AN EXPERIMENTAL INVESTIGATION INTO THE DESIGN AND PERCEPTION OF LINE SYMBOL SERIES ON ROUTE PLANNING MAPS

by

Graham Anthony Gill, M.A. Dip.Cart.

Thesis submitted for the degree of Doctor of Philosophy

to the

University of Glasgow

Department of Geography

October, 1986

CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

The work was funded through a Collaborative Award in the Social Sciences (CASS) by the Economic and Social Research Council (ESRC) in collaboration with the RAF Institute of Aviation Medicine at Farnborough, Hants. Special thanks are due to my principal supervisor Dr. Alastair Morrison (Glasgow) for his dedicated and thorough supervision, and to my collaborative supervisor Mr. Robert M.Taylor (Farnborough). I would also like to thank Mr. M.Spencer (Farnborough) for statistical advice and expertise, and the technical staff of Glasgow University Geography Department, especially Ian Gerrard and Mike Shand for their reprographic work in the preparation of various stimuli.

I should think that everybody who has experienced PhD research knows how invaluable are those people who provide the gentle breeze of encouragement through the all-too-frequent doldrums, and in this respect I am particularly grateful to Dr. Allan Findlay (Glasgow Geography Department), Dr. Barry Jones (Glasgow Psychology Department) and also Dr. Margaret 'supersubject' Robb, who took on the additional task of piloting my experiments, including a few which have fortunately not made it to the latter chapters of this thesis.

Thanks are also due to all the people who acted as subjects in any of the experiments, with minimal incentives. With respect to experiment 3 (chapter 9), I would also like to acknowledge the collaboration of the Automobile Association: the interest and practical help provided by Ralph Robbins and Vic Bates, and the assistance in arranging the interview locations provided by Mr. Peter Mouncey, Mr. E.Howard (Manchester) and Mrs.

M.Cawood (Leeds). I am also most grateful to Mr. Jim Barr, manager of Smiley's Partick depot, for allowing me to interview on his premises, and to David Miller, for conducting the 'Aberdeen' interviews so enthusiastically and conscientiously.

Finally, I would like to thank my parents, who, whenever you read this, are probably still wondering when I am going to get a 'proper job'.

NOTE ON LINGUISTIC SEXISM

Throughout this thesis, 'he' is used to refer to a member of a group of which the majority are males, and vice versa.

ABSTRACT

Recent estimates by the Department of Transport have suggested that up to £800 million of national resources are wasted in the U.K. each year through unnecessarily poor route choices. A valuable contribution towards the recovery of these losses could be made by the provision of more informative road maps, but the communication of this information will only be improved if it is adequately represented.

An investigation was consequently undertaken to discover ways in which road map information can be displayed in order to maximise the probability of efficient route choice. A pilot study confirmed that differences in symbolisation between two maps with the same road classification could significantly affect the routes chosen from them. A linked series of experiments was conducted to establish principles to enable the map designer to select a set of road symbols in which the potential conflict between order and distinctiveness might be resolved, so that the classification hierarchy might be unambiguously perceived within the constraints imposed by map reproduction processes. The prominence and conspicuity of line symbols of varying colour, width and casing was assessed by psychophysical magnitude estimation and visual search tasks. The prominence of uncased lines was found to be definable by a formula involving the logarithm of line width, and the lightness contrast, saturation and affective value of the line colour. The situation is, however, complicated by the introduction of casing lines and background colour.

In tests of specially designed experimental maps on a representative sample of British motorists, it emerged that the main cues used in decoding the classification were (in declining order of importance) the Department of Transport road class, the relative continuity of roads of

each class, particular graphical conventions (such as blue=motorway) and the prominence of the line symbol. Prominence does however become increasingly important where one or more of the other cues are not present. The most effective symbol series were where these cues operated concordantly. Colour provides the most immediate perceptual segregation of line classes, and should clearly be used to represent major discontinuities in road quality, such as between single and dual carriageways. The extent of built-up areas should also be clearly depicted. In general it is recommended that the size and evenness of the steps in the graphical hierarchy used should reflect the relative differences in road quality represented by the classificaion scheme used.

Symbols for the most common types of restricted access junction were also devised and successfully tested.

1. INTRODUCTION

For the majority of road journeys undertaken in the U.K., there is more than one possible route that could realistically be taken. Even when motorists set off on the same journey with the same objective (e.g. taking the route requiring the minimum travel time) they will often choose very different routes. In a study by Lunn (1978), a sample of 70 volunteer motorists was asked to drive from the Transport and Road Research Laboratory (Crowthorne, Berkshire), either to an address in Chertsey, Surrey or one in Bedford, as if they had an appointment for which they had allowed the right amount of time. The results demonstrated both the diversity of routes chosen in such circumstances and the large amount of unnecessary mileage driven. On average, the distance taken 'en route' (i.e. as far as the edge of the destination town) was 12.3% and 10.5% more than the minimum possible for Chertsey and Bedford respectively, with similar excesses (15% and 10%) for journey times as well.

Various surveys of representative samples of traffic in different areas (reported by Outram,1976; Carpenter,1979 and Jeffery,1981a) have produced estimates of overall excess at between 5.5% and 6.6% of the total distance travelled. Discounting journeys

1) less than 5 Km long,
2) regularly-driven (e.g. commuting trips),
3) where 'non-optimum' (e.g. scenic) routes are sought, or
4) where there is no real alternative route,

the figure for the amount of mileage that is realistically avoidable is reduced to about 2%. However, recent estimates have put the cost to national resources of this avoidable wastage as £800 million per year (Jeffery,1985).

Thus an £800 million loss is caused entirely by the extra length and duration of trips where this is both unnecessary and unwanted by the traveller. Clearly many drivers seeking optimum routes fail in their objectives: in one of the surveys, involving inter-urban trips in Gloucestershire, the proportion of journeys where drivers actually used the route which best fulfilled their stated objective varied between 36.2 and 53.7%, depending on the type of journey involved. The added cost amounts to about £40 a year for the average car and £130 for each lorry. Moreover the majority (70%) of this wastage occurs en route rather than in the terminal search for the specific destination, where the average motorist is more conscious of potential inefficiency. Consequently most of the savings are to be made on longer distance inter-urban or cross-country trips, where road maps have an important influence. The benefits of an overall reduction in aggregate distance and travel time could accrue from a better distribution of traffic across the road network, except where it is already operating at capacity, as in many metropolitan areas during peak hours. A corresponding reduction in the number of accidents would also be expected (Russon,1986). The cut-back in new road construction in the U.K. has made the efficient use of the existing network all the more important (Jeffery,1981b).

In order to improve the efficiency of route planning, two basic methodologies are available. One is to use a route guidance system that is able to calculate the optimum route (in terms of shortest time, shortest distance or some composite measure of 'cost') and tell the traveller how to follow it. The second is to provide the road user with better quality information from which he can make his own route decisions according to his own specific circumstances. Conceptually, route guidance systems can be divided into four major categories:

1) instructions supplied by information displays beside or above the carriageway (e.g. Keller and Cremer,1982),

2) routeings obtained before travelling either from

motoring organisations or from dial-up Viewdata systems such as ROUTE-TEL or CARTRIPS (Robb,1985) in the U.K., which supply route listings and strip maps respectively,

3) self-contained in-vehicle navigation systems using dead reckoning techniques (measurements of distance and direction travelled) to keep track of vehicle position: some of these systems do not plan routes (e.g. the Honda Gyrocator in Japan and the VDO Citypilot in Germany) while others use network knowledge to calculate optimum routes and prevent cumulative positional errors (e.g. the Etak Navigator in the U.S.A. and the Phillips CARIN system in Europe),

4) in-vehicle systems involving communication with outside units for position fixing and potentially real-time routeing in response to changing road conditions, the units being either roadside beacons, loops buried beneath the road, radio transmitters or satellites (e.g. Plessey 'Autoguide', Bosch/Blaupunkt ALI and several systems being developed by American car manufacturers). Some of the in-vehicle systems include map output on VDUs, while others merely use textual or spoken instructions. Obviously at the current state of development, all these systems are highly capital intensive and involve very high costs either to the consumer, to public funds or to both (Jeffery,1981b).

Information enabling travellers to make their own route choices (both before and during the journey) is derived from five main sources:

1) people's existing knowledge
2) personal advice
3) broadcast traffic and weather information
4) signposts
5) maps.

While the first two of these cannot be directly improved, the Department of Transport have attempted to evaluate the financial benefits that might derive from improved maps and signposts, compared with various types of automated route guidance system (Jeffery,1981b). Currently, signposts appear to have the most influence on route choices made in transit (Wootton et al.,1981), while maps are by far the most important source of information for pre-journey planning of long routes over unfamiliar roads (A.Morrison,1979a). In the area of the aforementioned Gloucestershire survey, Wootton found that at no signed junction were the signposts identical to a computed optimum, and that on average a poor direction was indicated at every sixth intersection. Relating this to the survey travel data, an optimised system could achieve savings of up to £100 million per annum (at 1979 prices) with a much higher cost-benefit ratio than any of the route guidance systems. However, Jeffery estimated that despite yielding the lowest aggregate return (£20 million p.a.), improved maps would provide the highest cost-benefit ratio of all, because their development and production would be so cheap and involve minimal use of public funds, even where new information had to be collected. The most important improvement required on maps is information to enable better assessments of relative journey times (Jeffery,1981b), given the wide range of likely travel speeds prevalent on the current road network.

The objection might be made that in presenting such information on a stimulus which can be updated only infrequently (i.e. the road map), routes chosen from an intelligent reading of the map may in fact be slower, once a certain level of usage of that map is attained. For example, in a choice between a slow, congested trunk route and a faster cross-country alternative, the map might encourage sufficient traffic to divert onto the latter to make it eventually slower than the trunk route. However, until that use level is reached, benefits are obtained both by the drivers who divert and by those who remain on the trunk road with its somewhat reduced traffic load. Beyond that point, a virtual map display with a dynamic response to changing traffic conditions would be required. According to Janssen (1975,p.19, transl.), 'traffic flow will always be benefitted by supplying the person involved in the traffic with complete information about the alternatives at his disposal.'

To improve the cartographic communication of information, better map design and better user training should ideally go hand in hand (M.Wood,1972; Eastman and Castner,1983). Given the potential problems of trying to train an almost entirely adult target population of drivers and navigators, improving the maps is clearly the line of less resistance with the more immediate benefits. The problem then for the map designer is defining the 'normative values for the user population' so as to maximise the possibility of the message being understood by those to whom it is directed (Easterby,1984,p.24).

1.1 Maps and Route Planning

The unique property of maps as a source of route planning information is their ability to show, using the inherent visual organisation provided by two-dimensional space (Haber,1981), the totality and complexity of spatial relations between the roads and places in the area of interest, so that routes can be selected from an informed synoptic overview of the possibilities. Bartram (1980) has demonstrated the superior performance in the planning of multiple-bus journeys of people using maps rather than verbal route lists, suggesting that the map is the format most compatible with people's mental representation of spatial information. While route listings give no information about the surroundings of a route, maps enable the skilled user to make diversions or detours, and possibly to recover his route if he loses his way. From a commercial point of view, maps have a further advantage over route guidance systems with no graphical output-namely that the producer is not responsible for the quality of the route choice. Until highly sophisticated systems involving full real-time monitoring of traffic conditions are available, any system will not be able to guarantee the quality of the decision it makes in every single case. This is a substantial disincentive to companies considering the development and/or marketing of such systems.

However, 'the inadequacy of many road maps forcibly strikes every traveller sooner or later.' (A.Morrison,1966,p.17) Often this is because the information they contain is either irrelevant, inaccurate, out-of-date or poorly presented, or because more useful information is not shown. Armstrong (1977) considers that the use of out-of-date, inaccurate or insufficiently detailed maps is one of the major causes of resource wastage through sub-optimal route planning. Developing technology has opened up some new ways in which improved map information can be provided. For example, individualised maps adjusted to the user's specific requirements could be rapidly plotted from a centrally updated database (e.g.Robb,1985), and the above-mentioned in-vehicle map displays can plot the vehicle's current position. However, while several studies have investigated road map users' information requirements and the (in)adequacy of maps to meet these (see, for example, section 2.2 below) the manner of representation of the information has been relatively ignored.

This imbalance of attention was first noted by Robinson at the start of 'The Look of Maps' in 1952 (p.4): 'The ability to gather and reproduce data', he said, 'has far outstripped our ability to present it.' Petchenik (1983,p.45) has suggested that 'in the past most data mapped were sufficiently simple and the range of graphic techniques employed so conventional that methods of depiction were essentially transparent to their users.'

With the growth in the amount and complexity of mapped data, more demanding uses of graphic variables have been required. Generally the making of a route choice from a map involves the weighing up of several factors, such as the trade-off between shorter distances and better roads. Whatever relevant information is shown, improved route choices will only result if the manner of symbolisation enables it to be clearly and unambiguously communicated to the map user. Various studies have shown that the design of a road map can have a considerable effect upon the routes chosen from it (section 2.4.3). As Dobson (1983b) has stressed, where a map display is to be used for spatial decision making, it is extremely important that it has been designed to induce quick and accurate task performance. This is perhaps especially significant with in-vehicle map use for route planning, where decisions may have to be taken in a hurry. Well-designed maps clearly do not compel their users to make good decisions with them, but they can maximise the probability of efficient route choice occurring.

Information on road classification is clearly fundamental to the route planning process. Whatever criteria are used for the classification, all schemes carry connotations of hierarchy. Consequently the main problem for the cartographer is how to assign coloured line symbols to road classes so that they form a clearly intelligible series, from which the ordering of the classes can be assessed at a glance. According to Robertson and O'Callaghan (1986,p.24-25), while some studies 'stress the value of supporting aids such as legends...to achieve...levels of comprehension, it is of primary importance that quantitative variables be represented by graphic variables which are in some way intuitively or perceptually significant.' Lloyd and Yehl (1979,p.150) state that 'providing the map reader with an orderly map with the order appropriate to the information and visually apparent to the map reader, should increase the probability of a successful map.' Clearly the map will communicate a hierarchical message most successfully if the psychological ordering of the symbols corresponds with the relative importance of the elements of information they represent.

The problem remains as to how this graphical hierarchy is achieved. Where one class of line symbols is perceived to be more important than another, what features are responsible for this, and how much control does the cartographer have over them? One particularly poorly understood area is that of colour. According to Haber (1981,p.13) 'colour alone is a powerful visual feature to provide immediate segregation of areas of details, such as to differentiate countries on a map...More complex but as yet unexplored uses of colour are possible...Is it possible to assign colour to density in such a way that the viewer can map the order of the colours onto the...values, without having to learn them?' The demands made of colour are particularly acute when different classification criteria demand the removal of some hierarchical cues. 'On conventional road maps...the pattern of the roads themselves tells the reader which is the highest category, for the roads of a lower class are always tributary to those of a higher class. It is therefore possible to publish successfully maps in which the colours representing the road classes do not lie in the order of their psychological impact.' (A.Morrison,1971,p.6) However, when the colours represent the likely travel speed on each road section, 'the fast, medium and slow sections may be isolated; they do not necessarily form a pattern', and the visual impact of the line symbols becomes all the more critical.

Consequently the aim of this thesis is to investigate the perception of line symbol series on road maps, and to discover design principles which might enable map users, in the amount of time available to them, to choose routes which fulfil their own objectives significantly better than they do at present. As there are no specific precedents to this particular study goal, it will be necessary to synthesise ideas and techniques from a wide variety of disciplines.

2. THE DESIGN OF ROUTE PLANNING MAPS

The road map is almost invariably a highly commercial product. In many countries, road maps and street plans are the only cartographic products on general sale. For many publishers they are the major or sole line; for several national survey organisations they are revenue-raising bulk sellers which subsidise less profitable products. Many are given away free as advertising or for the promotion of tourism. In any case, the ultimate aim of the product is to make money. Not only is this reflected in the content and presentation of the maps, but also in the relative lack of consideration given to them in the academic literature. The image of the road maps as a cheap, basic, easily worn-out and disposable product, encouraged in North America in particular by a history of free distribution by oil companies and now state authorities, must be partially responsible for this. According to Falk (reported by Schiede,1968) road maps do not enjoy the academic respect their wide circulation deserves, and the main design questions are too often left to oil companies and their agents, who generally possess little cartographic expertise. Additionally, according to Brian Coe, formerly of the NERC Experimental Cartography Unit (reported by Braidwood,1981,p.53), cartographers themselves are 'less interested in the principles of information design - getting people's attention and conveying ideas simply - than in measurement and accuracy.'

The map producers too have generally not been overanxious to reveal details of the process of design and evaluation of their own products. In many countries the road map market is highly competitive. For example, the U.K. map and guide market, worth £28 million in 1980, is served by at least 14 different companies who are each responsible for the preparation of at least one road map base. In such circumstances, with the majority of the companies being very small, few have undertaken much formal market research or product evaluation, and when they have, it has often remained strictly confidential. Other researchers have found it difficult to extract information from British commercial map producers. In one case where a constructive reply was received, from the leading U.S. map publisher Rand McNally, it revealed that their map specifications were often drawn up, with little forethought to consumer requirements, through committee decisions based on the skills and experience of their staff, modified through trial and error (R.D.Wright,1967).

However, a certain amount of literature on road map design does exist, and it can be divided into five main categories:

1) reviews or surveys of the informational requirements/ preferences of road map users (e.g. A.Morrison,1966; Connal,1983),

2) descriptions of the selection and representation of map contents for specific maps, sometimes by the maps' producers themselves (e.g. Mair,1963; Goldring,1978),

3) studies of users' preferences for particular designs of published or experimental map (e.g. Consumers' Association,1983b; Sheppard and Adams,1971),

4) studies of users' task performance on different designs of published or experimental map (e.g. Sheppard and Adams,1971; A.Morrison,1974),

5) traces of the history and development of road maps (Ristow,1964; Nicholson,1983).

2.1 The Route Planning Map

Road maps are put to a wide variety of uses. Because of their profusion, they are often the nearest or only map to hand in many situations where one is required. Whilst it is basically true that 'in comparison with some types of map, it is easy to evaluate a motoring map, because it has only two or three well-defined purposes and is intended for a definite population of users,' (A.Morrison,1975,p.120) these maps may well also fulfil other largely unintended functions. The following uses of road maps are suggested in the literature (Kirby,1970; Anderson,1975; A.Morrison,1975; Hilliard,1977; Thake,1979; Gill,1982; Connal,1983):

-route planning: the map helps the user to discover the best route to his destination, and how good it is;

-navigation, or route following: the map provides details which promote awareness of location, enabling the user to keep to his chosen route without getting lost;

-location of places or areas of recreational/ touristic significance for trip destinations and drives, or simply as points of interest,

-location of relevant services (e.g. fuel, food),

-general reference, especially when stored in map collections or used as wall maps,

-as gazeteers (road atlases in particular),

-providing information on large centres of population,

-calculation of distances (independent of any of the aforementioned uses), and

-walking (not recommended!).

Given this diversity it is hardly surprising that, in Keates' (1973) definition, road maps are not strictly 'thematic', but are 'special-purpose' versions of topographic maps with road information emphasised. This relates too to the origins of road map designs. Many of the 'commercial' producers' maps started as simple generalisations from topographic maps with little added information. Others are 'general purpose' maps which have gradually become more specialised: a prime example is provided by the slow but accelerating evolution of the Ordnance Survey 1:250,000 and 1:633,600 maps to meet the needs of the road user (Griffith and Kelly,1965; A.Morrison,1979b,1980b). This specialisation should be apparent in information priorities on the map, both in terms of what is selected and omitted and in its classification and representation. For example, Lundqvist (1963) suggests that features such as junctions with limited access, staggers of junctions and the by-passing of settlements by roads should be emphasised on road maps to assist route planning and navigation.

Some studies have differentiated road maps by scale in terms of the uses for which they are appropriate. In the U.K., maps at scales larger than about 5 miles to the inch (1:316,800) have been labelled 'touring' maps (Consumers' Association,1963,1983b) as they generally contain sufficient detail to enable navigation on rural minor roads. The smaller scale 'route planning' or 'long distance' maps are so called because, with the inevitable loss of detail, they are less useful for navigation or as a spatial inventory, but are suitable for planning long distance journeys on major roads. It does not mean that route planning is not or cannot also be carried out on maps of larger 'touring' scales. In fact the majority of road maps and atlases now sold in this country and used to plan routes are at 'touring' scales. The concern of the present study is with the task of planning inter-urban routes from maps. Consequently all maps and all scales which are regularly used for this purpose are of relevance, and in Kirby's (1970) survey, this included a whole range from one-inch (1:63,360) maps to general reference atlases.

2.2 Relevant Information for Route Planning Maps

In 1963 the Consumers' Association reviewed 51 contemporary British road maps. Despite the wide variety available, they were forced to conclude that the maps were 'generally graphically amateur', and that little improvement had been made to the content and presentation of motoring maps 'since the Michelin ones of the 1920s.'

However, over the next decade, the very rapid increase in the number of cars on the roads and the increasing length of holidays led to a massive expansion in the demand for road maps (Drewitt,1973). Many maps were redesigned with completely new specifications; extra detail such as dual carriageways and tourist information was gradually added. However, few map producers evaluated their products in terms of anything other than sales, an index influenced as much by the quality of marketing as the design and content of the map (Fullard,1965), which can only really be tested after purchase (Anderson,1977). Some researchers began to call for a more rigorous form of road mapping adapted to the tasks for which it was used.

In Germany, Schiede (1968) called for road maps which reflect the specialised needs of the motorist in terms of content, presentation and updatedness. Information on road condition, safety, traffic level and travel speeds is, he claimed, more important to the motorist than tourist information. He called for a road classification based on the quality of the road rather than just its administrative status or arteriality, and the depiction on the map of all the factors which can play a role in route choice, such as road width, bendiness, gradients and surface. From these the user can make the best decision for his own circumstances. Preston (1980) and Connal (1983) however have both distinguished between two types of road map user:

1) one seeking the fastest and/or most direct route to a specific destination, his only interests in transit being fuel, food and places which might slow him down or require a change in his concentration level. He requires only 'basic' information from the map.

2) one prepared for a leisurely drive with time for sight-seeing, who additionally needs tourist information and details of the environment.

Connal defined 'basic' information as trip origins and destinations (settlements etc.), a road classification based upon road quality, marked road distances (or times) and the location of services.

A number of surveys have also been undertaken (e.g. Astley,1969; Kirby,1970; Drewitt,1973; Anderson,1977; Connal,1983) which have asked road map users about the information they require from the maps. The results are, however, often of limited applicability for two main reasons.

1) Users are rarely asked about the relationship of map content to map using activities. For example, a road classification based on road quality may be desirable for route planning, while an administrative and/or arteriality classification related to signposting may be more useful for navigation.

2) When users are asked about map content outwith an actual map using task, there is a tendency to demand an amount of information that is, either because of the cost of procuring it or the technical problems of representing it without cluttering the map, unrealistically high (Connal,1983). This could be expected to be a particular problem with 'closed' questions.

Thus, according to M.Wood (1968), it is hardly surprising that many map editors have resorted again to their better judgment.

Astley's (1969) survey undertaken in the cafeterias of a motorway service station is however particularly interesting as his sample of 300 was largely composed of lorry drivers and salesmen who might be expected to fit into the 'fastest-route planning' category (A.Morrison,1979b). In closed questions, a majority favoured the indication of dual carriageways, danger spots, and road sections likely to be blocked by traffic, snow or floods, and agreed that the depiction of the extent of built-up areas was inadequate. When comments were invited, the most common calls were for a code to distinguish narrow, congested roads from fast, clear ones, and a specialist heavy haulage map showing height and weight limits and tight bends. In contrast, Kirby's (1970) sample of 1553, largely members of such organisations as the Royal Scottish Geographical Society and the Ramblers' Association, was far more orientated towards leisure drivers. Many wanted more landmarks (such as church steeples, windmills and tall chimneys) to be shown as an aid to navigation, and a 'functional', rather than simply administrative, classification of roads. Michelin road maps were often mentioned as an example of superior cartography.

In Canada, Anderson's (1977) survey of 395 users of the official Alberta road map also recommended a road classification based on quality (number of carriageways and surface type), a finding confirmed by McGrath (1971). Additionally, many people wanted scenic roads to be indicated. However, for 71% of Connal's (1983) sample of 224, their most usual use of a road map was for planning the quickest route, and environmental information was of little use. In a ranking of 18 categories of information, road surface, road number and number of lanes were amongst the four most important influences on fastest route selection. Services were relatively unimportant, but overall in the surveys there was considerable demand for the marking of public conveniences, followed by 24-hour petrol stations, police stations and restaurants (Astley,1969; Kirby,1970; Drewitt,1973; Anderson,1977).

Overall the surveys seem to indicate that the Preston/Connal model relating content requirements to use type is generally accurate.

