturing the beat and rhythm of the city: Using GPS devices to visualise individual and collective routines within Central London. *The Journal of Space Syntax*, 1(2):336, 2010.

[158] A.S. Niaraki and K. Kim. Ontology based personalized route planning system using a multi-criteria decision making approach. *Expert Systems with Applications*, 36(2):2250–2259, 2009.

[159] Cynthia Nickerson, Robert Ebel, Allison Borchers, and Fernando Carriazo. *Major uses of land in the United States, 2007*. US Department of Agriculture, Economic Research Service, 2011.

[160] C. Nilsson. Heuristics for the traveling salesman problem. Technical report, Tech. Report, Linkping University, Sweden, 2003.

[161] K. Oh. A perceptual evaluation of computer-based landscape simulations. *Landscape and Urban Planning*, 28(2-3):201–216, 1994.

[162] T. Ohya, M. Iri, and K. Murota. Improvements of the incremental method for the Voronoi diagram with computational comparison of various algorithms. *J. OPER. RES. SOC. JAPAN.*, 27(4):306–336, 1984.

[163] M.J. O'Neill. Effects of Signage and Floor Plan Configuration on Wayfinding Accuracy. *Environment and Behavior*, 23(5):553–574, 1991.

[164] M.J. O'Neill. Evaluation of a Conceptual Model of Architectural Legibility. *Environment and Behavior*, 23(3):259–284, 1991.

[165] H.L. Ong. Approximate algorithms for the travelling purchaser problem. *Operations Research Letters*, 1(5):201–205, November 1982.

[166] OpenStreetMap Wiki contributors. Points of interest - OpenStreetMap wiki. `http://wiki.openstreetmap.org/wiki/Points_of_interest`, 2013.

[167] Ordnance Survey. Business and government | Ordnance Survey. `https://www.ordnancesurvey.co.uk/business-and-government/`, 2015.

[168] Ordnance Survey. OS OpenData products | Business and government | Ordnance Survey. `https://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products-grid.html`, 2015.

[169] Ordnance Survey. Points of Interest hierarchy of classes, categories and groups. `https://www.ordnancesurvey.co.uk/business-and-government/help-and-support/products/points-of-interest-hierarchy.html`, 2015.

[170] A.J. Orman and H.P. Williams. A Survey of Different Integer Programming Formulations of the Travelling Salesman Problem. In E.J. Kontoghiorghes and C. Gatu, editors, *Optimisation, Econometric and Financial Analysis*, volume 9, pages 91–104. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[171] N. Owen, N. Humpel, E. Leslie, A. Bauman, and J.F. Sallis. Understanding environmental influences on walking. *American Journal of Preventive Medicine*, 27:67–76, 2004.

[172] M. Padberg and G. Rinaldi. A Branch-and-Cut Algorithm for the Resolution of Large-Scale Symmetric Traveling Salesman Problems. *SIAM Review*, 33(1):60–100, March 1991.

[173] S. Pallottino. Shortest-path methods: Complexity, interrelations and new propositions. *Networks*, 14(2):257–267, 1984.

[174] C.H. Papadimitriou. The Complexity of the LinKernighan Heuristic for the Traveling Salesman Problem. *SIAM Journal on Computing*, 21(3):450–465, June 1992.

[175] U. Pape. Implementation and efficiency of Moore-algorithms for the shortest route problem. *Mathematical Programming*, 7(1):212–222, December 1974.

[176] K. Park, M. Bell, I. Kaparias, and K. Bogenberger. Learning user preferences of route choice behaviour for adaptive route guidance. *IET Intelligent Transport Systems*, 1:159, 2007.

[177] D.W. Pearce, G. Atkinson, and S. Mourato. *Cost-benefit analysis and the environment: recent developments*. Organisation for Economic Co-operation and Development, Paris, 2006.

[178] W.L. Pearn and R.C. Chien. Improved solutions for the traveling purchaser problem. *Computers & Operations Research*, 25(11):879–885, November 1998.

[179] H. Petrie, V. Johnson, T. Strothotte, A. Raab, S. Fritz, and R. Michel. MOBIC: Designing a Travel Aid for Blind and Elderly People. *Journal of Navigation*, 49:45, 2009.

[180] T. Pikora, B. Giles-Corti, F. Bull, K. Jamrozik, and R. Donovan. Developing a framework for assessment of the environmental determinants of walking and cycling. *Social Science & Medicine*, 56(8):1693–1703, 2003.

[181] C. Prato and S. Bekhor. Applying branch-and-bound technique to route choice set generation. *Transportation Research Record*, 1985(1):19–28, 2006.

[182] P. Pruch and E.A. Lapin. Route knowledge in different spatial frames of reference. *Acta Psychologica*, 84(3):253–269, 1993.

[183] T. Ramesh. Traveling purchaser problem. *Opsearch*, 18(1-3):78–91, 1981.