2.3 Road Classification for Fastest-Route Planning

A.Morrison (1966,p.18) gave specific consideration to the information requirements of the map user seeking an optimum route by attempting to minimise 'the adverse properties of the journey, namely journey time and cost,' and maximise 'the favourable properties, namely safety, mental ease, physical comfort, certainty of following the route, amenity and the availability of services.' The particular significance of each of these aims will of course vary between people and specific circumstances. Some of the information needed to fulfil them is clearly available from existing maps: certainty of route following, for example, is largely a function of sticking to the more continuous roads and minimising the number of turns to be made.

Journey times, however, are less easily assessed from the map. Isoline techniques have been used for the direct mapping of likely journey times (Muller,1978), but only to one particular point on the map, and with no indication of the routes taken to achieve these times. Given that road maps show the network distances between places, indications of likely travel speed could act as a substitute for the direct representation of journey times. Speeds could either be represented

1) directly, in which a classification based on likely travel speed is depicted (A.Morrison,1971), or

2) indirectly, in which information is provided about the major factors influencing speeds. One of these is the driver's preferred speed in conditions where he is constrained only by national/ state speed limits or vehicular performance. A combination of interrelated aspects of carriageway and traffic conditions may prevent him from realising this speed, and they can be listed as follows:

1) the presence of local statutory speed limits, in built-up areas etc.

2) road width/ number of lanes/ number of carriageways, affecting the ease of passing and overtaking other vehicles

3) road surface type and quality

4) the bendiness of the road and the proportion of its length lacking overtaking sight distance

5) gradients

6) the frequency of junctions where priority is not protected

7) (potential) compulsory stops such as traffic lights, level crossings and tolls

8) the relationship of traffic volume to road and junction capacity

9) the presence of particular classes of traffic such as HGVs and slow vehicles (tractors etc.)

10) temporary stops or reductions in road capacity caused by road works, breakdowns and accidents

11) the prevailing weather conditions.

Clearly several of these factors, especially the last four (8-11), vary considerably from time to time and with unforeseeable circumstances. Thus some map producers are wary of direct indications of speed as, in addition to the usual problems of obtaining updated information, they can in no way guarantee speed designations in all circumstances. On occasions where they might turn out to be unhelpful (e.g. extensive road works on a usually fast road), the user may well lose confidence in the map. The direct representation of speed categories may also reduce the flexibility of the map for uses other than 'fastest route' planning (Connal,1983).

If an indirect indication is adopted, this puts the onus back onto the map user to make his own assessment of the prevalent conditions according to his own circumstances. Clearly not all of the more permanent factors can be incorporated into elements of the road symbol at once; nor would this be desirable as the user's information processing capacity would immediately be overloaded. However, classifications with up to four dimensions have been depicted, such as the scheme recommended for state road maps by the American

Association of State Highway Officials (AASHO,1962), where road arteriality, width and surface quality are shown by variations in symbol colour, width and configuration, and administrative status is indicated by the road number marker. Fortunately, several of the above factors can be incorporated on the map outwith the road classification. Gradients, level crossings, roundabouts and tolls are often indicated by point symbols. Bends can usually be depicted fairly accurately by the alignment of the road symbol at 1:200,000 or greater (A.Morrison,1966): at smaller scales appropriate generalisation is required.

Built-up areas can also be shown by numbers of repeated point (building) symbols or by coloured and/or textured area filling. Some maps emphasise the effect of urban areas on speeds by altering the line symbols within towns. Where no area fill is used, it might be considered desirable to make urban roads prominent and distinct, to indicate that they should clearly be avoided if possible, because of the disproportionate effect they have upon journey time. However, in practice an increase in width, for example, is generally impossible because of the increased density of information to be shown in urban areas. Thus it is more normal to depict the built-up area, within which the road symbols are sometimes narrowed (e.g. Esso maps in the British Isles) and/or deprived of their coloured filling. The remaining criteria- road width/ number of carriageways and road surface quality- are often incorporated into the road symbol, although the latter is of less significance in the densely populated countries of Western Europe where the vast majority of through roads are paved.

Studies by A.Morrison (1980a) and Taylor and Jeffery (1981) have pointed to the relative significance of the factors affecting driving speeds. Morrison tested the correlations between many of these individual factors and traffic speeds, derived from his own surveys, for the main (basically M,A and B) roads in Southern Scotland and Northern England. The strongest absolute correlations were with built-up areas and 30 m.p.h. speed limits, indicating that the urban/rural split was the most fundamental influence on speeds. He also found the classification with the best overall correlation with speeds which could be derived from existing information used in map compilation, i.e.

1) rural motorway
2) rural dual major road
3) rural single major road
4) other rural road
5) partly built-up road
6) wholly built-up road.

(The best definition of 'major' varied slightly according to the generalisation and urban/rural split of the particular network portion used.) This was clearly superior to any of the 18 classification schemes used on existing maps which he also tested.

Taylor and Jeffery, using a network around Reading, Berkshire which included more (2-lane) 'unclassified' roads, attempted to predict speeds for each road section using only data found on Ordnance Survey 1:50,000 maps plus information on the positions of 30/40 m.p.h. speed limits obtained from county councils. They divided the roads into five basic categories:

	Default speed (Km/h)	Reduced for....
Motorways	100	
Unrestricted dual	85	hills
Unrestricted single	75	hills, bends
Restricted roads	44.6/59.5*	peak hours
Town centre roads	35-40	peak hours

* 30 and 40 m.p.h. limits respectively

In addition, junction delays were estimated at 15 seconds, rising to 35 seconds for town centre junctions in peak hours. These predictions were tested by 26 subjects who drove routes which together provided a representative sample of driving conditions across the network. The model basically performed very well apart from an underestimation of peak hour delays in central Reading, where traffic levels were at near saturation. Remaining systematic errors were compensated for by an overall correction. Cartographically, the most unusual aspect of such a classification would be the depiction of speed limits. These were incorporated into certain American road maps of the 1930s (Ristow,1964), and other studies have called for their inclusion (e.g.Schiede,1968), but for most map producers extra expense and effort would be involved in procuring the appropriate information. Clear depiction of built-up areas may be considered to be an acceptable substitute.

Once realistic criteria for classification have been determined, there are still choices to be made with respect to representation. The special need, according to Aurada (reported by Schiede,1968) and Connal (1983), is for the portrayal of the road hierarchy in a graded sequence of colours and line widths, so that the driver can clearly see which roads are potentially the fastest, whilst at the same time maintaining an appropriate visual balance between the classes.

2.4 THE REPRESENTATION OF THE ROAD CLASSIFICATION

2.4.1 Practice

According to Bonacker (Schiede,1968,p.164 transl.) 'the creative spirit looks for map symbols to endow the main roads with conspicuousness in the map image, and take the road map to a higher level of development.' The humble tools for such a lofty endeavour are the variations in line width, colour and composition used to depict different line classes, and the colour of the background. In order to maintain sufficient contrast with the background, fully-saturated colours are generally used in preference to thin screened lines. However, red and black lines are occasionally shown screened, and screened line fillings, where contrast with the background is maintained by the casings, are fairly common, sometimes alongside other classes depicted by solid versions of the same ink. This can lead to problems in discriminating between the two, especially where

1) the parent colour is not particularly strong (e.g. reddish brown on Reader's Digest/AA,1986), or

2) the product is overinked in printing, as on several pages of a print run of the Shell Road Atlas of Great Britain (George Philip,1983), where single carriageway primary routes (screened red) are barely distinguishable from the dual versions (solid red).

Financial constraints may restrict the total number and choice of colours which can be used across the whole map. Process colours may be a viable alternative, and are necessary if the map cover is integral with the rest of the map and includes a colour photograph printed in half-tone. A drawback of process colours however is that any colour other than the three primaries and black has to be created by combination (overprinting). Thus if the register is not particularly good, the sharpness of the line might be lost and the perceived colour may alter, especially for fine uncased lines with screened combinations. Consequently these are often avoided, although the Ordnance Survey have made up a 0.2 mm brown line reasonably successfully for the portrayal of B/'unclassified' roads on the 1:625,000 Routeplanner.

One colour convention that seems to be fairly well-established worldwide across many different cultures is the use of red to represent major roads, often as a bright infill between black casings. Hopkin and Taylor (1979) consider that while red has no particular semantic associations with good roads, it may well be the most sensible colour for the purpose. Dark casing lines themselves are commonly utilised, presumably because they provide extra edge contrast and an added dimension of variation for the symbol. In a study by the Consumers' Association (1983a) it was considered that casings, with or without a filling, enable road junctions to be shown more clearly than those where minor roads are depicted by a thin uncased line. Casings also reduce the possibility of confusion between roads and boundaries of area features, but cause a certain amount of extra visual clutter. A further consideration in designing a line is ease of revision. It may be considered most practical to use cased lines and vary the class symbols primarily by filling colour, as if the class of a particular road section should change, it is easier to revise (on the masks) than if a change in width (requiring rescribing or redrawing) were required. However, for the addition of a new or realigned road, less work is required with uncased lines.

A survey of the literature reveals some interesting points about the representation of road series. In North America, various people have commented on or investigated the remarkable lack of standardisation between road symbols on road maps, even those prepared by the same companies. Schultz (1963) reported that the official road maps of each of 11 states on a roundabout journey from Florida to North Dakota used completely different symbols for Interstate Highways, involving 5 different casing colours, 5 filling colours and 3 configurations (uncased, cased and cased with a centre line). An attempt to remedy the situation by the American Association of State Highway Officials, who issued a standard specification for official state road maps (AASHO,1962) was unsuccessful as the overwhelming majority of states did not adopt it. In mid-1970s a thick red line still had 13 different meanings on U.S. state maps (Ward,1977). Including the equivalent Canadian maps, the number of road classes represented varied immensely from 4 to 24 (Anderson,1977). Anderson called for a reduction in the number of classes on these maps and the standardisation of the symbols in order to correlate with the more standardised road signs.

Despite this diversity, certain trends in the use of colours can be detected. Anderson (1975) looked at 80 different North American road maps and discovered that the range of colours used for the top three classes on each map were as follows:

	Class 1	Class 2	Class 3
Red	57	47	46
Green	17	12	7
Yellow	3	11	7
Black	1	5	16
Blue	1	3	2
Orange	1	2	2
Total	80	80	80

Red is clearly the most commonly used colour. However, as it is also used extensively to represent tourist information, the relative importance of road classes and tourist symbols is often not clear (Anderson,1977). The most common colours for the bottom class (not tabulated) are grey, black and blue (Anderson,1977; Connal,1983). As these minor roads are rarely numbered or labelled, blue ones could certainly be confused with drainage features where these are also shown in blue. Anderson also found that, contrary to the opinion of some road map publishers (A.Morrison,1966), most users (>80%) claimed to have consulted the key. This is perhaps not surprising given the diversity of symbol meanings. Despite this, Voisin (1976) has also suggested that the average American, whose overall map reading is poor, can generally pick out specific road classes fairly easily, because of his familiarity with road maps in general (about 5,000 million had been issued free by oil companies in the U.S.A. up to the mid-1960s).

Maps produced in continental Western Europe are somewhat more standardised in design, with a yellow line with thick red casings being the most common symbol for motorways (Consumers' Association,1983a), and usually a decline in colour strength through the series (e.g. red, yellow, grey). Several European producers have described the process of design and/or production of their road maps, most notably Michelin (Bourdon-Michelin,1952; Michelin,1963), who have been experimenting continually with different symbol designs since they produced their first road map in 1910. Their road classification is 3-dimensional, with arteriality, road width (based on the ease of passing and overtaking other vehicles) and hilliness all symbolised. Their colour specifications are tested under different lighting conditions in the laboratory and in the field, with line symbols checked for their distinctiveness in a moving car. Even a relatively generous specification of 8 printing colours is a constraining factor in colour choice. Experiments were conducted in the 1950s with orange as an intermediate road infill between red (main roads) and yellow (other tarred roads), but the idea was dropped as it would have necessitated the compensatory loss of a grey, and the currently-used red-yellow-unfilled progression for the 1:200,000 series was adopted.

Castiglioni (1961) described various maps produced by the Touring Club Italiano which provide an interesting example of the use of colour to promote legibility and visual hierarchy. Base information was clearly subordinated graphically in order to maximise the legibility of the road network. On the 1:200,000 series, roads were depicted in 'active' colours- red, green and yellow- a choice owing more to the 'intensity obtained in a harmonious accord' rather than 'violent contrasts.' On a new edition 'the most saturated colours' (red and orange) indicated touristic importance, while road width was represented by symbol width, untarred roads were shown by interrupting the filling colour, and a green band adjacent to the symbol was used for scenic roads, as on the Michelin maps. On the 1:500,000 series, the spectral (green-yellow-brown) hypsometric tints were changed to 'cold', weakly chromatic greys to facilitate the 'energetic detachment' of the red and orange main road network.

On the Shell-Mair maps at scales of 1:500,000 and smaller (Mair,1963), the administrative classification was abandoned, and a new top category of 'fast roads' was introduced and marked in violet. In Germany this was so successful in attracting new traffic that in some instances the roads ceased to be fast; the classification was revised downwards, the traffic level dropped, and traffic speeds increased again. This points to

1) a very high market penetration of this one particular map, and

2) the dangers of depicting a single class too prominently relative to the others, creating a visual discontinuity in the line symbol series.

The classifiction (and symbol) have nevertheless been retained on some more recent maps of other contries (e.g.Mair,1981). Mair himself (in Schiede,1968) also suggested that roads with high levels of traffic should be shown in a striking colour to encourage people to avoid them, with relief roads with surplus capacity marked in order to spread the traffic load more equally across the road network. This would be of special use at peak periods when drivers are urged to take alternative routes. However, he notes that great care must be taken in determining the balance of visual weight between the road types.

A few other road classifications on overseas maps are described in the literature. Stams' (1964) review of Eastern European road atlases found a great diversity of both the grounds of classification and its presentation. The latter differed from the Western European and North American models mainly because of the paucity of colours employed, possibly for economic reasons. Several used only variations in width, or width and casing state, within one colour. The Road Map of Finland GT (Niemelä,1979) used a graphically 2-dimensional system by depicting

1) the roads' administrative status by the colour of the filling and the road number markers (red with red numbers, red with black numbers, orange, grey)

2) road width/ number of lanes by symbol thickness, including a solid or dashed centre line and/or casing.

Finally, Igbozurike's (1974) suggested specification for the Africa road map used a classification based on road quality (number of carriageways and surface type) represented by 2 widths and 4 fairly similar colours (dark red, light red, orange and brown). Overall it can be seen that the long wavelength end of the spectrum is clearly favoured for road symbol colours in many places.

In Great Britain the classification criteria are normally based upon Department of Transport classes, and this has encouraged more uniformity in their manner of presentation. This standardisation is largely due to the fragmented state of the road map market: individual producers cannot afford to undertake their own extensive surveys, unlike the large foreign companies, and are mostly reliant upon Ordnance Survey (O.S.) mapping and their own consultations with relevant (public) authorities. For example, variations in road width have generally played a minor role in classifications, except on maps specifically designed to show them (Geographia,1980): the single track A and B roads in

Scotland and dual carriageways distinguished by the O.S. are generally about the limit, and the latter have only become common over the last 15 years.

In terms of representation, there has been a distinct historical evolution, starting with the earliest colour printed road maps- the cycling maps of the late 19th century- where the 'best roads' were often coloured red (occasionally brown or yellow) while all other detail was in black (Nicholson,1983). The most influential scheme for many years was introduced onto the O.S. 4th series 1-inch maps in 1918, with thin black casings and red, brown, yellow and white fillings. As the road system developed and traffic levels increased, this was later extended by the addition of thicker casings for dual carriageways and pecked fillings for single track A and B roads in Scotland. In the last 20 years, standardisation of colouring has increased further with the adoption of a blue code for motorways and green for primary routes in line with signposting, to the extent that every widely available specialised road map now incorporates these conventions, either in the colour of the lines, road number markers or names of primary route destinations.

The designs of the O.S. quarter-inch/1:250,000 and tenth-inch/1:625,000 maps have gradually become more specialised since the days when they were designed as general purpose maps but used mainly as road maps. On the 5th series quarter-inch, line widths were increased (and alignments consequently further generalised) to make the road symbols more prominent, with the embryonic motorway network depicted by even thicker red lines (Griffith and Kelly,1965). After metrication, the series became known as the 'Routemaster', and the 1:625,000 map became the 'Route Planning Map' and later the 'Routeplanner', undergoing major design changes which induced mixed reactions from reviewers. Motorways and Primary Routes were changed to blue and green respectively, so that only the 'other A roads' were still marked in red. All three classes shared the same line width. According to A.Morrison (1979b), this represented a reversal of the order of their conspicuity (red, green, blue) making it difficult for the user to decide which were the best roads. The use of blue for motorways still continues to attract criticism (e.g.Dixon,1986) now that it is a well-established convention. Process printing was also introduced, so that the green Primary Routes were created by overprinting and the built-up areas were represented in a stronger solid yellow. This led to a further design change on the 1980 map (A.Morrison,1980b) where overprinting was avoided for the major roads by the use of process primaries- cyan for the wider motorway symbol and magenta for A roads, with the Primary Routes almost imperceptibly wider than the 'other A roads'. According to one reviewer (Dixon,1986) this design lacks clarity because of a lack of distinction between the magenta and brown (minor) roads, and the introduction of a pale yellow background.

Some articles on non-official road maps also deserve mention. Gatrell (1966) described his design for the Mobil map series. He explained that he chose triple red lines (thick casings and a thin centre line) with solid yellow infill for motorways, so that they would be the most visually prominent feature of the road network. The next category, trunk roads, were red within a black casing, these two colours being necessarily part of the 4-colour specification as the Mobil logo had to be reproduced on the cover of each map. The uncased red line for A roads reflected its relative importance and obviated the need for the tedious and time-consuming retouching of junctions that would be necessary if scribed cased lines were used. Reference was also made to a 'Which?' report which had found the colours to be unsuitable for night use, presumably because of their similarity with each other and with the amber hue of sodium street lighting.

Goldring's (1978) review of the development of Esso road maps in Britain makes the following points about colour (p.71):

'Colour is of more fundamental importance than the casual map-reader might realise. Without his being fully aware of it, it contributes a great deal to his understanding of the map he is using. Different colours are used to separate important information that every user wants, such as roads and towns, from details which are only occasionally required. Immediate information is printed in black and red, colours which catch the eye quickly. Background information, not essential in planning a route, like rivers and forests, is printed in blue and green, more recessive colours. Strong, solid blue for motorways breaks this rule: in this case the blue relates to the colour of the motorway signs.'

He also noted the use of single, uncased lines for most road types for ease of revision, and what he claimed was map-makers' obsession with the whiteness of paper, given the necessity of a white background to maximise the clarity of the map. Thake (1979) disagrees, commenting that in the design of the Automobile Association (AA) road maps, background layer colouring is useful to <u>offset</u> the dominant visual influence of the road network. In fact, much of the diversity of colour specifications of British road maps comes in the background colour, which varies between maps according to producers' assessments of the usefulness and attractiveness of areal information on relief, scenic areas and woodland, and its effect on map clarity. The inclusion of such information often varies between differently packaged and marketed versions of the same map.

An interesting example of an integrated programme of market research, design and evaluation for a route planning map is provided by Andrew Holmes' design of the

Central London Bus Map (Braidwood,1981; LPTRG,1985). One of its features was the colour coding of routes into four categories according to their cardinal direction using maximally contrasting colours (blue, red, green and black) for clear distinction even under sodium lighting. However, research showed that the significance of the colours was not understood by map users, who generally ignored the compass-sign key, although it was useful as an elimination factor, as the numbers of the bus routes were printed in their respective colours. The green colour was eventually changed to maroon to avoid confusion with the 'Green Line' bus services.

In summary, it seems that with the variety of different colours used for road classes, even within one country, no associated convention is as sacrosanct or as basic to map reading as for example blue=sea. The only one that appears to be almost universal, and violated in many other designs only by the addition of recently-constructed high-speed roads, is red= major roads, implying that the colour has some innate advantage.

2.4.2 Preference

Road map users have occasionally been asked to evaluate different map products, either by planning routes on them or just inspecting them. In most cases, the criterion used to evaluate them is people's subjective impression of the clarity and ease of use of the maps. The Consumers' Association have conducted several surveys in this vein (1963,1971,1979,1983b). The first three involved a small panel of motorists, but in the most recent test 63 drivers were each asked to choose routes on four different maps. The most general conclusion that can be drawn from these surveys is that maps where the road classes are very distinct and difficult to confuse are generally preferred. Special complaints were made about the monotony of the smaller scale maps in the 1971 tests. An interesting effect of background colour was observed in 1963 on the 'Auto-mapic' atlas, which showed some A roads in red and others in green and used varying background colours, with the result that the green roads appeared to be the more prominent in East Anglia, while the red roads stood out in Cornwall.

In 1983, distinct and saturated line colours were again preferred to the 'anaemic' colouring of the Reader's Digest/AA New Book of the Road, on which the slight difference in casing width to distinguish between single and dual carriageways was considered to be very unclear. Complaints were made about the 'unfamiliar' colouring of roads on the O.S. Motoring Atlas, a bad example of process printing where the magenta A roads appeared to be a very weak almost violet colour, and were considered to be 'difficult to follow.' The Collins Road Atlas, utilising Gatrell's Mobil design, was considered to 'bring the main roads to life well', but the less distinct line colours were 'unfamiliar and unpopular', and detail was obscured in urban areas by the dark grey built-up area symbol. It is interesting that designs without green primary routes and blue motorways are now considered to be unfamiliar.

Carpenter (1979) found that the most commonly-used maps by her interviewees were those published by Esso, B.P. and Shell, which were preferred as they were considered to be 'clear and easy to use.' Sheppard and Adams (1971) conducted a comparative test of the O.S. quarter-inch map and one of the same scale produced in 1969 by George Philip for the Shell Touring Service. The O.S. map had layer colouring while the Philip map had a white background and generally more prominent road symbols and place names. While A roads on both were red, Philip also had trunk roads printed in a contrasting colour (blue). 128 male vehicle-owners were asked, as an open question, which map they preferred and why. Interestingly almost all the subjects (96% overall), including those who preferred the O.S. map, thought that showing trunk roads in a different colour from other main roads was a help, although it is not known what they understood by the word 'trunk'.

Of the 83 (65%) who preferred the Philip map, 41% mentioned that the roads were clearer and 23% said that they preferred the white background. Of the 34 (44%) preferring the O.S. map, only 9% said that the roads were clearer on it, and the main reason for preference was the better idea of the countryside given by the relief information (52%). In general there is an interesting correlation with the model of required information presented in section 2.2. Overall the pro-Philip group expressed a preference for a less detailed map with the 'basic' road and place information clearer and the main roads more obvious, as if their normal use of maps was for fastest-route planning. The pro-O.S. group preferred more detail of terrain, minor roads and villages, as if they normally used maps in more leisurely circumstances. Other studies have also found this sub-group who prefer an 'information-rich' map.

A.Morrison's (1974) tests included questions on subjects' preferences with four different designs of speed map, assessed in terms of how confusing and how pleasing the colour schemes were perceived to be. Maps with colours ranging widely across the spectrum were generally considered to be less pleasant, although no problems were mentioned in distinguishing between them. Overall the red, orange, yellow, grey, black scheme (the first four colours having a black casing) was preferred, and green-cased yellow, green, orange, grey, black (black casings on the middle three) was disliked.

Overall, it would seem that with respect to user preferences, the most important design criterion is that the symbols for the road classes are clearly and immediately distinguishable, although on speed maps, where the colour (i.e.class) often changes along a stretch of road, colour harmony may also be important.

2.4.3 Performance

More objective performance-based evaluation of road maps is not without precedent: a few investigations have compared road map designs in terms of the routes chosen by their users. The normal method used is to ask half the sample to choose a route on one map, and the other half to choose a route between the same places on another map of the same scale but of a different design. The routes chosen can be compared in two ways:

1) the difference between the actual roads used by the two groups. This is shown most clearly by synoptic pairs of flow-line maps, and is more difficult to analyse by formal statistical methods.

2) the difference in some overall measure such as average distance or time.

In Carpenter's (1979) study, the 39 subjects were shown maps either from the AA (19) or RAC (20) handbooks. These were in fact fairly similar in design, the main differences being that on the RAC map relief was shown, urban areas were paler and fewer place names were marked. Subjects were asked to plan a route from Burton-on-Trent to Scunthorpe as if they were going to a wedding and had plenty of time, and whether they would change the route if they were in a hurry (only four would). This particular journey was selected as it required the subject to choose whether or not to use the motorway and whether to by-pass or go through certain towns. Altogether 28 different routes were selected. The routes chosen on the AA map were slightly less diverse with more use of the motorway. Carpenter attributed this to its more detailed and cluttered appearance encouraging people to take the more straightforward (and slightly further) motorway route. Thus the average distance of the AA routes (95.4 miles) was nearly 3% further than that of the RAC (92.7 miles); overall the longest route chosen was more than 35% longer than the shortest. Carpenter concluded that the map used for route planning could be an important factor in route selection, and called for a larger scale survey.

Sheppard and Adams (1971) also asked their subjects to choose the route(s) between Stanhope and Hexham in Northumberland that they would take in fine weather and with snow falling. The quality of the routes was not evaluated statistically. However in snowy conditions, the difference between the routes selected from the O.S. and Philip maps was almost entirely explained by the representation of relief on the O.S. map. In fine weather, the only big difference was between Edmundbyers and Hexham, where Philip users tended to take the B road via Blanchland, while most of the O.S. users headed for the slightly less direct route on the A68. Unfortunately, Sheppard and Adams do not relate this difference back to the map symbology, so it is not possible to determine its cause(s).

The most comprehensive survey of the effect of map design on route choice was undertaken by A.Morrison (1974), the main purpose being to determine the relative effectiveness of 'conventional' and speed-based road classifications (i.e. different information represented by the same line symbol series). However, 36 subjects were each shown a set of maps with six different symbol series. The specific objective in each case was to reach the destination 'as soon as possible', and the routes were evaluated by the time they would take to drive according to data from Morrison's own speed surveys. The aim of the symbol series design was to order colours in terms of their 'psychological impact', utilising part-spectral sequences (e.g. red, orange, yellow) and particular colour associations such as green= go/fast, and black= town/slow (A.Morrison,1971). In two of the schemes the speed data were depicted by line width rather than colour.

The results showed no statistically significant differences between the schemes in terms of performance, although the two series based on red, orange, yellow, black performed the best. Morrison concluded (1974,p.108) that 'it merely shows that all the schemes considered worth testing were reasonably sensible. One might also conclude that individuals are quite flexible in their ability to assign meaning to colours.' An interesting observation was that the performance of the better schemes was made significantly worse by the presence of a key. Morrison interpreted this as an indication that the schemes were self-explanatory, and the complex key may have confused. The key had little effect on the more familiar 'conventional' maps. Another interesting point is that the red, orange, yellow, grey, black scheme which performed the best was also the most preferred. Otherwise, unfortunately, the literature can provide no specific guidance as to whether people perform route choice tasks better on maps which they prefer.

Clearly map design does affect route choice. However, it has not yet been demonstrated that differences in line symbology alone, between two maps using the same road classification and of otherwise comparable design, can have significant effects on the quality of routes chosen. Consequently, this was investigated in a pilot study described in chapter 3. It can also be seen from the above that several problems concerning the representation of road classifications remain to be studied. What are the causes of the relative 'psychological impact' of colours, and how important is the influence of newer associational (learned) colour conventions which may conflict with 'visual order'? What is the relative significance of variations in the colour, width and character of line symbols in creating a perceived hierarchy and influencing route choices?

3. THE PROCESS OF MAP-BASED ROUTE CHOICE

The vast majority of people intending to drive a considerable distance over unfamiliar roads use a map to help them plan a route. In surveys which have been conducted on representative samples of British motorists, the figure varies from 86% (Gray and Russell,1962) to 95% (Sheppard and Adams,1971), with the remainder either relying on routes supplied by motoring organisations and/or personal advice, or not planning a route in advance at all. In a study by Connal (1983) in Canada, the map-using proportion was nearly 97%, while in the U.S.A. (according to Janssen,1975) practically 100% of motorists use road maps when out of their own state.

How are maps used in the process of planning a route? The first important influence is the kind of route the user is seeking. For each individual, this may vary with factors such as the particular purpose of the journey, whether any time constraints are involved, the overall length of the trip, the presence of children and the driver's mood (Carpenter,1979), and with the particular time at which the journey is to be undertaken. However, the basic route characteristics sought in a given situation seem to be that it is either quick, short, scenic, or easy to follow or drive, or that it avoids congestion, sticks to particular favoured road classes, or fulfils any combination of these requirements. In some situations there may effectively be no choice, either because there is no known alternative to the route being taken, or, as with some commercial vehicles, the route is specified by the transport operator (Outram and Thompson,1978). People with low spatial ability seem to be more concerned with finding a simple route with few turns, and less concerned with avoiding congestion (Streeter and Vitello,1986). However, in every study of route planning criteria, whether the information was derived from questionnaire surveys of inter-urban (Outram and Thompson,1978) or intra-urban (Benshoof,1970) travel, from verbal protocols obtained from in-depth interviews (Carpenter,1979) or postal questionnaires (Connal,1983), or even by inference from chosen routes (Outram,1976), the most commonly-sought objective by far, either singly or in combination with others, is the minimum time route. It would seem that in Britain about 75% of routes planned using maps are selected with this goal in mind (Morrison,1979a), a proportion that does not vary significantly with age, sex, social class, education or mileage driven, the only deviant group being drivers of road goods vehicles, who seek the minimum time route on 95% of occasions.

The problem with trying to satisfy a minimum time criterion from a map is that journey times are rarely shown on road maps (Michelin,1985 is a current exception). Even if a speed-based classification is shown, the map user must still evaluate the trade-off between shorter distances and faster roads. For example, in a study by Streeter and Vitello (1986) individual roads were generally not selected if using them would make the total route length >20% more than the straight line distance, but no information is supplied on relative road speeds. With short routes (<6cm map distance) in A.Morrison's (1974) experiment, subjects actually planned quicker routes with 'conventional' maps than with speed maps, implying that they were diverting too much in order to access the faster roads. Assessment of relative distance may be a problem in itself, especially where the roads concerned differ considerably in terms of bendiness. Numerical indication of distances may be very helpful, although it is generally only given comprehensively on M and A roads. With 'conventional' road classifications, the user must also make his own assessment from the information available to him, of the relevant road and traffic conditions which will constrain his travel speed. Faced with this problem, many people develop specific cognitive strategies through which they can process the available information to find either the fastest route, or effectively some compromise between it and the most direct or easiest to follow. The importance of cognitive strategies has been noted by Griffin (1983), in stressing that the map user must be accepted as an active participant in the process of cartographic communication.

The difficulty for the researcher is accessing these strategies. Verbal protocols or written reports are used, but are clearly reliant on the subject being conscious of, and able to verbalise, his strategy. Carpenter (1979,p.7) found initially that 'when questioned about route planning, people answered "We look at a map", and it was not particularly easy for them to explain how they gained information from a map, since it seemed such an obvious process.' The best method found was to ask the subject to talk through making an actual route choice off a map. Various specific examples discovered in this way have been quoted. Sheppard and Adams (1971) reported that in attempting a cross-country journey in fine weather, most people would take the most direct route on A and B roads, i.e. avoiding 'unclassified' roads. Carpenter herself mentioned that many of the people she interviewed would use only A roads and motorways when seeking the minimum time route, although some might use some lower class roads as short cuts on longer journeys. A further group looked for the 'fastest direct' route, which involved neither major detours to get onto a motorway nor short cuts on minor roads. Many people prefer to avoid minor roads in areas they do not know (Streeter and Vitello,1986). Cognitive strategies may also be based on specific methods of presentation: for example, people might approach a map with the preconceived idea of searching for thick blue lines, which they assume from experience to be motorways.

These strategies may vary considerably in terms of their detail and the complexity of the information the user hopes to glean from the map. Interestingly, 92% of the drivers interviewed by Gray and Russell (1962) considered road map reading to be 'easy' or 'fairly easy', although this could imply that they felt perfectly capable of accomplishing a relatively complex task, and does not necessarily mean that they considered it to be simple and self-evident. Another important user variable, which may reflect the complexity of his strategy, is the amount of time he is willing to spend on reading the map (Castner,1978), and therefore the number of alternative routes he is willing to consider in a given situation. This will presumably influence the quality of his route choice, but is an area about which no information is currently available.

The route choice itself is the product of the interface between the user's knowledge and experience, as embodied in his cognitive strategy, and the perceptual stimulation of the map, mediated by any other specific knowledge he has about the area concerned (his cognitive map). The specific manner in which map information is processed in route planning has received even less attention than the processes of memorising and following routes. Clearly the first requirement is to search on the map for the origin and destination of the trip. Where they can both be seen simultaneously, the general method used by Carpenter's (1979) subjects was to imagine a straight line (i.e. the shortest distance) between the origin and the destination, and follow the roads nearest to that line which met their particular criteria, making a special effort to avoid apparent bottlenecks. Where the endpoints are not seen simultaneously because they lie on different pages of a road atlas, or on different map sheets where they cannot be fully extended, further complications arise. In such cases many people appear to adopt more of a 'satisficing' strategy, planning to the edge of the page in what they perceive to be the general direction of the destination, and often not replanning the route if it becomes somewhat deviant when the destination is neared (Bartlett,1958), especially as without the global overview it may not even appear to be deviant. There is also a tendency, as with some other long journeys, to plan the route in sections between intermediate destinations. Armstrong (1977) considers this to be a significant cause of sub-optimal decisions.

Effectively there exists for each choice a broadly elliptical 'corridor of possibilities' centred on the straight line between the endpoints. Where a short route is sought, the corridor will be much narrower than in cases where the user is prepared to go well out of his way to get onto quicker roads. Evidence that people do tend to set off on the straightest route is provided by map 2 of our own experiment 3 (rear pocket) between points C and D, where there is a basic choice between a northern route via Venn and a southern route via West End. For journeys from C to D, 71% of subjects set off on the northern route, which points more towards D at the start. However, when the journey was planned in the reverse direction (D to C) on the same maps, 80% selected the southern route, even though it starts off on roads of lower classification, because again it heads initially in the better direction. The roads picked up from the origin will begin almost inevitably to diverge from the overall straight line, drawing with them the map user's focus of attention as he traces his possible route. Then, as he 'homes in' on his final goal, the relevant straight lines are those between decision points along the road and the destination.

The important question is, within that corridor, how does the representation of the map information influence the choice of route, and over which of the relevant presentational variables does the map designer have effective control? Given two lines of equal length and directness, what will determine which one is the more likely to be selected as the user makes his assessment of probable road and traffic conditions? Cartographic and psychological experience suggests that several factors are involved (in order of declining controlability):

1) the perception of the road classification through the perceived importance and continuity of the line symbols,

2) other information on road quality, such as road numbers, the extent of built-up areas and gradients,

3) the local density of the road network and the sinuousity of the line symbols, and

4) various factors over which the cartographer has little control, such as the directional continuity of a line, its orientation, position on the page and aspects of line/network shape.

In the first category, the saliency of the line symbol is clearly important. For a line to be prominent in a map image, it must be able to catch the eye, emerge visually from its context, and carry a relatively strong magnitude implication. In these respects, line width, colour, character and contrast with the background are crucial, and their perceptual and cognitive effects are examined in detail from chapter 4 onwards. Continuity cues (discussed in section 4.5) enable the decoding of the road classification without reference to the key, by examining which lines take precedence at cross-roads (occlusion) at least. Also, as mentioned above, on maps where the road classification is administratively based, higher classes tend to occur in longer uninterrupted stretches, to which lower classes are tributaries. Certain classes also have inherent figural continuity, such as primary routes (or where these are not distinguished, A roads in general) which together form a linked network that constitutes a clearly recognisable level of structure. In such cases, users can attend selectively to this structure at the cost of slower utilisation of other levels (Kinchla et al.,1983) such that a common cause of poor routes selected from 'conventional' maps might be too little use of the non-primary network. In classifications which incorporate aspects of road quality (and not just direct representation of likely travel speeds) this continuity will often be reduced, as class may change from section to section along the same administratively-defined road (e.g. from single to dual carriageway) such that continuity becomes a less useful cue, and symbol prominence becomes more important.

The map designer also has potential control over the representation of other relevant information. The influence of road numbers may be affected by their prominence, which can be varied by the colour, size and weight of the characters, and by their inclusion in outline or solid boxes. These often straddle the line symbol and can disrupt its perceived continuity. With administrative classifications, the redundant Department of Transport (D.Tp) prefix letters are sometimes excluded, which tends to deemphasize somewhat the basis for the classification. In order to give an appropriate weighting to the significance of built-up areas as places to avoid in fastest-route planning, the depiction of their extent in a prominent colour (without detriment to the legibility of the urban road network) may also act as an effective repellent. Some map users may attempt to avoid areas with a dense road network (as likely concentrations of congestion and delaying and disorientating road junctions) and the more apparently sinuous roads. At smaller scales, the cartographer can exert a certain amount of control here through generalisation, by retaining extra minor road detail in congested areas and stressing the relative bendiness of roads.

However, the factors over which the cartographer has no real control may also have a considerable influence upon route choice. The directional, or angular, continuity of a road may be important, because in any human spatial behaviour, people appear to continue along the line of least resistance for as long as possible, delaying turns until they are absolutely necessary. Aspects of position, orientation and configuration (e.g. an unusual or interesting spatial arrangement) of particular lines may also affect their attention value. For example, items in the upper left quarter of a field of view are more likely to attract attention than those in the bottom right quarter (Woodworth and Schlosberg,1955). For the remainder of this thesis, however, attention will be concentrated on the first two categories where the potential influence of map design is greatest.

Eventually a route is chosen. It must then be recorded in some way to enable recall when it is being followed. Sometimes it is marked onto the map, but more commonly it is either memorised or transcribed into a route list of intermediate destinations and/or road numbers that should appear on the road signs on the way. Surveys indicate that up to 50% of drivers/ navigators use route lists, generally compiled from their own maps (Astley,1969; Carpenter,1979), although Connal's (1983) figure for Canadian drivers is lower (28%). The content of the lists is fairly evenly divided between place names, road numbers and both (Gray and Russell,1962; Carpenter,1979). Clearly major roads (in an administrative sense) and major settlements provide the simplest and most certain cues, so listed routes are often likely to be more a compromise with an easy-to-follow route than strictly the fastest, with short cuts and cross-country alternatives generally being eschewed. The route list can either act as a direct navigational prompt for the driver (and/or navigator) to correlate with road signs, or as a source for memorising the sequence of operations to be undertaken. This is easier to do from verbal instructions than from map images given the inherently linear (sequential) rather than spatial (2-dimensional) nature of the task (Wetherell,1984).

This is not to say, however, that the map is of no use in the car. Irrespective of the navigational cues it can provide (which are not the present concern) there are many occasions where route planning is undertaken en route, either because of a change of plan or as a response to external circumstances such as unexpected road works, weather or traffic conditions, or losing the way. (In Gray and Russell's (1962) survey, 36% of drivers admitted to sometimes losing their way, including 4% who did so on most journeys over unfamiliar roads.) Maps are used in the car on about 75% of long journeys over unfamiliar roads (Morrison,1979a). About one half of drivers would have a navigator with them in such circumstances. Of the remainder, just under half would attempt to glance at the map whilst moving (Morrison,1979a; Connal,1983) while 66% of Connal's sample would sometimes stop and look at the map. Some in-car decisions clearly have to be taken on the spur of the moment, where the ability of the map design to elicit a fast and accurate response is most severely tested, and is potentially critical for road safety (where the driver is dividing his attention between map and road).

An important side effect which would hopefully materialise from improved route planning maps is the increased confidence of the driver/ navigator in his own route selection, so that he would be less likely to be deflected by conflicting signposts. However well the route is planned, the resource savings will not be fully realised if it is not properly followed.

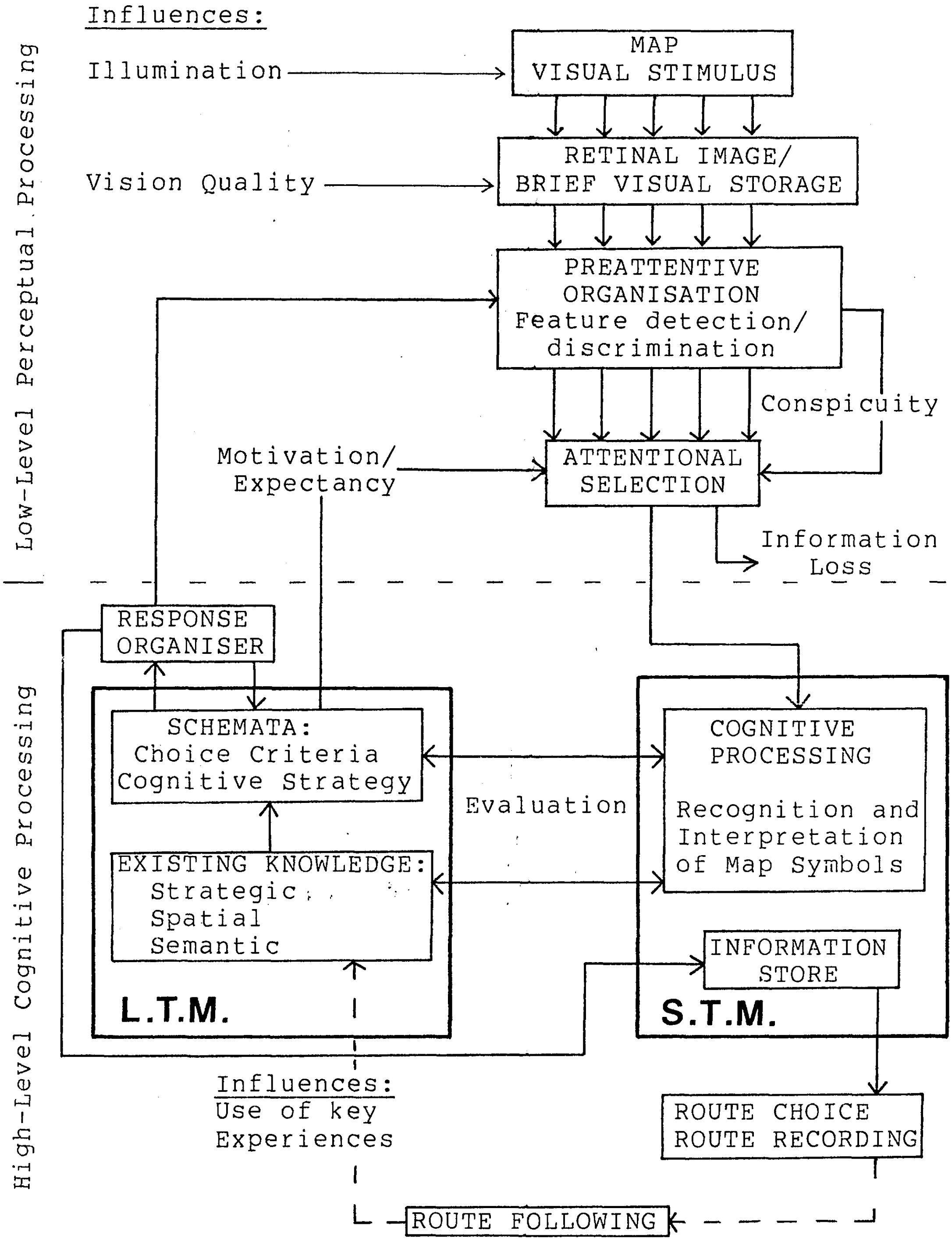
3.1 Visual Information Processing in Map-Based Route Planning

A model of the visual information processing system as applied to the process of route planning is presented in figure 3.1. The boxes are not intended to represent the physical locations of activities within the eye-brain complex, but simply describe stages of processing, each of which takes time and has a limited capacity. A useful distinction can be drawn between the early perceptual (low-level) processes, which are performed spontaneously and are unavailable to consciousness, and the later cognitive (high-level) processes, which involve a conscious effort of mind in the scrutiny of display elements, and the application of the map reader's own knowledge (Castner,1978). However, even this distinction is flexible, as cognitive skills can become perceptual (i.e. effectively automatic) by frequent rehearsal.

As Dobson (1979b) has stated, it is a fundamental assumption of map design research that changes in design can attract the reader's attention and allow the cartographer to orchestrate the map reading process. If this is the case, it is in the perceptual processes that he has the most immediate control. Phillips (1984) has argued that in all the studies which have found a large effect of map design on subject performance, the main differences occur at the lower levels of processing where the information load on the system is greatest, and the brain must progressively reduce, structure and sample a huge amount of raw data. The operation of these stages appears to be largely innate as it varies little between individuals (except those with physiological anomalies such as defective colour vision) and is minimally affected by subject factors such as age, education level and culture.

On the other hand, map users' experience is not instantly manipulable by the map designer, although there is clearly a slow feedback present as users become accustomed to specific map designs and perhaps begin to improve their performance with them. Behaviour at this level is learned and potentially modifiable (Keates,1962). In the higher levels of processing, variations between

Figure 3.1 Visual Information Processing in Map-Based Route Planning



L.T.M. Long Term Memory S.T.M. Short Term Memory

Sources: Engel,1971,1977; Haber and Hershenson,1973; Neisser,1976; J.Morrison,1976; Dobson,1979a, Eastman and Castner,1983; Head,1984; Eastman,1985.

individuals also become far more marked, not only because of the uniqueness of their experiences (so that factors such as age, education and culture may exert an influence) but also because of variations in their 'visual memory ability', or capacity to retain spatial information, which appears not to be improvable through training, and limits the performance of spatial tasks (Thorndyke and Stasz,1980). The more complex the task and the stimulus, and the longer the user looks at the map, the greater the relative importance of cognitive processing (Castner,1978; Eastman and Castner,1983), and consequently the greater the importance in experiments of obtaining a representative sample of users.

Following through the model, control of the focus of attention is clearly a critical stage in terms of map design, as the scarcity of further processing resources, owing to the small size of the high-acuity fovea, imposes a severe constriction on information load, with objects selected if their appearance suggests they are worthy of further inspection and likely to yield useful information (Castner,1978; Dobson,1981). This selection, effected by the guidance of fixations, is sensitive to both cognitive factors- the purpose of viewing the map (motivation) and prior information held by the user (expectancy)- and the perceptual attraction of display elements in the visual periphery (Yarbus,1967; Dobson,1979b), modulating between the two influences, according to the content of the cognitive strategy, in a form of 'intelligent scanning' (Dobson,1979b). It might be expected that the perceptual influence (conspicuity) would be strongest amongst individuals with the least developed cognitive strategies.

In subsequent processing, symbols are interpreted and evaluated with reference to the individual's knowledge of the area concerned (spatial), the meaning of the map symbols (semantic), and his cognitive strategy of how to use the map information to realise his route choice criteria, and he will reinterrogate the map until a satisfactory conclusion is reached. The resulting route choice is either memorised, recorded or both, and if it is then followed by the route planner, adds to his experience.

3.1.1 Conspicuity

For the map designer, conspicuity merits special consideration. There is firstly some confusion in the literature about the meaning of the term. Easterby (1980,1984) has clarified the distinction between

-'visibility', or 'legibility', the ease with which elements can be detected and discriminated from each other, and

-'conspicuity', the relative emphasis of visual messages, or the ease with which they are detected in the presence of competing information. In other words, conspicuity is relative visibility, scaled across the visual field, or a kind of signal-to-noise ratio.

However, Easterby's definitions ignore the basis for this distinction, namely the process of attention. Engel (1971) defines visibility as the extent to which an object can be seen while attention is directed towards it and it is in the principal area of focus, whereas conspicuity is 'that combination of properties of a visible object within its surroundings, by which the attention of the subject is attracted via the visual system' (Engel,1969,p.90), and is consequently the more primitive phenomenon. In terms of map-based route choices, it might consequently be said that conspicuity is the more important property with respect to an overall route catching the eye, whereas visibility would relate more to specific choices made in tracing the route where the prominence of the symbols is important.

Thus conspicuity is the 'external determiner' of attention (Engel,1969), inducing involuntary eye movements and attention shifts by a spontaneous perceptual process which cognitive schemas are too slow to suppress (Engel,1977). In this way, objects are 'acquired' from peripheral vision for fixation by foveal vision, which enables them to be processed to the level of identification (L.G.Williams,1967a; Dobson,1983c). A conspicuous object is easily noticed in its context and can be found quickly (Engel,1971). Obviously the visual emphasis of objects is relative in any given view. The locus of selective attention can only be directed towards a small part of the visual field at any one time; consequently only a very small proportion of the features we see are actually noticed (Engel,1971).

The pattern of attention allocation is particularly important in the reading of graphical displays: in text reading the location of fixations is largely internally determined by knowledge of the specific linear arrangement of the stimuli (i.e. left to right and down the page for English) and peripheral vision is barely used (Dobson,1979a; Phillips,1981), but it is virtually impossible to read a map without peripheral vision (Phillips,1984), as the conspicuity of the objects structures visual search (Engel,1977). According to Woodworth and Schlosberg (1955), the size, intensity, colour and motion of the stimulus are all external determiners of attention which determine its conspicuity. Other important influences are the density of items in the display, and the location of the stimulus in the visual field (Dobson,1981): a peripheral stimulus will not normally be able to compete successfully for attention if a similar one is present nearer the point of fixation, or if it is further out than about 12° of visual angle (Phillips,1981).

Conspicuity is very much related to concepts of graphical dominance and visual hierarchy (Dobson,1979b), and has often been discussed on maps under the guise of other highly interrelated terms such as contrast and the figure-ground relationship (Dobson,1983c). The operation of stimulus variables in the creation of conspicuous lines is discussed in the succeeding chapters.

3.2 Map Design and Route Choice: A Pilot Study

A pilot study was set up with the aim of discovering whether differences between maps in terms of graphic design alone could create significant differences in the routes chosen from them. It was also hoped

1) to obtain a fairly detailed picture of the way people use map information in route selection,

2) to isolate the proportion of the variation in chosen routes over which the cartographer can exert an influence, and

3) to gain an impression of the relative significance of the design variables. This information would be a useful guide in setting up the subsequent main series of experiments.