[184] L. Ran, S. Helal, and S. Moore. Drishti: An Integrated Indoor/Outdoor Blind Navigation System and Service. *Pervasive Computing and Communications, IEEE International Conference on*, page 23, 2004.

[185] C. Rego, D. Gamboa, F. Glover, and C. Osterman. Traveling salesman problem heuristics: Leading methods, implementations and latest advances. *European Journal of Operational Research*, 211(3):427–441, June 2011.

[186] E. Rippel, A. Bar-Gill, and N. Shimkin. Fast Graph-Search Algorithms for General-Aviation Flight Trajectory Generation. *Journal of Guidance, Control, and Dynamics*, 28(4):801–811, July 2005.

[187] S. Rogers, C.-N. Fiechter, and P. Langley. An adaptive interactive agent for route advice. In *Proceedings of the Third Annual Conference on Autonomous Agents*, pages 198–205, Seattle, Washington, United States, 1999. ACM Press.

[188] D.J. Rosenkrantz, R.E. Stearns, and P.M. Lewis, II. An Analysis of Several Heuristics for the Traveling Salesman Problem. *SIAM Journal on Computing*, 6(3):563–581, September 1977.

[189] R.A Ruddle, E. Volkova, B. Mohler, and H.H. Blthoff. The effect of landmark and body-based sensory information on route knowledge. *Mem Cognit*, 39(4):686–699, 2011.

[190] S.J. Russell and P. Norvig. *Artificial intelligence: a modern approach ; [the intelligent agent book]*. Prentice Hall series in artificial intelligence. Prentice Hall, Upper Saddle River, NJ, 2. ed., internat. edition, 2003.

[191] E.K. Sadalla and L.J. Staplin. The perception of traversed distance: Intersections. *Environment and Behavior*, 12:167–182, 1980.

[192] W. Sadowski and K. Stanney. Presence in virtual environments. In *Handbook of virtual environments: Design, implementation, and applications*, Human factors and ergonomics., pages 791–806. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2002.

[193] B.E. Saelens, J.F. Sallis, J.B. Black, and D. Chen. Neighborhood-Based Differences in Physical Activity: An Environment Scale Evaluation. *Am J Public Health*, 93(9):1552–1558, September 2003.

[194] Saturday Walkers Club. Handheld GPS devices, Mapping Software, GPS Accuracy, GPX and KML. `http://www.walkingclub.org.uk/gps-map-apps/GPS-devices-apps-accuracy-OS-mapping-GPX-KML.shtml`, 2015.

[195] A. Schilling, V. Coors, A. Schilling, M. Giersich, and R. Aasgaard. Introducing 3d GIS for the mobile community - technical aspects in the case of TellMaris. In *Proceedings of IMC 2003*, pages 86–92, Rostock, 2003.

[196] M.J. Schuemie, P. van der Straaten, M. Krijn, and C.A.P.G. van der Mast. Research on Presence in Virtual Reality: A Survey. *CyberPsychology & Behavior*, 4(2):183–201, 2001.

[197] P.N. Seneviratne and J.F. Morrall. Analysis of factors affecting the choice of route of pedestrians. *Transportation Planning and Technology*, 10(2):147 – 159, 1985.

[198] M. Ian Shamos and D. Hoey. Closest-point problems. In *16th Annual Symposium on Foundations of Computer Science (sfcs 1975)*, pages 151–162. IEEE, October 1975.

[199] A.W. Siegel and S.H. White. The development of spatial representations of large-scale environments. *Advances in Child Development and Behavior*, 10:9–55, 1975. PMID: 1101663.

[200] S. Siegel and J.N. Castellan. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, New York; London, 1988.

[201] J.A. Sinden. A Utility Approach to the Valuation of Recreational and Aesthetic Experiences. *American Journal of Agricultural Economics*, 56(1):61, 1974.

[202] K.N. Singh and D.L. van Oudheusden. A branch and bound algorithm for the traveling purchaser problem. *European Journal of Operational Research*, 97(3):571–579, March 1997.

[203] R.C. Smardon. Perception and aesthetics of the urban environment: Review of the role of vegetation. *Landscape and Urban Planning*, 15(1-2):85–106, 1988.

[204] M.M. Sohlberg, S. Fickas, P.-F. Hung, and A. Fortier. A comparison of four prompt modes for route finding for community travellers with severe cognitive impairments. *Brain Injury*, 21:531–538, 2007.

[205] W. Souffriau, P. Vansteenwegen, J. Vertommen, G. Vanden Berghe, and D. Van Oudheusden. A Personalized Tourist Trip Design Algorithm for Mobile Tourist Guides. *Applied Artificial Intelligence*, 22(10):964–985, 2008.

[206] Wouter Souffriau and Pieter Vansteenwegen. Tourist trip planning functionalities: StateoftheArt and future. In F. Daniel, F. Michele Facca, et al., editors, *Current Trends in Web Engineering*, volume 6385, pages 474–485. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[207] M. Southworth. Designing the Walkable City. *Journal of Urban Planning and Development*, 131(4):246–257, 2005.