Thus it was important that the experiment should involve the making of route choices from maps, both in order that the chosen routes could be assessed and, following Carpenter (1979), to act as a catalyst for subjects to comment in some depth about their methods of route choice. To this end, fairly detailed interviews were carried out (average duration 15-20 minutes) in relaxed circumstances, using four pairs of existing published maps. The subject was asked to choose the minimum-time route between a pre-marked origin and destination on one of the maps from each pair. Four routes per interview was considered to be feasible without making undue demands on individuals' time or introducing experimental fatigue. For each pair of maps, half the sample saw one while the remaining half were shown the other. People were interviewed in their normal work environment. This was a less pressured and more realistic location and atmosphere for pre-journey route planning than would prevail in on-street interviews, as the individual was not in a hurry to get away, and was able to take as long as he normally would over route selection.

32 volunteers, 22 males and 10 females, acted as subjects. They were all postgraduates or final year undergraduates in the Geography Department at the University of Glasgow who were drivers and/or navigators. At this stage external validity (cf. section 7.1) was not sought. It was more important to have a relatively homogenous sample in order to isolate the effects of the target variables by reducing the variance from other known sources. Use of this sampling frame controlled age (which ranged from 20 to 32), level of education and journey purpose (i.e. nobody drove as part of their work), and also tended to homogenise levels of experience and regularity of route planning (2 to 15 times a year with a mode of 5) and probably spatial ability. It also enabled relaxed indoor interviews to be carried out in standardised illumination ('warm white' fluorescent strip lighting), and allowed the experimenter to reaccess the subjects after preliminary analysis of the interviews to check his assessment of their cognitive strategies.

The journeys to be planned were located in parts of Great Britain remote from Glasgow and likely to be unfamiliar to the vast majority of the subjects. In fact, no more than two subjects had any significant knowledge of any of the four areas. The areas varied in terms of the density of settlement and roads, but each included a variety of road types providing a number of realistic alternative routes. In every case, both ends of the journey could be viewed simultaneously, the straight line distance between them varying from 12.9 to 24.9 cm. In each pair, the maps were of the same scale and, except in the case of pair 3, had the same road classification as far as possible (i.e. excepting the odd error) and were equally up to date. Consequently the main differences were not in the road information itself, but in the manner of its representation by line symbols, and in the depiction of other relevant information such as road numbers and settlements.

Pairs 1 and 2 were at the larger 'touring' scales of 1" to 3 miles/1:200,000, and three of the four maps were from road atlases. In recent years this has gradually become the modal scale and format for British road maps, with the increasing market penetration of the large format softback road atlas in particular. Pairs 3 and 4 were at the smaller, less detailed scales of 1:625,000/ 1" to 10 miles, three of the four maps providing single sheet coverage of Britain or England and Wales. Pair 3 was unusual as the maps do not share the same classification, 3B being a map of road travel speeds produced by Alastair Morrison under contract to the Transport and Road Research Laboratory. Otherwise the choice of maps was very much constrained by the availability of comparable pairs of published maps, with the unfortunate consequence that several interesting designs such as the Collins Road Atlas, with an unusual line symbol series, could not be incorporated into the experiment. The specifications of the maps are presented in appendix A.

A self-reporting questionnaire (overleaf) was compiled to guide the experimenter through the two stages of the interview. Firstly, personal information was obtained from the interviewee about such factors as age, colour vision, frequency of long-distance route planning, the particular maps he tended to use and the type of route he generally sought. Then the subject was shown the four maps in turn, and was asked to select what he considered to be the minimum time route from the marked origin to the

PILOT STUDY QUESTIONNAIRE

NUMBER: date:
Location/illumination:

Sex: Status:
Age: Nationality
Normal colour vision?

1. Do you drive?
If YES, how long? own vehicle?
Estimated annual mileage?
2. Do you ever plan routes or navigate for other drivers?
(If answer to 1 and 2 both NO, terminate interview.)

3. Roughly how many times a year do you use maps to plan a route (exc. in towns)?

4. Particular maps used

5. Normal route choice criteria

4 road maps, one route on each, quickest route (by car), daylight off-peak conditions

ROUTE 1 Map: Key used?:

8. Reasons/ comments

ROUTE 2 Map:

9. Reasons:

ROUTE 3 Map:

10. Reasons:

ROUTE 4 Map:

11. Reasons:

marked destination in off-peak daylight hours and in the absence of abnormal weather conditions. He was not shown the key to the map symbols unless he specifically asked to see it, although on the speed map (3B), where the key was prominent, the classification principle was explained to him if he was unfamiliar with this type of map. He was then encouraged to comment on his reasons for choosing his particular route. The interviewer noted the route using a simple numerical coding system, and used a cassette recorder to record the verbal protocols, unless the individual concerned did not wish to be taped.

A simple experimental design was used to determine which map the subject saw from each pair and the order in which he saw them. The possible map sequences were arranged into 4x4 Latin Squares, and for every odd-numbered subject, the order of the maps was changed by one-step cyclic permutation (Kirk,1968,p.153) while the even numbered subjects saw the other maps from each pair. For example, subject 9 was shown (in order) 2A, 1B, 3A and 4A, so consequently subject 10 saw 2B, 1A, 3B and 4B. Thus the experiment would be balanced after every two subjects, with equal numbers having viewed each map in each pair. This was important as at the start it was not known how many people could be persuaded to participate in the experiment. Overall the design could be used without repetition for (4!x2) =48 subjects.

3.2.1 The Chosen Routes

The routes chosen in this experiment are shown in figures 3.2 to 3.5. There is however no single obvious method for determining the significance of the differences between the routes chosen from the different maps. It is possible to look at the whole network of routes chosen from both maps for each journey, and treat each link within it as a separate category, within which the count of the number of people using it on map A is compared with the equivalent count for map B in a chi-squared test. The summed chi-squared figure provides an overall measure of the difference between the routes. Statistical problems were caused, however, by the fact that several links were chosen by only one or two subjects, leading to unacceptably low expected values which necessitated the merging of categories.

In terms of the map design, the use of different road classes is more important than the use of individual links, so an alternative method is to count the number of links of each road class used on either map, and conduct a chi-squared test. However, this still takes no account of the length of each link (which varied from 0.2 to 42.9 Km). Thus the most sophisticated approach was to add up for each subject the length of route used on each class of road, and conduct a two-way analysis of variance using road class and map type as the explanatory variables, so that the class/map interaction term was the relevant measure. Even this method has drawbacks, however, in cases such as map 2, where the route involved a basic choice between two mutually exclusive options: it was either mostly on a primary route with no 'other A' roads, or vice versa. The high variances caused by this all-or-nothing type of situation in the route length figures for these classes made it very difficult for the interaction to be statistically significant.

It was clearly possible to measure the overall distance for each route (link lengths were measured off OS 1:50,000 maps), and for pairs 3 and 4 it was also possible to estimate the time it would take to drive the routes. This was because in these areas a D.Tp database known as the Present Year Network File (PYNF) is sufficiently detailed. PYNF includes details of likely travel speeds (24-hour averages) for the links it covers, and these were in fact used as the basis for the compilation of the speed map 3B. There are however inaccuracies in the data, but the figures used here have been corrected for obvious errors.

Journey 1: Swansea to Carmarthen

Map 1A: Ordnance Survey Motoring Atlas, 1:190,080
Map 1B: Map Productions RAC South Wales, 1:190,080
Direction of journey: NW

The line symbols for the main road classes on the OS Atlas contrast much less with each other than those on the RAC sheet. M, A and B roads are depicted with a uniform filling width, and the colour contrast between magenta A roads and brown B roads is particularly unclear. On this particular page of the test atlas the brown B4306 appears to be very dark and perhaps more prominent than the magenta A48, which is marked as a trunk road only by the (T) suffix to the road number. B4306 is more direct than the A48, but has 5 marked gradients. Thus the basic idea behind this particular choice of journey was to see how much this apparent reversal of prominence would affect a route choice that seems on the RAC map, with its more hierarchical line symbols (the motorway and the green/yellow primary route (A48) being particularly prominent), to be relatively straightforward, especially once Swansea is left behind.

Overall (see table 3.1 overleaf) the effect of map design on the distance covered on each road class was highly statistically significant. The main differences (figure 3.2) were that on 1A, 3 people chose the B4306 and 4 more considered it, while no user of 1B even mentioned taking it into account. Those who chose it were attracted by its straightness (3 mentions) and prominence (2), and two of them did not even notice the A48. Those who considered the B4306 but decided against taking it were

Table 3.1. Aggregate Distances (Km) Covered by Chosen Routes, by Road Class

	MAP 1A	%	MAP 1B	%
Motorway	139.2	21.3	180.0	26.5
Trunk/primary	306.8	47.0	390.7	57.5
Other A dual	16.5	2.5	9.9	1.5
Other A single	123.4	18.9	92.3	13.6
B	67.0	10.3	6.1	0.9
Total	652.9		679.0	
Mean	40.8		42.4	

Between-map differences in:
Link usage by road class (chi^2) 7.83, 4 df, n.s.
Distance covered by road class $F(4,150) = 4.69$, $p<0.005$

put off by the gradients (4) and discovering its administrative class from the road number (3). The other major difference between the maps was the diversity of routes leaving Swansea on 1B, apparently due to an eagerness to access the more prominent motorway as quickly as possible. The only subject who failed to use the motorway thought it was a river, and was attracted to the A4070/B4296, incorrectly shown as a primary route. On each map, half the subjects who considered accessing the motorway at junction 46 failed to understand the respective restricted access junction symbols- a white disc (1A) and a red disc with arrows (1B).

The clear hierarchical break on 1B between the symbols for Primary and 'Other A' roads clearly focussed subjects' attention and routes into the top two classes, and only four mentioned other routes that they had considered.

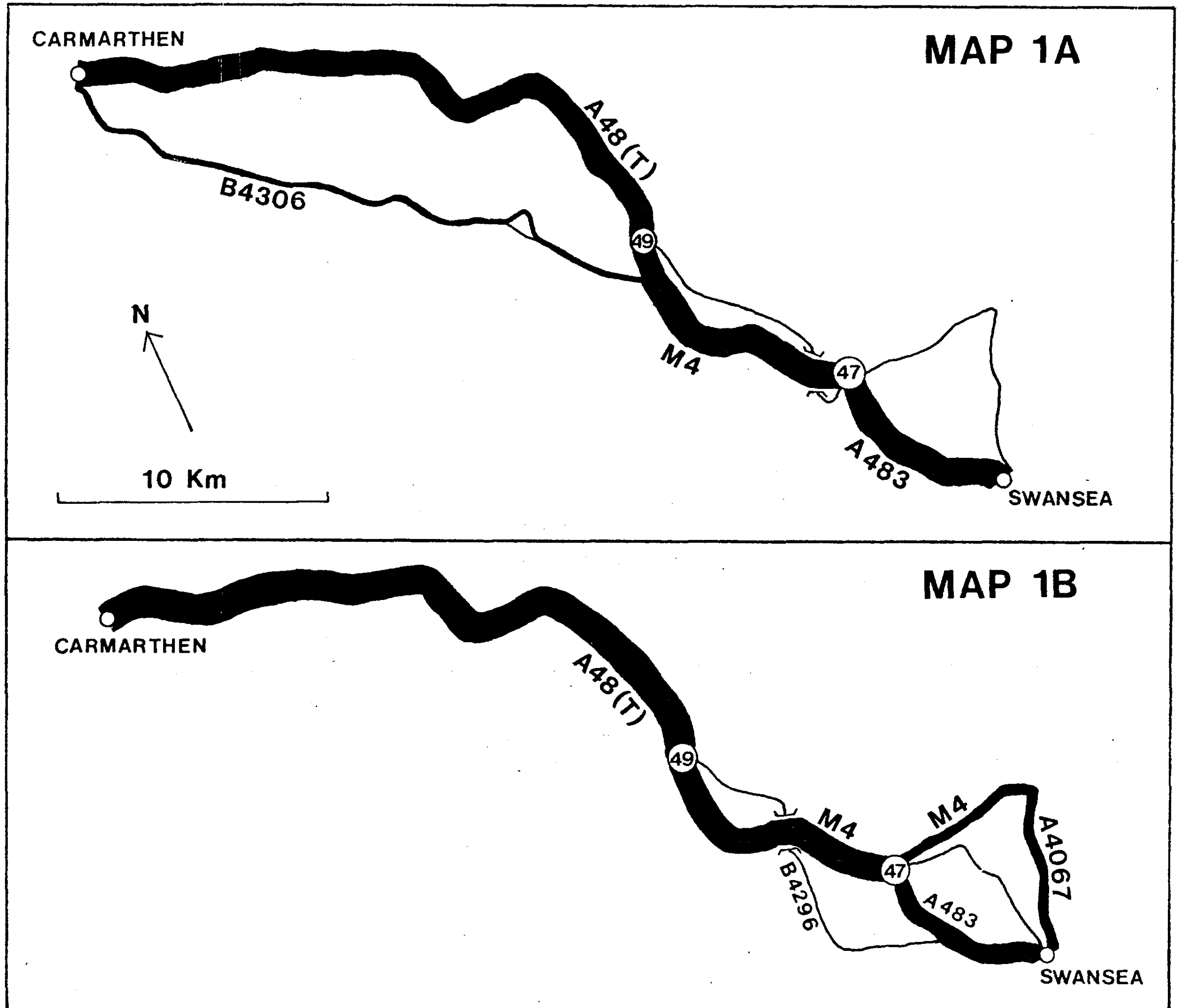


Figure 3.1

CHOSEN ROUTES

SWANSEA TO CARMARTHEN

1mm line width = 4 choices

16 choices per map

However, the prominence reversal on 1A disturbed a significant minority of users, although about half of these eventually relied upon the D.Tp classification, and 13 people mentioned considered alternatives, compared to 4 on 1B.

<u>Journey 2:</u> Llangoedmor (SN 2046) to near Bronant (SN 6567), Dyfed

Map 2A: Geographers' A-Z Road Atlas, 1:200,000
Map 2B: Shell Road Atlas, 1:200,000
Direction of route: NE

This journey was set in a more rural environment, with opportunities for short cuts and the use of 'unclassified' roads. The perception of the road classification is enhanced here by a pyramidal hierarchy: in the 'corridor of possibilities' there is only one primary route, but there are 5 'other A' roads, 15 B roads and numerous 'unclassified' ones. Consequently, even without the cues of symbol hierarchy and continuity, an individual line is clearly less conspicuous the lower its class. The differences between the symbol series on the two maps are that on 2A the primary/A/B distinction is made entirely by filling colour, whereas on 2B the primary route symbol is wider and the A/B/unclassified distinction is made by colour alone. While the colour scheme on 2B (pink, orange, yellow, white, on a white background) is clearly related to that used on OS 1:50,000 maps, on 2A (against a layered background) the primary route is green and 'other A' roads are in a more chromatic red.

Table 3.2. Aggregate Distances (Km) Covered by Chosen Routes, by Road Class

	MAP 2A	%	MAP 2B	%
Primary	288.8	27.7	446.3	44.1
Other A	512.4	49.2	310.9	30.7
B	204.7	19.7	174.9	17.3
Unclassified	11.2*	1.1	53.7*	5.3
Total	1041.1		1011.2	
Mean	65.1		63.2	

Between-map differences in:

Link usage by road class (chi^2) 9.47, 3 df, $p<0.05$

Distance covered by road class $F(3,120) = 2.34$, n.s.

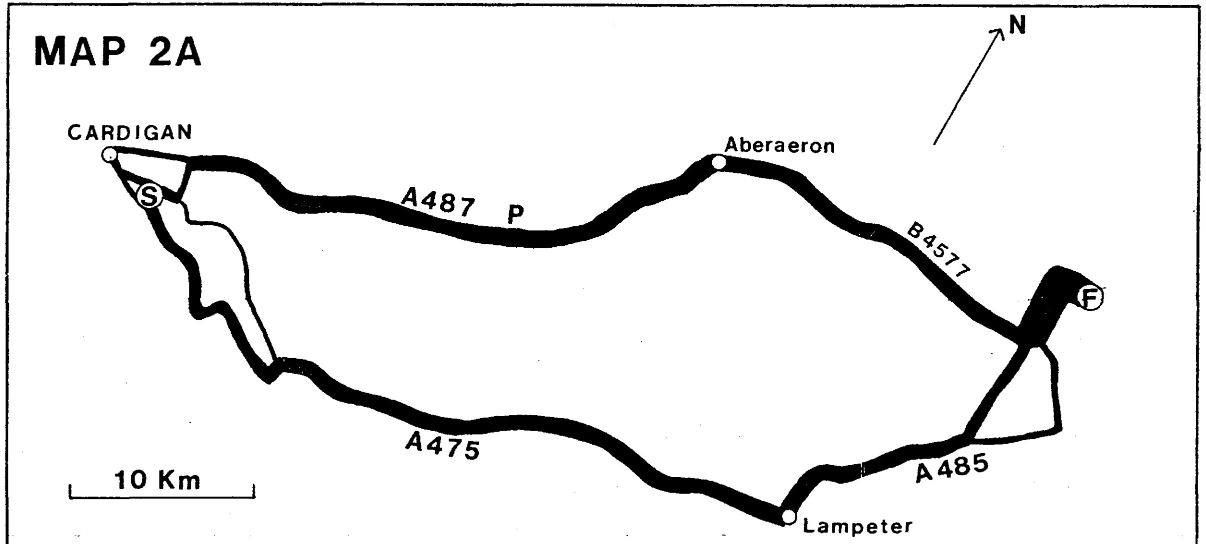
* Excluding short sections at either end of the journey where no choice of road class is available.

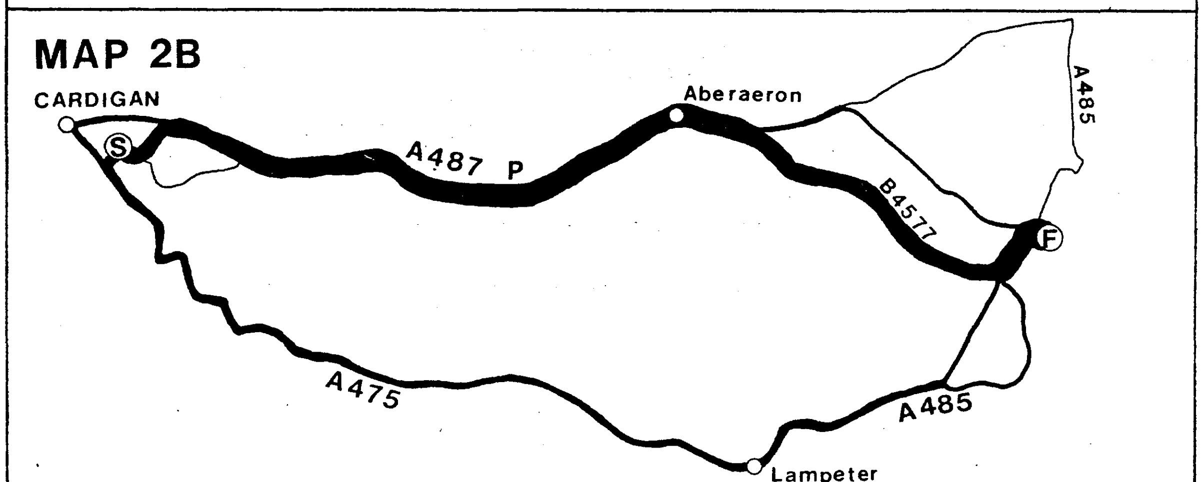
While the design-related differences are less significant than on journey 1, this is partially due to the aforementioned statistical factors, and it can be clearly seen from the above table that in relative terms, more use was made of the two extreme classes (and less of the middle ones) on 2B. The basic choice to be made on this journey was whether to follow the coastal primary route (A487) and cut inland on B or unclassified roads, or to take the longer inland route over fairly hilly and bendy 'other' A roads (A475/A485). The A487 was relatively less favoured on 2A, where of the 7 subjects who mentioned the relative classification of the two roads, the 4 who considered the A487 (correctly) to be more highly classified had all consulted the key, while the 3 who were of the opposite opinion relied on their instinct and/or knowledge about the superiority of red.

Figure 3.2

CHOSEN ROUTES

LLANGOEDMOR TO BRONANT





1mm line width = 4 choices

16 choices per map

P Primary Route
S Start
F Finish

Despite the bold A487 road number box, several people were clearly attracted inland by the red roads, so that overall the sample was equally divided between the two routes.

However on 2B, where because of the lack of road number class prefixes the graphical hierarchy might be expected to be more influential, the wider, reddish A487 was clearly seen to be the major road (10 specific mentions, compared to none for the A475/A485), and the only reservation users had about it was the necessity to cut across on the apparently lowly B4577 (9 mentions). Some people consequently considered the alternatives of going 'round the top' on A roads (A487/A485) or cutting across on fairly direct unclassified roads. Clearly unclassified roads were relatively favoured on 2B (e.g. at the start), whereas B roads (B4570/B4578) were used more as short cuts on 2A, where their relative boldness clearly carried connotations of quality.

Overall it can be seen that even in a highly structured network conducive to easy classification decoding, the graphical representation of the roads has an effect on route choice. It would seem that on each map the most significant perceptual breakpoint was provided by the change in width (between B and unclassified on 2A, and Primary and 'other A' on 2B), but that the attraction of red was also considerable. Neither scheme provided a smoothly-perceived hierarchy.

Journey 3: Sevenoaks to Bognor

Map 3A: Ordnance Survey Routeplanner, 1:625,000
Map 3B: Road Speed Map of South-East England, 1:625,000
Direction of journey: SW

This journey is the most complex of the four, as nothing like a straight line or continuous route is and considerable 'dog-legging' is required. In fact the shortest route chosen was nearly 27% further than the straight line distance. There is an interesting contrast in the map symbolisation, as 3A's derives mainly from expediency (the use of process printing) while 3B's speed classification was specifically designed with a graphical hierarchy to emphasise the fast and (to a lesser extent) slow roads over those of medium speed, so that the former might be selected and the latter avoided. There are large areas of graphical monotony on both maps. On 3A this is because all A roads are depicted in magenta, and the difference between single and dual carriageways and between primary and non-primary routes is none too clear. The magenta roads tend to dominate the map image, especially as roads of lower class are particularly inconspicuous. On 3B, the relative uniformity of expectable speeds on single carriageway cross-country roads creates substantial areas where almost all the roads are of the same class. Although the classification criteria on the two maps are different (so that they obviously cannot be compared in the same way as the other map pairs), there is considerable correspondence between roads classified as fast and slow on 3B and respectively motorway/ dual carriageway and roads within towns on 3A.

Table 3.3. Overall Measures of Journey Length by Map Type

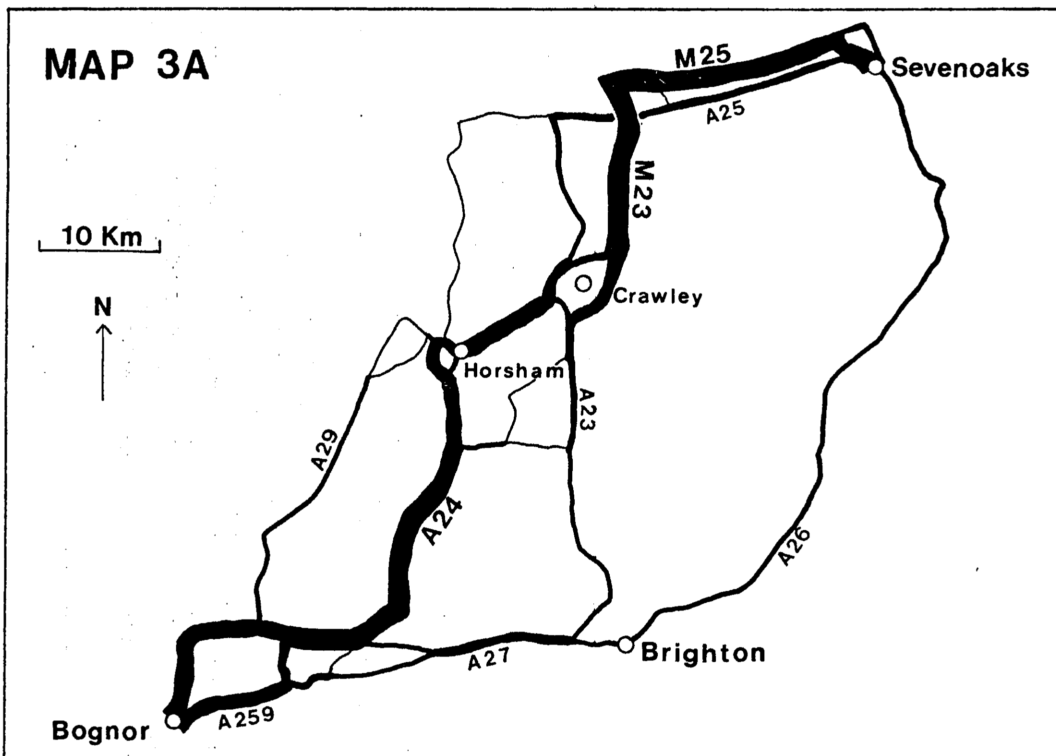
	MAP 3A	MAP 3B	Difference *
Aggregate distance (Km)	1829.3	1803.0	
Mean distance (Km)	114.3	112.7	t= 0.94, n.s.
Mean time (hrs)	1.52	1.41	t= 2.72, p<0.01
Correlation between route length and duration	-0.253	+0.403	

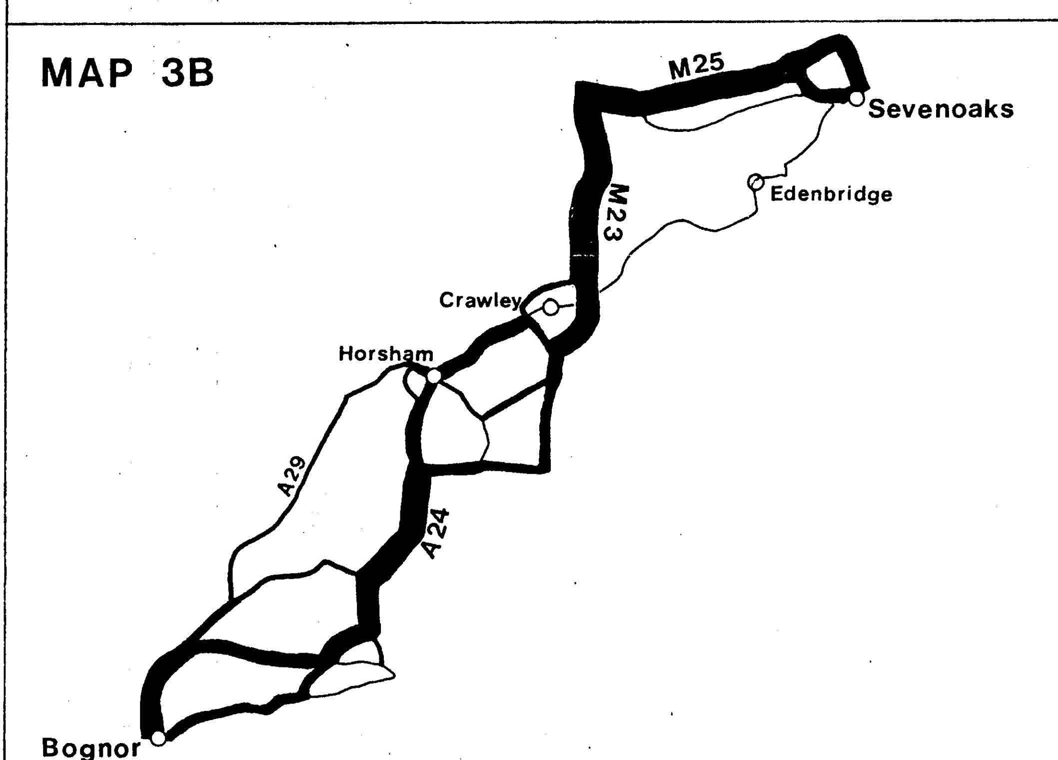
* Two-sample t-test

Figure 3.3
CHOSEN ROUTES

SEVENOAKS TO BOGNOR

1mm line width = 4 choices





Map 3B yielded significantly quicker routes, as would be expected given that it was compiled from the data used to assess it. It was successful in concentrating the routes onto roads of the top speed class (red), which accounted for 48.7% of the total route distance, while the slowest class (grey) was avoided completely (although a few links in Brighton and Worthing are the only relevant ones thus classified). However, this speed was achieved without going particularly out of the way to find fast roads, as overall the routes were no longer than those chosen from 3A. The difference between the coefficients for the correlation between route distance and time is in fact statistically significant ($z=1.47$, $p<0.05$) indicating a slight tendency for the more direct of the routes chosen from 3A to be slow, while 3B focussed the routes more tightly around the fast, direct M25/M23/A24 corridor.

On 3A, the lack of prominence of the motorway symbol presented a problem. Again one subject, who was used to maps with blue motorways, interpreted it as a river, and two did not even see it because of its weak contrast against the background whilst being very different from the dominating magenta network. Nobody even considered using the subdued B roads. Two subjects who visualised the route as being in a particular cardinal direction (west or south) chose poor routes as they were not attracted into the fast NE-SW corridor. All four people who considered or used the A29 also suffered from two short links being almost completely obscured by road number boxes. This was also the only map in the study which subjects (3 of them) complained was difficult to use.

On 3B, the relative graphical importance of roads was the major determinant of the route for all of its users, except for one who looked for a direct route to minimise dog-legging. This included four people who normally relied upon the D.Tp class letters. Although they all used the key, there is no evidence that the administrative classification based on casing thickness had any influence on the routes. Users appeared to be more concerned about going for the top speed classes (12 mentions) rather than avoiding the lower ones (0 mentions): town names were more often used to determine places to avoid, especially where there are few links entirely within the town, so that its retarding effect was spread graphically across semi-rural road sections. The lightness reversal in the symbol series (the yellow in the middle class being the lightest colour) also caused perceptual problems, and two people interpreted the greyish yellow class to be faster than the yellow. Complaints were also made about the lack of distinction between the red and brown fillings (similar to the problem on map 1A) and between the yellow and the greyish yellow.

Overall it would seem that many of the problems experienced with 3A were due to graphical monotony amongst relatively prominent roads, whereas on 3B localised monotony of more graphically subdued roads (e.g. around Edenbridge/ East Grinstead) helped to focus the routes onto the roads that are differently classified. This tends to be a property of speed maps, as in many rural areas the majority of marked roads tend to be of medium speed.

Journey 4: Buxton to Melton Mowbray

Map 4A: AA Touring Map of Great Britain, 1:633,600
Map 4B: Map Productions England and Wales, 1:633,600
Direction of journey: SE

This journey is in an area with a dense road network which includes two major cities (Derby and Nottingham), and requires a difficult assessment of whether the use of the M1 motorway, which runs somewhat transverse to the line of the journey but bisects the cities, is beneficial. The AA map is almost identical to that found in the much-used AA Members Handbook, as tested by Carpenter (1979), with a blue-green-pink colour scheme, and dual carriageways depicted by additional width and a thin centre line. The Map Productions specification is particularly varied with respect to casings, and apart from the uncased blue motorway, the remaining classes are of some configuration involving red or red and yellow, and carry road numbers without class prefixes.

Table 3.4. Aggregate Distances (Km) Covered by Chosen Routes, by Road Class

	MAP 4A	%	MAP 4B	%
Motorway	272.3	15.2	61.3	3.6
Primary dual	270.4	15.1	125.1	7.3
Primary single	904.5	50.6	1018.1	59.8
Other A dual	20.0	1.1	50.0	2.9
Other A single	319.1	17.9	449.2	26.4
Total	1786.3		1703.7	
Mean	111.6 Km		106.5 Km	
Mean time	1.62 hrs		1.56 hrs	

Between-map differences in:

Link usage by road class (chi^2)	19.69, 4 df, $p<0.001$
Distance covered by road class	$F(4,150)= 4.63$, $p<0.005$
Mean time (two-sample t-test)	$t = 2.51$, $p<0.01$

The effect of map design was clearly highly significant both in terms of the actual routes taken ($p<0.001$) and their quality, those chosen from 4B being significantly shorter and quicker. The best 'solutions'

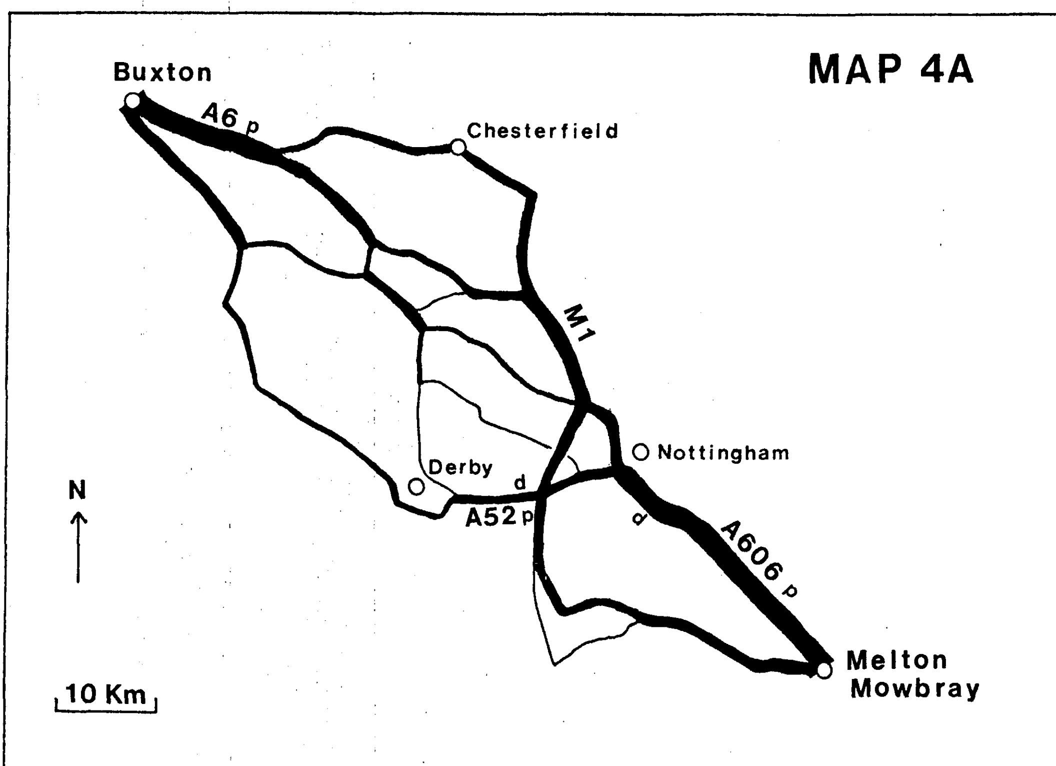
Figure 3.4

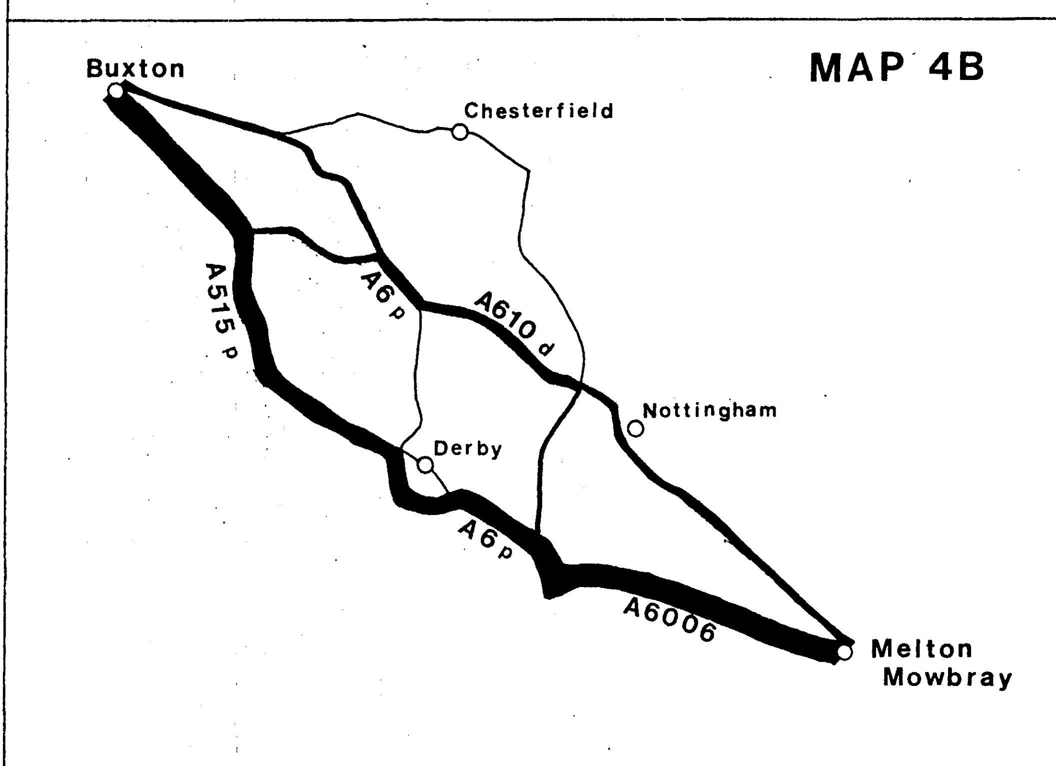
CHOSEN ROUTES

BUXTON TO MELTON MOWBRAY

1mm line width = 4 choices

p Primary d Dual Carriageway





were revealed by the corrected PYNF data, in that three basic routes are clearly faster than the alternatives. The main elements shared by these routes are that

1) if the motorway is used, no great detour is involved in accessing it,

2) at most one of the two major city ring roads is used, and

3) travel on the fairly slow, winding A6 (and certain similar tributaries) is minimised.

The main difference in performance between the maps was due to 10 of the 16 users of the AA map falling foul of these principles. Clearly there was more of a tendency on 4A to divert considerably to use the relatively prominent motorway, and several people had difficulty in assessing the relative classification of the green primary routes and the pink 'other A' roads.

The most stunning specific difference between the route sets is between Derby and Melton Mowbray, where all 4 of the viewers of 4A who had come via Derby used the A52/A606, while on 4B the equivalent 10 subjects all chose the A6/A6006. It would seem that the wide, 'triple-line' dual carriageway symbol for the A52 on 4A was more conspicuous than the thicker red casing used on 4B, and there is also a slight gap in the filling colour on 4B on the A52 side of the divergence, but perhaps the clearest design difference is in the clarity of the built-up area symbol, which faded into the cream background on 4A, but on 4B was darker than on any of the other tested maps. In fact 6 users of 4B specifically mentioned the need to avoid Nottingham and/or Derby compared to only 1 on 4A.

The success of the Map Productions map is interesting because of its unusual specification (half the sample needed to use the key) which relies on contrasts in the character of the line symbol as much as changes in colour or width. In this case it seemed to lead to a clearer visualisation of the more direct routes, focussing the attention mainly onto the primary routes, which accounted for 67% of the aggregate distance. The majority of the remainder was covered on the A6006, which forms the final link in what is otherwise the most direct primary routing, and was selected by over half the subjects. The less prominent motorway was not a significant cause of deflection. Interestingly, two people said they had missed seeing either the motorway or the more direct 'other A' route (A610) because of concentrating on the yellow primary routes. This backs up the aforementioned work of Kinchla et al.(1983) suggesting concentration on one perceptual grouping at the cost of lesser utilisation of information in other groupings. This is clearly possible in route choice in cases where the journey can be accomplished predominantly on one perceptually continuous road.

3.2.2 Overall Considerations

From the comments made by the subjects, it was clear that nearly three-quarters of the sample (23 subjects) misread a map at some stage during the experiment without resolving the error by subsequent use of the key or of non-graphical information such as road numbers. In total, 30 cases of 'graphical communication failure' were diagnosed, but no subjects made more than two errors. 12 (40%) of these were 'low-level' processing errors, where roads that would have been considered were simply not seen by the subject. Nine of them were caused by inconspicuous symbology, while in the other three instances the line symbol was obscured, notably by the road number boxes on the OS Routeplanner (3A).

The remaining 18 cases (60% of the total) were 'high-level' communication failures due to incorrect interpretation of the map symbols. Half of these were misinterpretations of the difference between the symbols for single and dual carriageways, including four people who made this mistake even after consultation of the key. These were on maps where the single/dual distinction is made by screening the filling for single carriageways (1A and 2B) or by adding a thin centre line for dual roads (1B). Clearly imprecision of colour memory between viewing the key and the map may cause interpretational problems. Other high-level errors were perceiving the relative classification of the roads incorrectly (5 cases), and considering blue motorways to be rivers (2) and bridges where one road crosses over another to be junctions (2). It is perhaps not surprising, in view of the subjects' comments and its design with little apparent concern for perception, that the map on which the largest number of errors was made is 3A. It would certainly seem that a map designed with consideration to the relative conspicuity and prominence of the road classes is liable to be misread less frequently.

Subjects' ability to judge distance generally appeared to be fairly accurate, in that those people who mentioned taking the shorter of alternative routes generally succeeded in doing so. However, there was an exception to this on journey 2, where two people selected 'shorter' (inland) routes that were in fact 13.5% longer than their stated (coastal) alternative. A possible explanation for this is that the former route can be undertaken almost entirely on a road of continuous class, albeit in an arc, while the psychological distance of the coastal route is increased by the need to use relatively minor roads to cross the gaps to and from the A487. If so, this represents an added perceptual advantage for a continuous classification.

The number of alternative routes considered clearly varied both between subjects and between maps, with more alternatives being considered on maps with a less strongly pronounced hierarchy of roads such as 1A. The greatest number of basic alternatives mentioned by any subject on any map was 3 by a user of 1A. On the more complex journeys, people were less likely to consider alternatives which were completely different from start to finish. On any of the routes mentioned, the point furthest away from the straight line route was at a visual eccentricity of 14° from a fixation on the straight line (at a viewing distance of 30cm).

From the taped protocols an overall picture was developed of the cognitive strategies used by subjects in order to find the minimum time route from the map (a criterion which was relatively unusual for 11 of the subjects). The strategies were summarised by the experimenter into their basic elements (a maximum of 5 per subject was required), which as mentioned above were checked with the subject to ensure that they corresponded with his own assessment of the situation. The single most significant element in each subject's strategy is tabulated below.

Use of graphically prominent roads	11
Use of M and A roads only	6
Detour for motorways/ most dual carriageways	5
Use of prominent roads amongst M and A only	4
Avoid towns	3
Directness more important than class	1
Use of A roads exclusively	1
Selection of the easiest route to follow	1
	--
TOTAL	32

Clearly there is a balance between reliance upon the D.Tp road class and the prominence of the road symbol, and although no conclusions can be drawn about the overall habits of the map using public because of the specificity of the sample, it is interesting that the largest single group relied on prominence. Adding in all the subsidiary elements to people's strategies, and counting the total number of mentions, a more detailed picture emerges.

D.Tp class/ road type:	
Use of motorways	17
-shortest routes to motorways	5
Use of dual carriageways	8
Use of trunk/primary rather than 'other A'	4
Use of M and A roads only	16
-except obvious short cuts	3
Use of M, A and B roads	1
Use of A roads exclusively	1
Use of unclassified roads where necessary	1
Graphical significance on the map:	
Use of prominent roads	14
-where an overall route clearly stands out	2
-within A roads	2
-red roads in particular	2
-where road numbers lack class prefixes	1
Directness more important than class:	
Within M, A and B roads	1
Within A roads	3
Directness and class carefully balanced	2
Other considerations:	
Avoid towns/ built-up areas	13
-larger towns only	8
Use of an easy-to-follow route	11
Avoid hilly roads	9
Avoid bendy roads	9
Avoid roads likely to carry heavy traffic	3
Avoid ring roads	1
Minimise the number of road junctions	2
Use roads with small numbers (e.g. A6)	2

Considerations such as avoiding towns, selecting a route which is easy to follow and avoiding hilly and bendy roads are clearly significant secondary elements to many people's strategies, indicating the importance of the representation of information such as built-up areas and gradients. A check was also made to see whether drivers and non-driving navigators differed in their strategies, but no obvious differences were found.

Overall several conclusions can be derived from this pilot study. Firstly, the conspicuity and prominence of line symbols have an important influence on route choice. Even where the D.Tp classification is labelled, prominence reversals in a line series (as in 1A) cause a significant disturbance to a proportion of map users. Discontinuities in a graphical hierarchy are also reflected in the chosen routes. Width appears to be a major determiner of prominence, as is darkness, such that a reversal of lightness order in a series again causes problems for some users. Red also appears to be a particularly attention-getting colour which connotes importance, while the prominent representation of built-up areas clearly aids the selection of a fast route. Continuity of road classification is also an important influence on route choice. Finally, graphical monotony amongst prominent road symbols and lack of distinction between road classes does not appear to encourage good route choice, which is encouraged when the range and smoothness of the graphical hierarchy reflects the relative differences in road quality.

4. THE CODING OF LINE SYMBOLS

Of the influences on map-based route choices over which the cartographer has control, the most important is the design of the series of line symbols to represent the road classification. For the series to be clearly understood without reference to the key, two sorts of visual organisation are required:

1) The symbols used for each class should be clearly distinguishable from each other and from other lines on the map such as those representing rivers and boundaries. These are qualitative differences as they represent distinctions with no hierarchical implications.

2) The order of the classes should be reflected in the relative saliency of the line symbols used. In Wrigley's (1985) terminology, road classification is an 'ordered polychotomous' variable- a hierarchy of discrete levels which do not directly represent numerical quantities. However, where it is necessary to distinguish them from qualitative differences, variations in the saliency of class symbols will be referred to as 'quantitative' as they do have hierarchical connotations.

4.1 Constraints of the Road Map Context

Before consideration of the graphical means by which these aims might be realised, some points on the perceptual nature of road map displays should be made. Certain processes in the perception of complex displays, such as the emergence of 'figures', operate in a manner peculiar to each individual situation. The perception of colour in particular is beset by countless contextual complexities (see Chapter 5). That this does not lead to an anarchic infinity of possible interpretations is because many of the potential variables are eliminated or constrained by the very nature of road maps and the manner and conditions in which they are used for route planning.

Notably,

1) Road map images are two-dimensional and are generally viewed orthogonally.

2) They are generally entirely static.

3) The 'corridor of possibilities' is generally narrow enough to involve only foveal and near-peripheral vision. (At a visual eccentricity of 14° , the most peripheral road considered in the pilot study is just within the range where, according to Kinney (1979), colour coding is effective for discrimination.)

4) Printed road maps are usually folded inwards or enclosed within atlases and are therefore not generally exposed to ambient illumination. Consequently ink colours do not fade significantly.

5) The number of ordered road classes required is generally about 5 or 6.

6) The map composition as a whole tends to be dominated by small areas of dark, saturated colour (lines) on a relatively light unsaturated background (or for negative contrast self-luminous displays, light, saturated colour on a dark, unsaturated background). While background colours vary to a certain extent across road maps, they are always relatively recessed. Thus

(i) the road network clearly emerges as figure given its continuity and the inevitably small ratio between its size (area) and that of its background. The visual organisation is much more stable than on a colour choropleth map or a painting with variable size ratios between areas of colour, where for example a small, light and highly colourful yellow patch amid dark, unsaturated browns and greys would stand out far more than it would if set amongst saturated oranges and reds. Consequently

(ii) the background colour(s) have the strongest influence upon the average brightness of the stimulus, to which the eye adapts (although where the figure is self-luminous it will exert some control), and

(iii) the main contrasts influencing the perception of a coloured line are with its immediate background and with lines of other classes.

Despite these constraints, however, there are still problems caused by the variety of conditions under which route planning maps are used- for example from leisurely pre-journey planning in the home in daylight, tungsten or fluorescent lighting, to periodic sideways glances at the map on the passenger seat by the nighttime driver-navigator under low-luminance sodium lighting. In most cases, however, the part of the map under consideration is at least relatively evenly illuminated.

4.2 Coding Dimensions for Lines

The foundations in semiology (the language of sign systems) of graphical coding have been laid by Bertin (1981,1983a). Apart from the two (x,y) dimensions of planar space, he defines 6 visual variables for the graphical display of information, namely size, texture, colour, value (lightness), orientation and shape. Moreover he demonstrates that each of them can be used individually for the coding of lines, although orientation and shape can clearly only be varied for constituent elements of the line symbol (e.g. dots/dashes) and not the overall line itself. His use of the term 'colour' also requires clarification, as colour itself has three psychological dimensions (see chapter 5 for definitions). Here, as in much of the human factors literature, 'colour' effectively refers basically to differences in hue, while lightness (value) is treated separately and chroma is not considered.

Bertin defines three levels of perception for which these variables can be used:

1) quantitative perception, which can yield a numerical ratio between two symbols, and is possible only with size

2) ordered perception, enabling the ranking of

categories without reference to a key, which is achievable with size, value and (to a certain extent) texture

3) selective or differential perception, where similar marks can be grouped together without being individually scrutinised, whilst different ones are clearly distinguishable (qualitatively). This applies for all the variables except shape.

In summary the levels of perception for each variable are:

Size	quantitative, ordered, selective
Value	ordered, selective
Texture	(ordered), selective
Colour	selective
Orientation	selective
Shape	none

According to Bertin, size and value are 'the variables of the image' which effectively form a third (z) dimension through a 'reduction in the amount of white' (i.e. background) that is visible. Because of their 'variable visibility' they are also 'dissociative'- that is they dominate visually over any other variables with which they are combined, and reduce their range of selectivity (e.g. fewer distinctions of colour are possible with small symbols). This is broadly equivalent to being 'non-integral' in psychological terminology. All the other variables are of 'constant visibility' and associative. In other words, according to Joel Morrison (1984,p.48), 'the addition of a differentiating variable can serve to enhance the ordering', but the converse is not true.