[208] M. St. John, M.B. Cowen, H.S. Smallman, and H.M. Oonk. The Use of 2D and 3D Displays for Shape-Understanding versus Relative-Position Tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1):79–98, 2001.

[209] A.E. Stamps. Some streets of san francisco: preference effects of trees, cars, wires, and buildings. *Environment and Planning B: Planning and Design*, 24(1):81–93, 1997.

[210] L.J. Staplin and E.K. Sadalla. Distance cognition in urban environments. *The Professional Geographer*, 33:302–310, 1981.

[211] R. Stoakley, M.J. Conway, and R. Pausch. Virtual reality on a WIM. In *CHI '95 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 265–272. ACM Press, 1995.

[212] H.A. Taylor and B. Tversky. Descriptions and depictions of environments. *Memory & Cognition*, 20(5):483–496, 1992.

[213] A. Teeninga and A. Volgenant. Improved heuristics for the traveling purchaser problem. *Computers & Operations Research*, 31(1):139–150, January 2004.

[214] TFL. Tube - Transport for London. `https://tfl.gov.uk/maps/track/tube`, 2015.

[215] TomTom. GO LIVE 1005 Europe - car navigation - TomTom. `http://www.tomtom.com/en_gb/products/car-navigation/go-live-1005-europe/`, 2011.

[216] K. Train. *Discrete choice methods with simulation*. Cambridge University Press, New York, 2003.

[217] S.-K. Tsai. WADER: a novel wayfinding system with deviation recovery for individuals with cognitive impairments. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Assets '07, pages 267–268, Tempe, Arizona, USA, 2007. ACM Press.

[218] A. Turner. Could a road-centre line be an axial line in disguise? In *Proceedings of the 5rd Space Syntax Symposium*, volume 1, pages 145–159, 2005.

[219] Barbara Tversky. Cognitive maps, cognitive collages, and spatial mental models. In A.U. Frank and I. Campari, editors, *Spatial Information Theory A Theoretical Basis for GIS*, volume 716 of *Lecture Notes in Computer Science*, pages 14–24. Springer Berlin Heidelberg, 1993.

[220] Barbara Tversky. Some Ways that Maps and Diagrams Communicate. In C. Freksa, C. Habel, W. Brauer, and K.F. Wender, editors, *Spatial Cognition II*, volume 1849 of *Lecture Notes in Computer Science*, pages 72–79. Springer Berlin Heidelberg, 2000.

[221] Department for Transport UK Data Archive [distributor]. National travel survey 2010 - statistics. `http://www.dft.gov.uk/statistics/releases/national-travel-survey-2010/`, 2010.

[222] R.S. Ulrich. Human responses to vegetation and landscapes. *Landscape and Urban Planning*, 13:29–44, 1986.

[223] University of Edinburgh. Digimap, January 2017.

[224] J. van Leeuwen and A.A. Schoone. Untangling a traveling salesman tour in the plane. Technical Report RUU-CS-80-11, Department of Computer Science, University of Utrecht, 1980.

[225] P. Vansteenwegen and W. Souffriau. Trip planning functionalities: State of the art and future. *Information Technology & Tourism*, 12(4):305–315, 2011.

[226] D. Veldkamp, F. Hagethorn, B. Krse, and P. De Greef. The use of visual landmarks in a wayfinding system for elderly with beginning dementia. *MedeTel 2008*, pages 161–165, 2008.

[227] S. Vo. Dynamic tabu search strategies for the traveling purchaser problem. *Annals of Operations Research*, 63(2):253–275, April 1996.

[228] T. Völkel and G. Weber. RouteCheckr: personalized multicriteria routing for mobility impaired pedestrians. In *Assets '08 Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 185–192. ACM Press, 2008.

[229] Walkingworld. Walkingworld - GPS. `http://www.walkingworld.com/Articles/General-info/GPS.aspx`, 2015.

[230] walkit.com. walkit.com The urban walking route planner. `http://walkit.com/`, 2015.

[231] J. Weisman. Evaluating Architectural Legibility: Way-Finding in the Built Environment. *Environment and Behavior*, 13(2):189–204, 1981.

[232] J. Wood, A. Slingsby, and J. Dykes. Visualizing the Dynamics of London's Bicycle-Hire Scheme. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 46(4):239–251, 2011.

[233] Y. Yang and A.V. Diez-Roux. Walking Distance by Trip Purpose and Population Subgroups. *American Journal of Preventive Medicine*, 43(1):11–19, July 2012.

[234] J. Zacharias. Exploratory spatial behaviour in real and virtual environments. *Landscape and Urban Planning*, 78(1-2):1–13, 2006.

[235] F.B. Zhan and C.E. Noon. Shortest Path Algorithms: An Evaluation Using Real Road Networks. *Transportation Science*, 32(1):65–73, February 1998.

[236] G.K. Zipf. *Human behavior and the principle of least effort.* Addison-Wesley Press, Oxford, England, 1949.