Colour provides the next best selectivity after size and value, especially when high chroma colours are used (see figure 5.2), and thus is excellent for distinguishing several classes. (The relative ease of use of the colour and black and white (texture and shape coded) London

Underground maps is clear evidence of this.) Because of its associativity (or integrality) it can combine well with other variables, and be used 'redundantly'- i.e. as a back-up to another variable which already provides enough information for a feature to be classified. For example, class 1 and 2 roads with respective widths of 1.5 mm and 1 mm could also be coloured so that all the class 1 roads are red while the class 2 roads are brown.

Joel Morrison's own framework redefines Bertin's specifications with slightly more precision. He resolves the somewhat shaky status of texture by fixing thresholds in pattern resolution. If the texture is finer than 72 lpi (lines per inch) its perception becomes dominantly one of lightness and thus ordered, while if it is coarser than 43 lpi it becomes one of pattern and is consequently only differential. Morrison also acknowledges the tridimensionality of colour, realising that chroma is a relevant and ordered variable. Consequently his classification is

1) ordering variables: size, colour lightness, colour chroma

2) differential variables: pattern arrangement, orientation, shape, colour hue.

Bertin (1983b) also affirms that ordered information <u>must</u> be represented by visual order in the symbols, or else a misleading impression will be given. Cartography, he suggests, is a finite rigorous language with <u>one</u> visual order imposed by physiology, and to adopt any other convention 'is to state that 2 is equal to 5.....or it is to be blind.' (p.78) Unfortunately the apparent simplicity and universality of this statement is illusory, or else the reader would be relieved by the imminent termination of this thesis. As Salichtchev (1983) has pointed out, the cartographic application of Bertin's work is limited by his exclusion of semantics- the relationship between symbols and the features being mapped. Lines seriously reduce the usable range of the ordering variables, and feature-related connotations and conventions may therefore become very significant in their perception.

4.2.1 Road Symbol Representation

Cartographers tend to use the coding dimensions for lines in particular ways. Texture coding is generally avoided, because for patterns to be sufficiently coarse to be adequately discriminable, small areas such as lines may be too small even for one pattern repetition (Phillips and Noyes,1980). A common ploy is the addition of casing lines, which may also vary in thickness, or the use of two or more parallel lines, with or without a filling. Throughout this thesis these will be referred to, along with any other variations of form within the line symbol (e.g. dots, dashes), as changes of 'character'. As lightness depends on area for its effectiveness, larger thresholds are required for ordering lines than for areas (J.L.Morrison,1984), but according to Crawford (1971) it can still be successfully applied on black and white maps as a partial substitute for colour.

Colour is however generally favoured where economically possible, and along with size (width) is the most commonly used variable for coding road symbols on maps. Often colour is used only qualitatively, with width providing the ordering. Various studies have reviewed the representation of roads on atlases (e.g. J.L.Morrison,1984) and military maps (e.g.Langran,1984). Morrison found almost universally that all types of road were shown in red (to distinguish them from black railways and blue rivers) and ranked by width, often with the addition of a double-line symbol for the top category and a dashed line for the bottom. An extended version of this specification was recommended by Rado and Dudar (1971) as a standard for atlas maps, with a filled triple-line at the top and declining widths of double and then single red lines. On the Eastern European road maps, mostly at around 1:500,000 scale, reviewed by Stams (1964), despite complete non-standardisation of the symbols, no more than three colours were used in any scheme, and the ordering was performed by width, often with the help of some character changes. His two recommended schemes involve three colours (purple, red and yellow), with width and character (double lines, pecked lines) used to represent changes in road width and surface quality. The difference between the two schemes is that one is cased throughout while the other is uncased. Standard NATO series use at most two colours (from red, brown and black) in any scheme (Langran,1984). Langran also mentions a scheme for non-redundant multidimensional coding with colour representing road surface quality and line width depicting the number of lanes.

Other qualitative information may also be shown (e.g. roads under construction, toll roads). However, where there is a need for a greater number of ordered yet distinct classes, as for example with larger scale road maps in developed countries, colour takes on a more demanding role. On specialised road maps where the road network is the most important information shown, line symbols especially need colour for contrast and emphasis as in terms of size and shape they are relatively inconspicuous (Arnberger,1966; Imhof,1982). A fairly wide spread of hues is common. Colour associations and conventions may also relate to particular classes of road, such as blue for British motorways. Where the amount of quantitative variation available with width and character is inadequate, Arnberger (1966) suggests that the quantitative implications of colour be utilised.

4.3 Perceptual Organisation

If, as Haber (1981) suggested, visual information has inherent organisation, how is this achieved, and how do the map designer's 'tools' (the aforementioned graphical variables) contribute to it?

According to R.M.Taylor (1985), map designers have been generally far more interested in the overall structure and organisation of maps rather than the design of individual symbols. Following the ideas of the Gestalt psychologists, they have stressed the importance of 'composing the graphic elements so that a hierarchy of structures is perceived that correspond to the functional and intellectual importance of the information to the user.' (p.194) Recently, following the work of Neisser (1967), there has been a resurgence of interest amongst cognitive psychologists in the Gestalt principles of organisation, first elaborated in the 1920s. According to the principle of Prägnantz, perception tends towards the greatest simplicity, with the simplest structures which require the least processing effort being perceived first. Complexity is minimised by the grouping of elements in the display according to the laws of organisation. As stated by Coren and Girgus (1980) items tend to be grouped through:

Proximity	Close together features
Similarity	Perceptually similar features
Closure	Items forming a closed figure
Good continuation	Items continuing in a common direction with the fewest sharp deviations (cf. section 3).

The recent interest has derived from the fact that while acceptance of these principles is universal, it has relied upon phenomenology (research by demonstration), and the underlying mechanisms were not originally understood.

They have now been scrutinised to the extent that an overall picture is emerging. It appears that perceptual organisation is an automatic and instant process that is performed at a glance in parallel across the whole field of view by low spatial frequency mechanisms, and is not available to consciousness. As a result of this process the field is parsed into 'perceptual units' according to the Gestalt principles. This grouping affects all subsequent processing by guiding eye movements and directing the allocation of conscious attention at a higher spatial frequency. A particularly conspicuous target or group may emerge spontaneously at this stage by calling attention to itself. If not, selective, focal attention can then operate serially on these units, and in visual search is allocated first to the units, and then by a controlled scan to the individual items within the units. The unit boundaries can only be transcended by a conscious attentional effort (Hoffman,1980; Kahneman and Henik,1981; Treisman,1982; Bundesen and Pedersen,1983).

4.4 The Qualitative Use of Coding Dimensions

Line width, colour and character are used to show qualitative differences between roads (Langran,1984). In the determination of Gestalt similarity/ dissimilarity, what is the relative saliency to the processes of preattentive organisation of the boundaries formed by these dimensions (or 'features')? A measure of subjective distinction is provided by the 'just noticeable difference' (jnd)- the smallest amount of difference a subject can perceive 50% of the time in a series of trials. For reliable perception in a map-using situation, however, greater differences are required (Castner,1983). Consequently Castner and Robinson (1969) have defined a separate threshold, the 'least practicable difference' (lpd), as the difference observed and reported by a given percentage of map percipients. 75-100% is generally recommended (Robinson,1982) and 75% was used in the original study. The example application mentioned by Castner (1983) is the determination of the series of line widths to be used in the symbolisation of a highway network.

Following the work of R.D.Wright (1967) it is possible to estimate lpds for line width. Wright recognised the lack of cartographic research devoted to lines, and the mixed results obtained by 'trial and error' map specifications. Using same/different judgements he conducted a series of tests of width discrimination between solid lines, ranging from simple pairs of line segments to integrated networks. In the simple test, 88% of the variance in the responses was explained by two measures

1) difference in width divided by the width of the wider line (the D/WW ratio) and

2) the width of the wider line (smaller relative differences are necessary between pairs of wider lines).

In networks, three variables appeared to be significant:

1) the D/WW ratio

2) the number of width categories present, and

3) network pattern: line width discrimination was harder in rectilinear networks than in irregular ones.

Neither the width of the wider line nor the density of the network had a significant effect. From his graphs, on which these relationships were represented by multiple Spillman curves, lpds can be calculated. The D/WW thresholds for the respective proportions of people to make correct discriminations are tabulated overleaf.

Discrim. Level	Simple Pair	No. of categories (Irreg. Networks) 2	3	4
75%	0.17 (1)	0.22	0.27	0.45
90%	0.25 (2)	0.34	0.43	0.61
95%	0.32 (3)	0.42	0.56	>0.7

(1) or a difference of 0.015" (0.38 mm) if smaller
(2) or a difference of 0.025" (0.64 mm) if smaller
(3) or a difference of 0.03" (0.76 mm) if smaller

These results were confirmed by duplication with an entirely different sample of 200 subjects by Crawford (1971) for both black and grey (30% screen) lines. However, as no comparisons were made across the lightness difference, no information was provided about possible interactions between colour variables and perceived line widths. An interesting comparison is however provided by a study on type size discrimination by Shortridge (1979), who recommended a minimum distinction of 2 point sizes (about 0.5 mm letter height) or 25% (equivalent to a D/WW of 0.33).

Clearly then size discrimination deteriorates as the number of sizes seen together increases. Bertin (1983a) recommends that for good selectivity, only about 4 or 5 sizes should be used together, as greater numbers of codes begin to make absolute identifications difficult. The problem is particularly acute with lines on maps, whose size can only really be usefully varied in one dimension (width), such that Bertin advises that 3 steps of line thickness is really the maximum for good visual groupings to be formed. Potash's (1977) review agrees, recommending spacing on a logarithmic scale to minimise confusion, and Robinson et al.(1984) consider three to be the limit for lettering sizes of a single type style. Dobson (1983c) mentions two additional limitations of size cues:

1) memory of size appears to deteriorate rapidly,

causing potential difficulties when comparisons with a key are required.

2) size cues are easily confused in the visual periphery because of reduced visual acuity away from the fovea. Small features, especially those lacking in contrast, may not even be perceived at all (Kinney,1979).

While the range of contrast that can be introduced by character alone is relatively small, cased, uncased and even double-line symbols of the same colour and width have been used together on specialised road maps. In such instances casings also create distinctions in colour through interaction with the filling (see section 5.4). However, the quality of discrimination by such means has apparently never been tested. The situation with colour is quite the opposite, although on maps the use of colour coding has evidently not been seriously questioned, perhaps because of their complexity and the shortage of other coding options for lines and areas in particular. However, applied psychologists have conducted a wide range of studies of the effectiveness of colour coding in visual displays.

According to R.M.Taylor (1986), colour and texture differences are the major determinants of boundaries of perceptual units. Treisman (1982) considers colour to be a simple feature which is detected automatically at the preattentive stage, while Haber (1981) has suggested that it is the only dimension which enables immediate preattentive segregation. Other studies have confirmed the quality of grouping of same-coloured items (Bundesen and Pedersen,1983), the efficacy of segregation by colour in focussing attention on the target grouping (Harms and Bundesen,1983), and the difficulty of inhibiting the processing of irrelevant or distracting colour (Kahneman and Henik,1981; Luder,1982). Even when the target was not coded by colour, other elements of the same colour interfered with target identification (Harms and

Bundesen,1983). As Phillips (1984) has suggested, the advantages of colour seem to occur in the early stages of visual processing where big differences in performance can occur. In simple visual search tasks, near perfect visual separability was achieved with codes of up to five colours (Green and Anderson,1956; Smith,1962). The overall density of features in a display has far less effect on search times than the number of items sharing the target's colour, indicating that even in more complex fields the appropriate colour grouping is found very quickly (Cahill and Carter,1976).

Other tests of the speed and accuracy of task performance have compared colour to alternative coding dimensions such as size, shape and sometimes brightness, again making the use of the term 'colour' all the more ambiguous. Precise colour specifications are rarely given, and while it often appears that the specific colours used are all highly saturated and differentiated solely by hue, variations in brightness and saturation, which may considerably affect discriminability, may also be present (Christ,1975). In his comprehensive review of 42 relevant studies, Christ showed that colour coding almost universally improved the performance of search tasks compared to other coding dimensions. Colour-coded symbols were located and counted faster than those coded by size, shape, brightness or letters/ numbers. This again points to the superior perceptual segregation provided by colour.

L.G.Williams (1967a,1971) stressed the importance of extrafoveal discrimination in determining how well structures separated, and evaluated colour, size and shape codes in this respect. Using eye movement recordings of search tasks, he devised an index called relative fixation rate (RFR), a measure of apparent similarity obtained by dividing the number of fixations made on items different from the target by the number of fixations on objects sharing its defining characteristic (e.g.colour). He found that when the target was specified by colour, there was a very strong tendency to fixate objects of that colour (low RFR). Size was less effective, although the largest objects were most easily discriminated from the rest, and selectivity by shape was particularly poor: the maximum distinction obtainable by shape (triangle and circle, RFR=0.35) was equivalent to a size ratio of about 1.6, and about 7% of the hue range (cf.section 5.2). With redundant coding, the less potent dimension was virtually ignored, so that with a redundant application of colour and size, colour was used to perform the task. Thus it can be seen that colour is more discriminable in the near periphery than size or shape. A further advantage of colour is that the memory of a particular colour deteriorates less than that of size, orientation or shape (Christ,1975). It has also been found that the presence of colour helps in the identification of line symbols (Wong and Yacoumelos,1973).

However, Dobson (1984) has claimed that while cartographers are hampered by their ignorance of the human factors literature on colour coding, the cartographic application of these experiments is severely limited. Of the 42 studies reviewed by Christ, only 3 actually involved map or map-like stimuli. Most of the remainder used small, discrete geometric or alphanumeric point symbols against an undifferentiated, neutral background. This he claims is very far removed from the subtlety and variety of map symbols with their far more demanding use of graphical variables, and the problems of the varied contexts and simultaneous contrasts created by different spatial configurations. Hopkin and Taylor (1979) agree that it might therefore be difficult to relate these results to map contexts. As, Dobson says, much of the human factors work is 'stimulus dependent', such differences could significantly affect the results. Background colour, for example, which is generally ignored or treated as a distraction in such studies, may improve the figure-ground relationship on maps, although in one small study (Christner and Ray,1961) no significant differences were found in the performance of symbol locating and counting tasks between maps with pastel layer colours and others with various achromatic backgrounds.

Overall, colour is likely to be <u>more</u> effective in the complex map environment, where organisation is of paramount importance. Evidence for this in demanding map-using tasks is provided by Stringer (1973) and notably Garland et al.(1979), with a minimum-time journey planning task conducted on various designs of urban bus map. Despite a large number of colours (11), colour coding produced a significant improvement in performance over an uncoloured map, equivalent to a reduction in the amount of detail. A study by Forrest and Castner (1985) of point symbol design for tourist maps, also confirmed the value of colour as the primary sorting characteristic for groups of cartographic symbols, while a major cause of errors in Streeter and Vitlello's (1986) experiment was the difficulty in distinguishing between roads and other linear features (railways, rivers) on a black and white road map.

Where redundant coding is possible, given the usable range of the variables and the number of classes required, discrimination is likely to be further enhanced. In map-based experiments with the size and lightness contrast of proportional symbols, when the two dimensions were used redundantly, Dobson (1983c) found improvements in the speed and accuracy of identification and categorisation tasks and in extrafoveal discrimination. Redundancy clearly led to a 'strengthening of the graphic message.' Used alone, colour allows the greatest range of distinctions of all the variables used for the qualitative coding of lines. The discrimination of specific colours and its relationship to the dimensions of colour space is discussed in section 5.5.

4.5 Figure-Ground and Visual Levels

With serial attention, units and features must clearly be scrutinised in some order, and priorities must therefore be established. Figure-ground segregation is achieved by the differential processing of these units, such that in addition to being distinguished qualitatively, they can form different 'visual levels'. What determines which units are the ones that appear to have simple structures, get selected first and become the focus of attention, while others form the background?

Gestalt psychologists have traditionally been associated with 'global precedence', the idea that the whole is different from the sum of its parts, and is recognised quicker. With several levels of grouping, the most wholistic is processed quickest. Recent work has shown that this is dependent on the natural tendency of small stimuli to be more degraded than large ones, and that the interaction between local and global levels of form depends on the relative visibility and quality of form information at each level. Consequently any particular level with 'good form' might attract attention (Hoffman,1980; Grice et al.,1983). Factors affecting form quality include stimulus size, contrast, continuity of contour (edge definition) and retinal peripherality. There may well be an 'optimum size' above which a stimulus is less likely to be seen as figure (Kinchla et al.,1983). However level precedence is not entirely externally determined. Although the perception of hierarchical structure is difficult to monitor as it includes covert attention processes not revealed by eye movement, it is clear that it is possible to decide to attend to a particular level and process it faster, at the cost of slower utilisation of other levels (Kinchla et al.,1983).

Cues which imply a third (z) dimension also assist in the segregation of figure (features cued to be nearer the eye) from ground. According to Haber (1981,p.6), it is possible that the visual system 'attempts to interpret all stimulation as if it is reflected from a three-dimensional scene even if the information comes from a flat surface.' In experiments with the apparent vertical (z) organisation of sets of three overlapping triangles, R.M.Taylor (1986) showed the importance of closure and overlay (occlusion) in the creation of figures. Moreover when one solid triangle was superimposed upon another, figure-ground segregation was not improved by hidden-line removal (figure 4.1) as the advantages of a more stable figure were outweighed by the loss of a closed outline. Ordering was also improved by a continuous brightness progression, and reinforced by the strong boundaries and distinctions provided by solid colour.

The results of this experimental work involving stimuli of much greater simplicity than maps compare interestingly with cartographic experience. On a road map, the overall line image meets the basic requirements to be seen as figure through

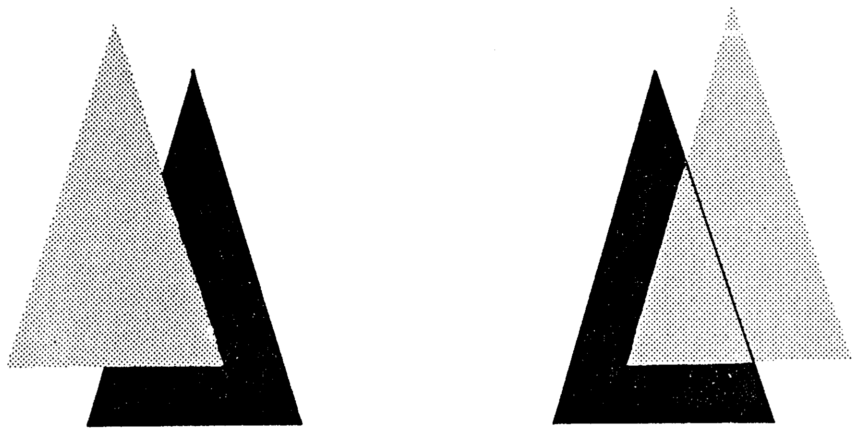
1) the relatively small proportion of the map area it covers,

2) its strong edge (brightness) contrasts, whether they are positive (as usual) or negative (Dent,1972; C.H.Wood,1976), and

3) its continuity.

This is further strengthened by the closed configuration possessed by almost every road network. The problem of the perception of road classes is therefore less one of figure-ground differentiation than one of the further segregation into ordered visual levels of a network of lines already perceived as figure. As Dent (1972) suggested, each map should have a range of visual significances to represent the relative intellectual

Figure 4.1 HIDDEN LINE REMOVAL



Hidden Line Removed Hidden Line Retained importance of the mapped elements, where 'each plane could provide for focussed attention and a good 'gestalt' be perceived.'

The means by which such a hierarchy of line classes can be achieved fall into two groups- those dependent on the spatial arrangement of road classes and those dependent upon their symbolisation. In the former category two factors are worthy of note:

1) closure. Certain individual road classes may form a linked network, like Primary Routes (and 'A' roads as a whole) in Great Britain.

2) continuity of contour. In general as mentioned above more highly classified roads are more continuous, and lower class roads appear to be their tributaries. This continuity can be strengthened by casings, especially where they are not broken at the junctions of side roads.

These factors are clearly circumscribed by the classification criteria used, so that the map designer may have no control over them. Their influence is greatly reduced where the class of a road can change from section to section.

Three relevant aspects of symbols have been noted which invoke depth cues to promote hierarchical organisation, and can be used redundantly for reinforcement (Saunders,1958; Bertin,1983a, Robinson et al.,1984; Lansdown,1985). These are, as they apply to lines, as follows:

1) superimposition (also referred to as occlusion or overlay). It is possible to decode a road classification without reference to the key as the higher class line symbols take precedence at cross-roads. (With cased symbols, the 'hidden lines' are generally removed.) The continued contour is perceived as covering the interrupted line (Ittleson and Kilpatrick,1951; Dent,1972).

2) progression of line weight. An association of size with depth was demonstrated by Ittelson and Kilpatrick

(1951), with the larger of two 'floating' objects appearing to be nearer the eye. Dent considers that this is not always entirely satisfactory without a variation in a more important depth cue, namely

3) intensity of contrast, both chromatic and achromatic, and particularly for edges. Edge contrast can be reduced by screening or increased by the addition of high contrast (usually black) casings, which can establish or strengthen contours, especially when brightness differences are weak (M.Wood,1968; Dent,1972). Wood (1972) has also noted that patterns with high internal contrast (e.g. dark casings and light fillings) appear to be nearer the eye. Variations are more difficult to achieve with negative contrast. Dent claims that edge contrast is successful as a depth cue because of 'aerial perspective', an association with everyday visual experience where, because of the further passage of the image through the air, distant objects appear to be fuzzy or (perhaps more correctly) faded, and are consequently less pronounced and less contrasting.

4.6 The Quantitative Use of Coding Dimensions

As Kaufman (1975) has stated, there is a distinct imbalance between the amounts known about

1) the visual processes involved in the discrimination and delineation of features, and

2) the factors which determine the relative saliency of stimuli, and the influence of saliency on their retention and subsequent processing.

The latter are by far the less well understood, and this is reflected in the rather disjointed nature of the evidence presented below. The conspicuity of a mark, its physical prominence in central vision and any quantitative connotations it carries all contribute to its saliency, but how are they related to the usable graphical variables?

Size is the dimension with the most obvious quantitative connotation. 'The larger of two lines of the same colour will appear more prominent, and without elaboration connote ordinal ranking (Robinson et al.,1984,p.312). The subjective estimation of stimulus magnitudes has been the domain of psychophysics for over a century since Fechner first suggested a general logarithmic relationship between physical and subjective intensity. Later work using different scaling techniques prompted Stevens (1957) to propose an alternative general law that equal stimulus ratios produce equal subjective ratios, the relationship being approximated by a power function

$$R=cS^n$$

where R= subjective intensity, S= stimulus intensity, c is a scaling constant (relating to the units of measurement), and n is the power exponent to be derived for each scale (e.g. loudness, brightness) by experimentation. This is referred to as Stevens' Power Law or the Psychophysical Law.

Stevens' work sparked off a rush of studies into the scaling of different phenomena including the apparent length of lines and area of 2-dimensional figures, but not line width. The resulting exponents for line length varied between 0.95 and 1.11, focussing around 1 (Stevens and Galanter,1957; Künnapas,1958; Ekman and Junge,1961; Stevens and Guirao,1963; Teghtsoonian,1965) indicating a simple 1:1 correspondence of physical and apparent length, although some judgments of large length differences were slightly underestimated (Clarke,1959). For the areas of regular shapes the exponents (reported by the above authors and R.L.Williams,1956) varied from 0.7 to 1.03. This variety may reflect the specific question asked. Teghtsoonian found that when subjects were asked to judge the <u>area</u> of circles, the exponent was close to 1 as they were able to estimate the linear dimension (diameter) accurately and square it, but when asked to estimate 'how big' they were the resulting exponent was 0.76.

However it is not possible from this to say how width affects the perceived visual weight of a line- whether it is judged as a single lineally-scaled dimension or whether the overall area of the symbol is estimated. In the construction of flow-line maps where a continuous variation in line width is used to represent statistical quantities, the conventional wisdom is to scale the symbol width in direct arithmetic proportion (Robinson et al.,1984). However, Wright's experiment (section 4.4) suggests the possibility of a logarithmic relationship, in that perceptually equal increments of width are basically predicted by the D/WW ratio, and therefore form a logarithmic series (nearer in shape to a power relationship with an exponent of <1).

In terms of conspicuity, size is clearly a major factor (Dobson,1983c). Woodworth and Schlosberg (1955) reported an experiment by Brandt in which doubling the size of the stimulus led to a 40-60% increase in its 'attention value'. In an experiment on the conspicuity of motorcyclists, Fulton et al.(1980) discovered a log-log relationship between the area of high-visibility fluorescent clothing worn (as seen by the subject) and the time taken to detect the presence of the motorcyclist. Clearly these findings relate to the conspicuity of individual specified items in an image, and not the more complex case of the relative conspicuity of many image elements seen together. Consequently the principle is likely to be more important than the specific statistical relationships mentioned. According to Kosslyn (1981), marks drawn with heavier lines tend to be noticed sooner, and also carry semantic connotations of 'more of' the symbolised information.

Variations of line character are conventionally used to symbolise stretches of dual carriageway, and as such carry hierarchical connotations. The addition of a thick black or red casing is common amongst road map producers, but other more mimetic symbols are also used, notably double line symbols (representing the two carriageways) and filled triple-lines, with the (usually narrower) centre line representing the central reservation. The influence of line character on visual importance has also been investigated in studies undertaken at the University of Wisconsin-Madison by Severud (1968) and K.Pearson (1971). However, most of the line types they/she? looked at were broken (e.g. dotted, dashed) forms conventionally not used to represent roads on specialised road maps. Factors such as the inked area, 'closure' (the perceived connectivity of symbol elements), complexity and compactness contribute to their visual weight. Solid lines have continuity as well as perfect closure and compactness, and consequently appear to be more prominent than broken lines of equivalent gauge (Pearson,1971). Their visual importance can be varied by width and brightness contrast.

Colour in general can have a big influence. In terms of promoting the conspicuity of features its superior performance in the near visual periphery is crucial. According to Easterby (1980), 'no single coding variable is capable of enhancing the conspicuity of specific elements in a display as much as colour can.' Saunders (1961) referred to the extreme conspicuity of a colour juxtaposed with achromatic items. Bertin (1983a) refers to the psychological attraction of colour and its ability to capture and hold the attention. This was confirmed in a study by Dooley and Harkins (1970) in which an apparently irrelevant graph, which was either functionally coloured, decoratively coloured or black and white, had been placed in a room where subjects were taking part in a separate and unrelated experiment. Post-testing revealed that colour, whether functional or decorative, significantly increased the amount of attention the subjects paid to the graph.

Colour, which has perhaps the greatest scope for manipulation of any of the map designer's tools (Robinson,1982), is however perceptually very complex, and in its cartographic and particularly quantitative application is imprecisely understood. A substantial literature on the use of colour design to organise colour displays does not exist (Dobson,1983b). When psychologists have explored colour space, their main concern has generally been to find large colour differences for the qualitative coding of information categories (e.g. R.M.Taylor,1977), for which colour is more often used. The perception of colour in lines and the selection from colour space of subjectively ranked series of colours merit special attention, and are discussed in chapters 5 and 6 below.

5. THE PERCEPTION OF COLOUR IN LINES

> 'The study of colour is generally confusing. The discrepancies found in physics, art and psychology offer no end of complication to the student who is conscientious enough to strive for a clear and exact understanding. Explanations and descriptions are difficult to reconcile.' (Birren,1930,p.271)

Since Birren wrote this a considerable amount of research has been conducted into the mysteries of the perception of colour, mainly within the fields of physiology, photography, physics, art and psychology, which has served to underline its complexity. Often the multiplicity of variables involved in the perception of colour is not fully understood or even recognised. For example the main factors influencing colour discrimination have been listed in rough rank order (Silverstein,1982 and Huddleston,1982) as follows:

- -wavelength separation (hue difference)
- -purity
- -brightness
- -item size
- -viewing time
- -adaptation state of the eye
- -number of colours present, and whether absolute colour identifications or just comparisons are required
- -brightness and colour of background objects and surroundings
- -foveal versus off-centre viewing
- -variations in individuals' visual systems (e.g. with age and colour vision anomalies) and experiences.

Presumably the ambient illumination characteristics are also of significance.

It is therefore not surprising that when the effects of one colour variable are being tested, the effects of the remainder are often inadequately controlled (Christ,1975). Moreover the necessity for consistent terminology for the comparison of studies both within and across disciplinary boundaries is apparently not realised. Colour clearly has three fundamental perceptual dimensions, referred to by Hunt (1979) as hue, brightness (the overall amount of luminous stimulation) and colourfulness (total chromatic stimulation). However a variety of terms are used for these and for related more specific concepts such as lightness and saturation (table 5.1). Sometimes slight distinctions in meaning are intended; in other cases the terms are just used loosely. The terms used in the present study are those above the dashed line in Table 5.1 and defined where ambiguous, with the addition of:

CHROMATICITY, for the plane of hue and colourfulness variations at constant brightness, and

COLOUR INTENSITY, for the plane of colourfulness and brightness variations at constant hue (Macioch,1983).

Additionally there is a vital distinction not always made in the literature between the perception of
1) (self-) luminous colours, seen in light sources and transparent films which emit or transmit light, and
2) surface colours, which are reflected off objects or surfaces such as conventional paper maps.

TABLE 5.1. COLOUR TERMS USED IN THE LITERATURE

		Dimensional Grouping	
Locus of Usage	HUE	BRIGHTNESS	COLOURFULNESS
All Colours:			
CIE System	Dominant λ	(Luminance)	Purity
Psychology	Hue	Brightness	Saturation
Surface Cols. only:			
CIE System		Reflectance	
Psychology		Lightness	Chroma
Munsell System	Hue	Value	Chroma
Conversation	Colour	Lightness	Brightness
Other terms:		Brilliance	Brilliance
		Intensity	Intensity
		Tint	Tint
		Luminosity	Density
		Albedo	Strength
		Visual effect	Vividness
		Tone	Richness
		Shade	Pallor
		Depth	Pastel/ archetypal

Definitions of terms as used in this study:

Lightness — The sensation of lightness/ darkness of a surface colour rated on a grey scale.

Luminosity — The capacity of radiant energy of a given wavelength to evoke a visual sensation of brightness. The luminosity of a colour (a combination of wavelengths) is thus its inherent brightness.

Saturation — The apparent chromatic concentration of a colour- the proportion of the total (chromatic plus achromatic) sensation that is chromatic. I.e. colourfulness relative to brightness.

Chroma The colourfulness of a surface colour relative to a neutral (grey) of equivalent lightness.

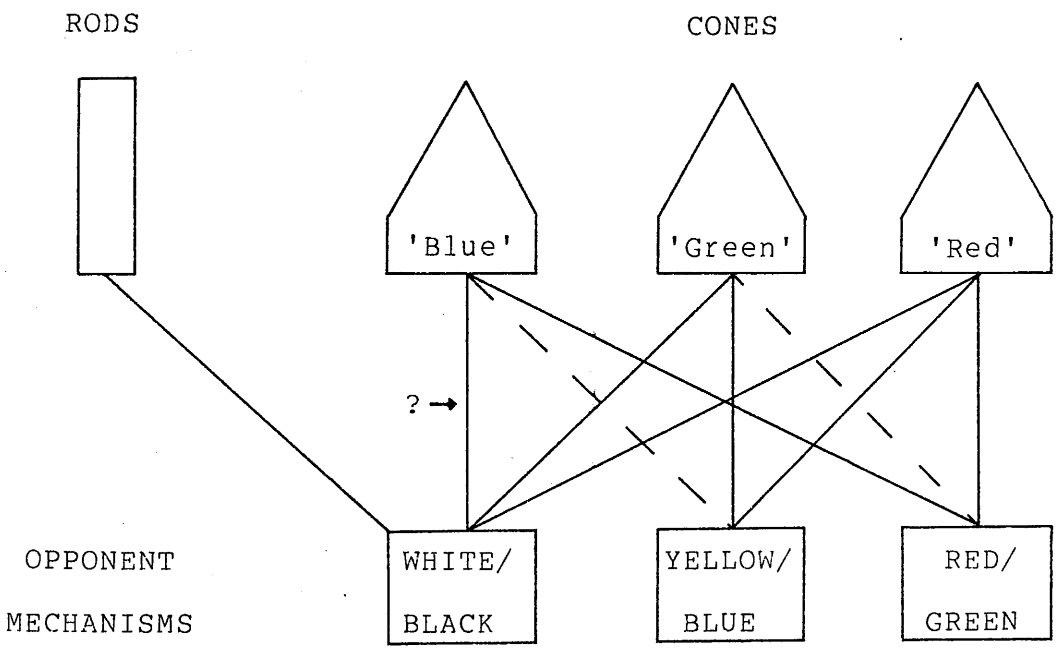
5.1 Colour Vision

There has been much argument among physiologists about the operation of the visual system with respect to colour. However in recent years a consensus view has developed on the basic principles involved, which are all that need concern us here. This consensus, known as zone theory (Wassermann,1979), differentiates a component zone containing the basic retinal photoreceptor cells, where responses to stimuli vary only in terms of intensity, from a central opponent zone, where the responses can change in polarity also. Photoreceptors come in two basic groups:

-rods, which are not spectrally selective, and are associated with the low intensity, low resolution, colourless (scotopic) vision that occurs at low illuminations as in twilight, and

-cones, which are spectrally selective and require somewhat higher levels of stimulation to fire producing full colour (photopic) vision. The cones are particularly concentrated into an area in the centre of the retina containing the fovea and about 2° of visual angle in extent, where colour vision has the highest sensitivity and acuity (J.J.Sheppard,1968). There are three types of cone which appear to respond maximally to wavelengths of around 440, 530 and 570nm, and are often referred to as blue, green and red cones respectively (figure 5.1). However, the differences between their spectral sensitivity curves (especially between 'green' and 'red') are quite slight, and have to be magnified in later processing (DeValois and DeValois,1975). 'Bumps' in these curves are unlikely to relate to any significant

Figure 5.1 COLOUR VISION: ZONE THEORY



Solid lines indicate connections of one polarity and dashed lines indicate connections of the opposite polarity. The core of this model is derived from Wasserman, 1979.

differences in the conspicuity of different wavelengths, as it would seem that their occurrence varies with conditions and between individuals.

The outputs from these receptors are relayed by optic nerves to the lateral geniculate nucleus (LGN) of the brain, where three sets of opponent mechanism are thought to be located. One system, utilising summed cone and rod information, transmits an achromatic (white to black) signal associated with brightness perception to the visual cortex. The other system involves two two-pole chromatic signals, the polarity of which changes with wavelength between 1) yellow and blue, and 2) red and green respectively. The red/green pairing, for example, is most active at 500 and 640nm, where the differences in sensitivity between the two sets of cones are greatest. These chromatic signals are associated with hue perception, while saturation is dependent on the relationship between the outputs of the two (chromatic and achromatic) systems. Thus the three perceived dimensions of colour space can be traced physiologically. The brightness signal is 'spatially opponent' and derives from <u>relative</u> intensity information (spatial brightness contrasts in the visual field) while the spectrally-opponent chromatic system signals the <u>absolute</u> intensity of colour. This, according to DeValois and DeValois (1975), is well adapted to a reflective world of light and shade whose sun has relatively invariant spectral characteristics.

However, the long-recognised phenomenon of colour constancy, whereby reflective objects appear to maintain their hue under illumination of differing spectral composition, has been investigated in long-term research by Edwin Land (Land,1977; BBC/Mollon,1985). He has suggested that the important factor in colour vision is the pattern of relative lightnesses across the visual field recorded by each of the cone systems, with each local signal scaled by the character of the illumination over its wider context. Recent neurophysiological work has shown how this might be possible, and thus modified and amplified the view of how this information is then represented to the visual cortex of the brain. Zeki (1977) first found evidence that brightness and colour information are stored in separate cortical 'maps'. Stimulated by Land's work, he has since discovered in the primary visual cortex (V1) double-opponent cells (i.e. both spatially and spectrally opponent, and thus capable of integrating colour information) which he suggests sort the messages received and forward them to the specialised sectors dealing with brightness, colour and movement which he had previously discovered.

Thus there is a physiological basis for the separate processing of achromatic and chromatic signals. As noted above, this is effectively assumed by most of the applied psychology literature which treats brightness and 'colour' as separate coding variables. Colour constancy also points to the significance of the adaptation of the eye to the stimulation of the total visual field, responding to light in some 'average' way which has been compared to selecting a camera stop. Adaptation can occur at a hierarchy of levels from the total illumination of the visual field down to local effects with features subtending only 1° or more (Hunt,1979). The adaptation level of a particular viewing context may become a neutral reference point from which perceived contrasts are measured (Cox,1973).

5.2 Colour Classification

As many thousands of colours can be discriminated by the human eye in juxtaposition, some form of classification is clearly required. While many different systems have been devised (Imhof,1982, chapter 4; Agoston,1979, chapter 8), they can be grouped into two basic categories- psychophysical (or 'objective') and psychological (or 'subjective').

In psychophysical colour measurement the definition of colours is based upon primary colour mixing, as on printers' colour charts, with colour scales devised technically (e.g. 10% tint intervals) rather than perceptually. The most comprehensive system of this type was defined by the Commission Internationale de l'Eclairage (CIE) in 1931, with three primaries known as X, Y and Z. These are slight mathematical transformations of red, green and blue (which can be mixed in light to produce any spectral hue) such that their values will always be positive across the whole range of colours. The CIE 1931 Standard Observer is defined by the amounts of each primary required for the average eye to match the wavelengths of the spectrum (colour-matching functions) in specified viewing conditions. The $\bar{y}$ colour-matching function is also known as the CIE photopic luminosity curve as it defines the level of brightness evoked in the visual system by a given amount of radiant energy of each wavelength). The function peaks in mid-spectrum (555nm) where it reaches its maximum luminous efficiency of 680 lumens per watt (figure 5.2). The composition of any colour can be determined by a spectrophotometer which measures the amounts of each primary (known as tristimulus values) present. Using a standardisation procedure where each tristimulus value is divided into the sum of the three values, chromaticity coordinates x, y and z are derived.

As these sum to 1, the chromaticity diagram, defining all realisable luminous colours for a given level of luminance, can be plotted using just the x and y coordinates (figure 5.3). Each point in the diagram

Figure 5.2 THE CIE PHOTOPIC LUMINOSITY FUNCTION

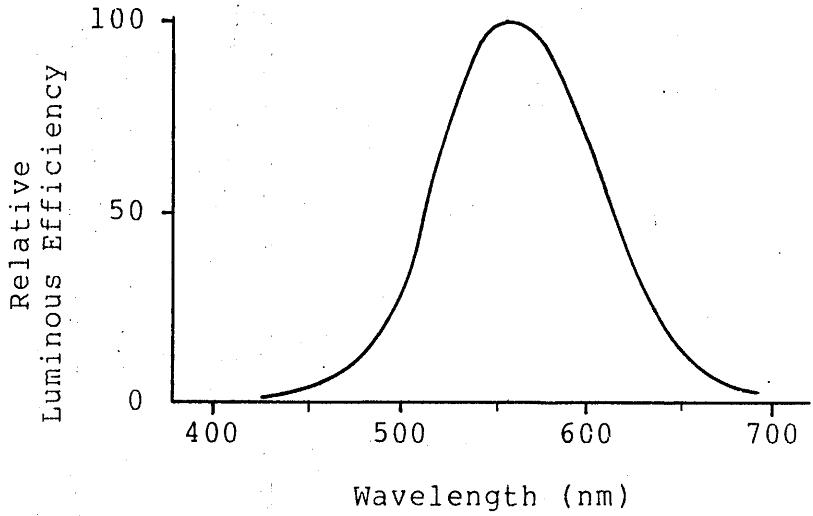
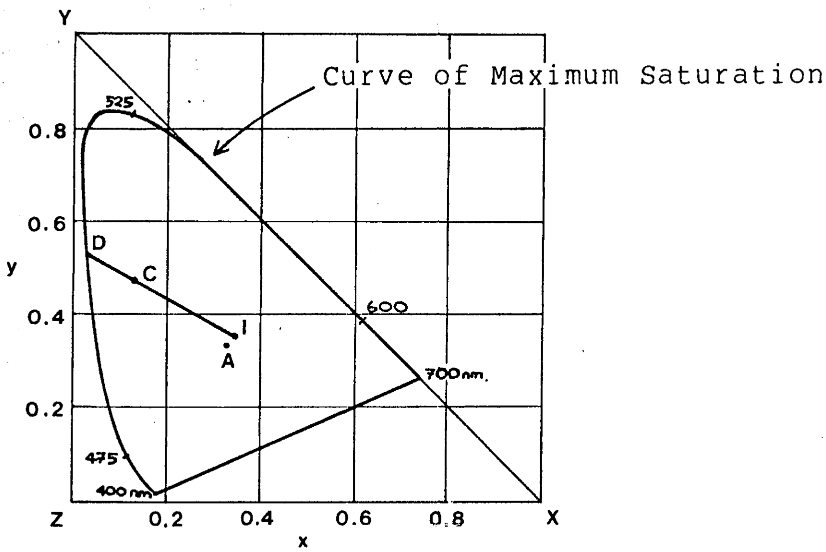


Figure 5.3 THE CIE CHROMATICITY DIAGRAM



A Achromatic point (x=y=z=0.333)
I Chromaticity of example illuminant
D Dominant Wavelength of Colour C.
Purity of Colour C= CI/ID (%)

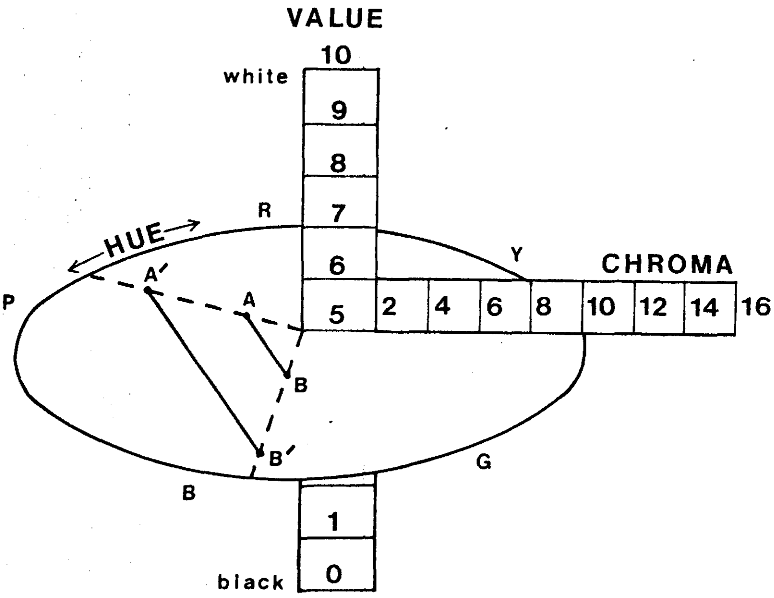
represents a 'shadow series' of colours of constant spectral composition but varying luminance. As the perception of surface colours varies with lightness (so that colours such as brown which do not exist in light can be seen), a third dimension is needed to define surface colour space, and this is provided by the Y tristimulus value which represents the luminous reflectance of the surface. Thus the x,y,Y surface colour solid is defined, the three axes being orthogonal.

The shape of this solid is defined by the MacAdam Limits- the maximum brightnesses theoretically attainable by a non-fluorescent surface colour of a given chromaticity, calculated by assuming perfect (100%) reflectance over the range of wavelengths required to produce that chromaticity (MacAdam,1935). Colours at these limits are known as 'optimal colours'. Here there is an inverse relationship between brightness and realisable saturation, because of the selective absorbtion of ranges of unwanted wavelengths necessary to produce a saturated colour. Thus the MacAdam Limits retreat progressively further within the curve of maximum saturation, like a contour map of a round-topped hill, as lightness is increased. With lights too, high luminosity colours appear to be less saturated: no yellow light appears as saturated as a red or blue one of the same purity and luminance (DeValois and DeValois,1975).

Luminous colours do not exhibit 'lightness' and two dimensions are sufficient to define each colour uniquely in CIE terms. However the colour perception can still vary with luminance and marked hue shifts have been observed (the Bezold-Brücke effect). CIE colours can alternatively be specified by their perceptual characteristics in terms of dominant wavelength, luminance/reflectance and purity (figure 5.3). However, although luminance is mathematically related to subjective brightness (Stevens and Galanter,1957) and purity to saturation, the gradations on each scale do not form perceptually equal steps. Various non-linear transforms of tristimulus values have been used to produce a psychologically-based system with perceptually uniform colour scales (UCS). One such is the 1964 u* v* w* system, developed out of the 1960 UCS diagram (Padgham and Saunders,1975), which incorporated the concept of the just noticeable difference. Two further refinements were presented by the CIE in 1976- the L* a* b* system recommended for surface colours (CIELAB space), and the L* u* v* system recommended for luminous colours (CIELUV space). In the latter, the plane of chromaticity for luminous colours is defined by u' and v', which are calculated from the tristimulus values, and from which figures for the perceptual attributes (hue, saturation etc.) can be derived. While no simple Euclidean space can perfectly match uniform colour space for given observing conditions, L* u* v* is probably the closest workable approximation (Shellswell,1976).

Of the other psychological colour classifications the Munsell system, first used in 1905, is by far the most common. In this scheme the three dimensions of perceived colour, named hue, value and chroma, are divided into perceptually equal spacings. These were recalibrated in the 1940s on the basis of 3 million ratio estimates of colour difference (Newhall et al.,1943). A painted card ('colour chip') is produced for any physically realisable surface colour at a given interval on each dimension. Thus a 3-dimensional colour solid can be produced with value forming a grey scale on the vertical axis, while chroma increases radially and hue varies around the circumference (figure 5.4). This is effectively a cylinder, and the specifications are effectively cylindrical polar coordinates. However the resulting shape resembles more a double cone, as because of limitations imposed by pigments the maximum realisable

Figure 5.4 MUNSELL COLOUR SPACE



The dependency of the distinction between two colours of a given hue separation upon their chromas can be seen by comparing the length of AB and A'B'.

chromas decrease towards the extremes of the value scale. Munsell specifications are assigned to colours by visual comparison with the standard chips in specified daylight conditions. The system is therefore particularly useful for choosing surface colours, and variants of it are also used for colour choice on self-luminous electronic displays (e.g. the Tektronix HLS (hue, lightness, saturation) colour model).

However, even with standardised viewing conditions, no colour classification system can take into account all the perceptual variables involved. Two surface colours with the same hue, lightness and saturation are not necessarily seen to be identical (Beck,1975). Factors such as the glossiness of the finish can come into play, creating 'colours' such as silver and gold which cannot be uniquely identified by standard classification systems. Even within the constraints of the map-using situation, different contexts can seriously distort the appearance of colours, and surface colours in particular. As Wyszecki and Stiles (1967) have stated, phenomena such as simultaneous contrast, successive contrast, colour constancy and memory colour have to date prevented the scientific prediction of surface colour appearance. The systems outlined above should be considered to represent the best available approximations for specific circumstances.

5.3 The Dimensionality of Colour Space

Doubts have been expressed about the orthogonality of the three Munsell dimensions. Cuff (1972a) suggested that hue, value and chroma are not independent because fully-saturated hues have inherent differences in lightness. The lightness for each hue at which its highest chroma is attainable varies (he suggests because of the spectral luminosity curve) from very light (value 9) for yellow to very dark (value 2) for purple-blue. This is clearly true, and such differences are of great importance to the present study. However it need not imply that the fundamental structure of the colour solid is wrong and its axes are really correlated, but merely that it is not elaborated symmetrically. This causes problems with the potential confounding of the effects of different variables. In fact several investigations have verified the perceived independence of the axes by different techniques. For example, in the study by Witzel et al.(1973) subjects varied the stimulus qualities in order to match one-dimensional colour differences, while Farmer et al.(1980) obtained single estimates of multidimensional differences from which the Euclidean equivalent of Munsell space clearly emerged by multidimensional scaling.

Are the three dimensions of colour therefore processed separately rather than as an integrated whole? According to Garner and Felfoldy (1970), dimensions are potentially integral if, as with colour, one can only be realised if a level of the other(s) is specified, and they coexist in one place. From experiments with lightness and chroma judgments, they suggested that colour variables were at a level intermediate between complete integrality and non-integrality, where they can be scaled separately in tasks requiring selective attention to individual dimensions, but are treated integrally in overall colour comparisons.

Although the scaling of each individual Munsell dimension is uniform, one value step does not represent an equivalent perceptual difference to one hue or chroma step. In order to derive a measure for overall colour difference, based on the length a straight line in colour space between the two relevant colours, the relative sensitivity of the value and chroma axes must be determined. Although various estimates have made one value unit (one tenth of the scale) equivalent to anything between 2 and 8 chroma units, most studies put it at between 3 and 4, or about a quarter of the used scale (Godlove,1951; L.G.Williams,1967b; Farmer et al.1980). Thus in terms of the eye's ability to discriminate differences, chroma is overall the weakest dimension in colour space, although it may be stronger with respect to other aspects of colour perception such as various emotional connotations (Wright and Rainwater,1962).

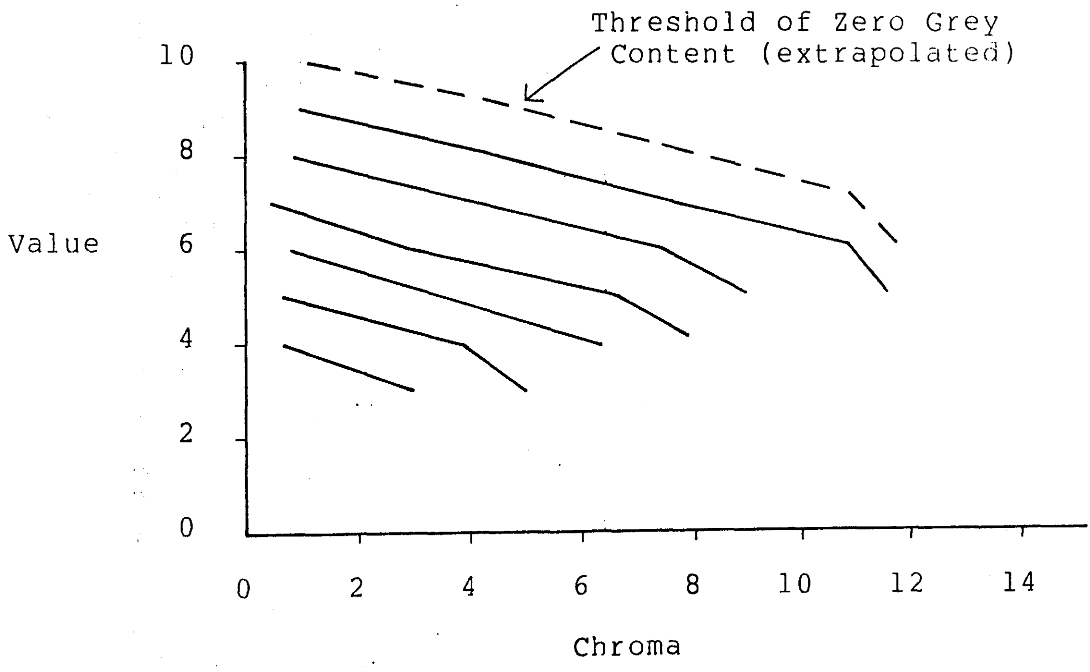
However, lightness and chroma do often vary together- for example in a standard lithographic tint series within a single hue. Davidoff (1974) studied people's ability to distinguish between the two separate sources of variation in such circumstances. He too used a multidimensional scaling technique, but asked his subjects to judge the distances between pairs with respect to lightness only, and subsequently to compare samples for lightness directly to a grey scale. In both cases, chroma differences interfered very strongly with the judgment of lightness. Evidence of an additive subjective connection between brightness and saturation is provided by the Helmholtz-Kohlrausch effect. It was noted in the 1920s that although the luminance and perceived brightness of a light source generally correlate well, an increase in saturation of the source caused a dramatic rise in brightness for most observers, especially at low luminances and with blue (Padgham,1971). The corresponding effect for surfaces is that darker colours are judged to have higher chromas, and vice versa (Newhall,1939; Evans,1959; Davidoff,1974). An inhibitory effect has also been noticed for figure/background colour combinations, where the chroma contrast appears to be bigger when the lightness contrast is small (Graves,1952; Silverstein,1982). Graves suggests the use of a thin bounding line to 'subdue and integrate' this effect. Overall it would seem that although people can make quantitative judgements demonstrably built up from two independent dimensions, they would normally consider those variables integrally. Clearly care must be taken with the precise specification of colourfulness used in experimentation.

Are there however more than three dimensions to surface colour perception? It has been suggested that variations in illumination effectively provide a fourth dimension to Munsell space, not only because of its effect on the degrees of pronouncedness of items in the visual field but also because it determines what intensities they require to become fluorent (Katz,1935). 'Fluorence' itself (otherwise known as 'brilliance') is, according to Evans (1974), the fourth dimension of non-isolated surface colours. Although it is related to lightness and chroma, and may provide some insight into their subjective connection, it must additionally be specified to fully describe the colour perception. It is a context-dependent effect which has been described as relating to artists' use of the words 'vivid' and 'bright' as in 'bright red' (Agoston,1979), or more precisely as the lack of perceived grey content in a colour.

The phenomenon was discovered by Evans (1959) in experiments involving visual matching of Munsell chips on a white background. He found a threshold of colours of apparently zero grey content (figure 5.5), above which (in value and chroma) samples appeared to become brighter than their surround and take on a fluorescent appearance intermediate between the surface and self-luminous modes of perception. He related it to an excess of overall stimulation above the adaptation level (background brightness) associated particularly with highly saturated stimuli slightly less bright than the background. He stressed that this apparent fluorescence was completely different from actual physical fluorescence, and that many

Figure 5.5 PERCEIVED GREY CONTENT OF MUNSELL CHIPS ON A WHITE BACKGROUND

Isolines of perceived grey content



Source: Evans (1959)

colour samples do fluoresce physically but do not appear to. Fluorence can become particularly clear in some road map printing inks in certain types of illumination (e.g.sunlight) at specific combinations of illumination and viewing angles.

Below this threshold colours were perceived as having a grey component in addition to their chromatic component. He obtained very consistent results indicating an increase in perceived grey content (i.e. a decrease in brilliance) with decreasing value and chroma, one value step being equivalent to about 3 to 4 chroma steps. Thus his 'gray-fluorent' scale is a measure of brilliance. In fact he claimed that the Munsell chroma scale is not a simple correlate of saturation but of a combination of saturation and brilliance. It would seem that the potential significance of this variable has not been considered in cartographic studies.

5.4 Extra Problems with Linear Colour

A considerable amount of research has been conducted into colour perception under conditions of stimulus degradation, which occur at low illumination levels and with very small stimuli. However, the objects used in the latter case are generally discs or similar 'point' shapes defined in terms of the angle they subtend in the eye. Narrow lines, which obviously subtend a very small visual angle in one dimension but a much greater one in the other, are clearly different- lines may be reduced in width below minimum acceptable diameters for point symbols as their length enhances their visibility (Robinson et al.,1978). Evidence for this is provided by Hecht and Mintz (1939) who found that against their highest background illumination level (30.2 mL), a line 0.5" of subtense wide was just visible, compared to a threshold of 0.5-1' diameter for 'points'. They concluded that with optical blurring caused by the diffraction of the pupil, this contrast was just sufficient at the retina to lower the stimulation intensity of a single row of cones just sufficiently (by 1%) to be perceived.

However coloured lines have been comprehensively ignored in the literature, as have the added problems of colour contrast created by the integration of differently coloured lines into networks. The experimental evidence mentioned below, often derived from work on 'points' or 'small areas', must be seen in this light. The following specific problems are however suggested by cartographic research and experience:

1) The range of colours that can economically be reproduced on a printed map may be severely restricted by technical considerations. For fine uncased lines in particular, overprinting of inks is often not practicable because of registration problems, and any kind of screening will reduce the sharpness and consequently the legibility of the line (Imhof,1972). Maps printed in process colours with uncased lines, such as the Ordnance Survey Routeplanner, are consequently particularly limited in terms of linear colour choice.

2) Small areas such as lines appear less saturated (Burnham,1951) as their subtense is reduced from 1°, so that high chroma inks and fully saturated hues are necessary for clear identification and coding (Arnberger,1966; Hopkin and Taylor,1979). For the route network to dominate visually on a paper road map, the hues should additionally be of low lightness (Keates,1973) for strong contrast against white/pale backgrounds. Imhof (1982) has stressed the need for pure and rich line colours for clear differentiation between lines and contrast against the background. He has also suggested that the only printing colours which should be used for lines are black, dark grey, dark brown, and saturated

purple, blue, red and green (Imhof,1972). Consequently, as noted above, most road symbols on road maps are not screened. Sometimes, however, screened road fillings are used alongside their solid parents to create extra distinctions within road classes (e.g. single/dual carriageway) although sufficient distinction can only be maintained within inks of high chroma (e.g.red) or low lightness (e.g.black).

3) As line width decreases, and apparent saturations drop, colours become less distinctive, and positive identification of specific colours becomes more difficult. Variations in perceived lightness and chroma are first to become too subtle for visual differentiation, so that colour ordering (chapter 6) becomes especially difficult. With further narrowing, qualitative distinctions are also lost as hue differences gradually disappear (Robinson et al.,1984). Hunt (1979) has stated that as the angular subtense of surface colours decreases to 15 minutes (corresponding to a symbol diameter of 1.3mm at 30cm viewing distance) sensitivity to blue-yellow colour differences gradually reduces to nothing because of the paucity of blue-sensitive cones in the fovea ('artificial tritanopia') and the chromatic aberration of the eye which causes blue light to be poorly focussed. At 5 minutes of arc (0.4mm) red-green sensitivity disappears too and only light-dark differences can be detected. These results are confirmed in studies by Judd and his co-workers (1969,1971) although their thresholds were 10 and 4 minutes respectively. The equivalents for line widths are unfortunately not known. Bright luminous colours are affected to a much lesser extent.

It has also been suggested (Kaiser,1968) that because of artificial tritanopia, blue and green become more confused and yellow tends towards white as features are attenuated. This corresponds with the findings of cartographic research and experience. Phillips and Noyes

(1980) noted such problems in matching small areas of colour to the key. Hiatt (1982), testing discrimination between fine coloured lines by partially sighted students, found green and cyan to be the least legible combination, and M.Wood (1968,p.56) noted how narrow yellow lines 'merge into an off-white background.' Red also emerges clearly as the easiest colour to distinguish in conditions of stimulus degradation and spatial dispersal of like symbols (Imhof,1972). It is the first colour that can be distinguished as illumination levels are increased from pitch black (Boynton,1979) and it clearly outperformed the competition in Hiatt's tests. It is perhaps not surprising therefore that colours other than red and black are very rarely used for narrow single lines on road maps.

In fact, apart from yellow and 'bright red' all linear colours on a white background tend to be perceived increasingly as black as illumination levels and/or line widths decrease (M.Wood,1968). Dark colours are attractive for lines because of their high contrast with the common white background- Imhof considers that black lines have the greatest visual impact. However, the darker the colours become, the more difficult they are to discriminate (Imhof,1972). This appears to be partially an effect of induction (also referred to as simultaneous contrast or Mach banding), a phenomenon thought to be caused by the physiological process known as lateral inhibition, whereby contrast is enhanced across edges. Heinemann (1972) has shown that because of brightness induction perceived brightness is constant for all objects more than a certain amount darker than their surround. Induction can occur both achromatically (with brightness contrast) and chromatically, where the line appears to adopt the opponent colour of the background (often known as simultaneous colour contrast). These have been shown to be separate and independent effects (Kinnear,1979). Opponent colours in juxtaposition will therefore enhance each other's apparent saturation. Kinney (1962), using self-luminous fields, demonstrated strong chromatic and brightness induction effects. Clearly these are less striking on conventional paper maps, where background colours are generally weak and chromatic induction is rarely observable.

However, Kinney's work shows that lines on these maps are in many ways particularly susceptible to induction effects, with high background-to-line size and brightness ratios and plenty of contiguity between the two fields. According to Imhof (1972), dark and highly saturated colours interact less with their (generally light and unsaturated) backgrounds. Given the aforementioned problem of discrimination between dark colours, he has consequently recommended the use of saturated colours such as strong red and intense green. The problem of discriminating between line colours is clearly greater than if the same colours were used on a choropleth map, because of the areal degradation of the stimuli and the very limited contiguity between the line classes (i.e. only at junctions). Changing the background colour can reduce the severity of these effects. Keates (1973) has recognised that colours often used in the line image such as red, blue, brown and purple, and black, may be similar in saturation and/or lightness, and has suggested that the introduction of chromaticity contrast through pale yellow or yellow-green backgrounds could be used to maintain their distinctiveness. Another more widely used and conventional technique is to edge the linear colour with thin black casing strips.

4) Casings introduce further problems to the perception of linear colour. While they reduce interactions between line and background colour, they themselves interact with the line fillings. This interaction often results in noticeable assimilation effects, with the adjacent thin strips of colour appearing

less contrasting, making the use of multicoloured lines problematic. For example, a yellow line between red casings can take on an overall orange appearance. Whereas black lines separating coloured areas seem to make the colours appear brighter and more saturated (Phillips and Noyes,1980) adding a black casing to a coloured line reduces both its apparent lightness and chroma. This can be clearly seen on Esso road maps, by comparing the 'primary routes' and 'other A roads' symbols.

However there seems to be no clearly observable relationship between the magnitude of this effect and casing/filling width (ratios) or colour contrast. Thus the problem is unlikely to be caused primarily by the limited spatial resolution of the visual system. In fact these symptoms, also referred to as Von Bezold spreading effect, are not easily explainable. However, a small inducing surround does seem to be a requirement (Helson and Rohles,1959). DeValois and DeValois (1975) have suggested that because of optical blurring the visual system tends towards spread unless there are powerful contrary neural interactions, and that if the effect of the surround of a cell's receptive field is lessened by narrow borders around an area (e.g. casings) then spread is to be expected. The effect seems to occur when a coloured pattern (such as is created by two parallel casings) is imposed on a differently coloured background (Walraven,1984). Assimilation effects could also cause problems in networks of differently coloured lines, with short links taking on the colour of the lines they join.

5) With further stimulus degradation, lines of some colours may not be seen at all. Blue and violet are recognised as being especially poor defining colours for fine lines (Taylor and Belyavin,1980), and this has been confirmed by recent work on self-luminous displays (Huddleston,1982) which use blue phosphors of necessarily low luminance. In fact only 10% of retinal cones are of

the 'blue-sensitive' type. Unlike the 'red' and 'green' cones, there is a whole range of visible wavelengths to which they do not respond at all. Various investigations have suggested that, especially in small fields and at the brightness levels of surface colour (J.J.Sheppard,1968), they make little or no contribution to the perception of contour (edge contrast). As this appears to be mainly a function of brightness contrast, the brightness signal in the LGN may contain only the outputs of the red and green cones, plus the rods (DeValois and DeValois,1975; Tansley and Boynton,1976; Boynton,1978). This would seem to support M.Wood's (1968) contention that 'hard colours' such as red are best for acuity/definition. Robinson et al.(1978) also consider that less saturated colours which mix a variety of wavelengths such as brown, are poor for the definition of fine detail, both in lines (e.g. contour lines) and as a background to them. Thus yellow, which is monochromatic and of high lightness (i.e. contrasts well with standard line colours) is Robinson's ideal background colour, although it may be thought desirable to reduce contrast and 'tone down' excessively conspicuous line work' (R.M.Taylor,1984; Thake,1979).

Overall it can be seen that with lines the objective of a hierarchically perceived series of clearly contrasting fully-saturated hues is, as A.Morrison (1971) discovered, very difficult to realise.

5.5 Colours and Distinctiveness

Colour difference formulae based on the geometry of psychological colour spaces are readily available for determining the straight line distance between a pair of colours. For Munsell colours, the formula used throughout this thesis is:

$$\text{Colour Difference} = \sqrt{[2C1C2(1-\cos(3.6\Delta H)^\circ)+(\Delta C)^2+(3.76\Delta V)^2]}$$

where C1 and C2 are the chromas of the respective colours, and ΔH, C and V are the differences in hue, chroma and value steps respectively. As this is a modification by Farmer et al.(1980) of a formula proposed by Godlove (1951), differences calculated by it are subsequently referred to as 'Farmer differences'. For qualitative colour coding, large colour differences are clearly desirable to maximise the distinction between classes, and various sets of a given number of maximally contrasting colours have been derived from colour catalogues (e.g.Kelly,1976; R.M.Taylor,1977). When an ordered use of colour is required, the problem is more that of the <u>minimum</u> amount of distinction that can be used between colours whilst maintaining their perceived individuality and non-confusability. Several studies in the human factors literature have looked at this problem from the point of view of the maximum number of different colour codes that can be reliably used for targets on one display. The resulting recommendations range widely from 3 to about 70 depending on a number of factors, such as whether the colours are seen together or isolated, self-luminous or surface, whether absolute or relative judgements are required, the variety of contexts in which any colour is seen, the size and brightness contrast of the targets, the level of accuracy taken to be acceptable and how much the observer has practised the task (Bishop and Crook,1961; Huddleston,1982; Silverstein,1982).

The most basic level of distinction is provided by the 'psychological primaries', or 'unitary hues'- red, green, yellow and blue, plus white and black, the three opponent pairs in the visual system- which each appear to be completely individual colours rather than mixtures of others (DeValois and DeValois,1975; Agoston,1979), although as an artist Arnheim (1954) finds the status of green arguable, while Robinson et al.(1978) add brown to the list. For effective coding at a visual eccentricity of 15°, only these hues (in highly saturated versions contrasting highly with their backgrounds) should be used (Kinney,1979). Much of the work has examined how far this hue code can be extended before absolute identifications from memory become insufficiently reliable, and there is a consensus of about 10 named spectral hues, plus achromatics and 'purples' (Bishop and Crook,1961; Grether and Baker,1972).

Bishop and Crook also examined the role of brightness and saturation in the distinction of self-luminous colours, and found that up to 3 levels of luminance and 2 of purity were potentially usable, but not for every hue. Overall there are 125 just noticeable differences of hue, and up to 200 jnds of brightness and saturation in some hues (depending on the shape, size, texture and context of the coloured object). In the aforementioned experiment by L.G.Williams (1971), the subjective differences on each Munsell dimension represented by a RFR of 0.5 were 4.8 hue units (at an unspecified but high chroma), 1.1 value units, and 2.7-3.1 chroma units. In terms of proportions of the maximum available range, these represent 9.6%, 11% and 16-19% respectively. Especially for lines on maps, with fully-saturated colours of often poorly discriminable lightness, hue is clearly the dimension which allows the widest range of distinctions, with the assistance of those inherent variations of lightness that are clearly visible.

Experiments where the entire processing effort is devoted to discrimination are clearly somewhat unrealistic, and for real map using tasks only 6 or 7 hues are recommended. Miller's (1956) widely-cited 'magic number' of 7 plus or minus 2 as the limit for the human retention of unorganised data items may be relevant here. In a search experiment conducted by Cahill and Carter (1976) using up to 10 maximally distinct colours, overall search times began to increase very rapidly when more than 7 were used. Bertin (1983a) recommends a maximum code for lines of 6 colours of equal value (grey, violet, blue, green, brown, red). Maps of rail networks where colour is used solely for the qualitative coding of lines also seem to confirm this. The well-known and highly commended London Underground map, introduced in 1934, now has to cover 9 lines, which is achieved by the use of 7 chromatic (red, brown, yellow, green, cyan, blue, purple) and 2 achromatic (grey, black) codes. The map of the Trans-Clyde rail network uses the additive and subtractive primary sets- red, green, blue, magenta, yellow, cyan (plus the much thinner orange underground line)- but the addition of a further code (maroon) creates potential distinction problems. With narrower lines, the usable code becomes still smaller.

Dobson (1983b) has however suggested that for maps with a key somewhat more classes are usable because the key acts as an 'anchoring stimulus'. Also where colour is used redundantly (e.g. with line width or casing) smaller colour differences may be used as distinction between classes is reinforced by the other dimension. It appears that even on maps with keys, a colour can be identified far more positively if it can be allocated to a basic colour name. In tests of layer tints produced by lightness/chroma gradations within one hue, the classes were insufficiently distinct for reliable absolute height judgments (Audley et al.,1974; Phillips,1982). Audley suggested that in map/key comparisons, verbal information becomes more important where there is a longer interval between viewing the map and the key, as the sensory information decays rapidly. For a colour to be named identically by the majority of people it must be within a certain hue range and of sufficient saturation to be recognised as being of that hue. Additionally within certain hues sufficient lightness and saturation contrast is available to create separately named colours (e.g.brown). Berlin and Kay (1969) have noted that the colour terms of every language draw from precisely eleven basic categories- white, grey, black, red, yellow, green, blue, brown, purple, pink and orange.

The precise minimum requirements for qualitative distinctions between colours for road symbols remain to be tested.

5.6 Limitations of Colour Use

Apart from obvious considerations of reproduction costs, the implications of two basic groups of limitations must be considered in any use of colour on maps, namely:
1) variations between people's visual systems and their perceptual responses to colour, and
2) variations in colour appearance caused by the conditions of map usage.

In the first category the most serious variations are caused by colour-defective vision (colour blindness), which affects about 8% of the male population but less than half a percent of females (Judd,1943). This appears to be related to a scarcity or complete absence of certain cone photopigments, making distinction between reds and greens on the grounds of their chromatic content difficult or impossible. Any other colour defect is extremely rare. In terms of map design, it may therefore be considered desirable to avoid coding with equally bright reds and greens (Kaufman,1975). Unfortunately the fully-saturated reds and greens commonly used on road maps are usually of very similar lightness, and in tests by the Consumers' Association (1963,1971) several of the maps tested were problematic in this respect. Colour-defective users also found maps with a white background generally easier to use. A useful design guide is provided by Kelly's (1976) listing of 22 maximally distinct colours, of which the first 9 (white, black, vivid yellow, strong purple, vivid orange, very light blue, vivid red, greyish yellow and medium grey) also provide maximum contrast for the partially colour blind. However, degradation of the visual system with age is a further problem. Old people gradually lose in particular both their sensitivity to blue, as the 'blue' cones appear to be more vulnerable to the diseases of the retina (Lansdown,1985), and their ability to accommodate it for adequate focussing (Dobson,1983a).

In the second category, colour is also differentially affected by variations in the quantity and quality of illumination. Road maps are sometimes read in 'mesopic' conditions (e.g. in cars at night) where the illumination is inadequate for full cone vision. While this clearly produces an overall degradation of the stimulus, reducing the efficacy of any coding variable, it causes specific, consistent distortions for colour because of the so-called Purkinje shift, whereby the peak of the CIE luminosity curve moves from 555 to 507 nm for scotopic vision. This could clearly affect both the relative perceived brightness of colours and how long they stay recognisably chromatic as illumination decreases. For example, red loses its saturation and vividness faster than blue (Katz,1935). However, an interesting anomaly is that around the threshold of chromatic perception, the two curves remain similar above 560 nm implying a more persistent involvement of the cones (J.J.Sheppard,1968). This could account for Boynton's (1979) observation that red is the first colour that can be seen as light levels are gradually increased from pitch dark.

However, the spectral quality of the illumination is equally important, and under amber sodium street lighting, the relatively pale reds found on smaller-scale road maps have been found to fade and even disappear at these illumination levels (Consumers' Association,1971). Colours in the yellow-red region, commonly used for road symbols, suffer particularly, and information relying on chromatic differences (such as road type) tends to become illegible. The best maps in such situations are consequently those with white backgrounds and the largest lightness contrasts (Consumers' Association,1979). Otherwise use of only those colours which maintain reasonable distinctiveness in these conditions (basically green, blue and black) would severely limit road map legibility in daylight.

6. THE ORDERED USE OF COLOURS

As mentioned above, colour is often applied to line symbols on maps in order to make solely qualitative distinctions, generally by using widespread hues and colour associations (Cuff,1972a). On road maps, however, where the road categories are ranked, any hierarchical implications of the symbol colours are at once significant. The map designer can either ignore them at his peril or attempt to take advantage of them.

In a colour sequence representing an ordered series, the colours chosen should be intuitively and/or perceptually appropriate to the dual aims of 'order' and 'separation' (Trumbo,1981). However, while it is relatively easy to employ colour space to produce very different colours, maintaining order at the same time presents serious difficulties. Arnberger (1966) suggests that hue differences should be used for separation, and 'colour weight' (see figure 6.1) for order, and this kind of relationship appears to be conventionally accepted wisdom (R.M.Taylor,1985). Macioch (1983), in his specification for a 'system of cartographic denotations', mentions the properties of colour intensity (brightness and colourfulness) and chromaticity, and describes the latter as the qualitative function for distinguishing information classes, and the former as the ordering function for ranking information within those classes. This would seem to be eminently feasible as brightness and colourfulness are perceptually ranked scales (i.e. one object can be brighter and more colourful than another) while hue is simply a distinguishing scale. However, hue may have a quantitative influence in a non-linear way: certain specific hues, for example, have strong cognitive associations with quantitative concepts such as strength and importance.

Consideration must be given to five ways, at successively higher levels of processing, in which the use of different colours may affect the saliency of otherwise identical linear stimuli:

1) variations in the relative visibility of colours and their conspicuity in road map contexts.
2) the potential creation of visual levels by the sensory and/or cognitive effects of colours on depth perception.
3) the magnitude implication of a colour, effectively its density or intensity.
4) variations in the perceived pleasantness, or affective values, of colours.
5) the observer's experience with respect to the indirect associations and conventional usages of colours.

At first sight this list may appear daunting to someone attempting to find a single consistent specification of the quantitative significance of colours. However, not only are some effects far stronger than others, but it seems that the first three at least may well be driven by the same colour properties.

Interpretation of previous work is however often difficult, as results are quite commonly presented in the form of ranked lists of colour names, with the colours used being inadequately specified. In themselves such studies are just a consideration of a few discrete points (normally highly-saturated spectral hues) from the vast ocean of colour space. Without colour specifications they may be of little use in an attempt to find a single, unambiguously interpretable line through colour space representing the underlying dimensions of colour which are producing these effects.

Figure 6.1 COLOUR WEIGHT AND BRIGHTNESS OF SPECTRAL COLOURS

(after Arnberger, 1966,p.278)

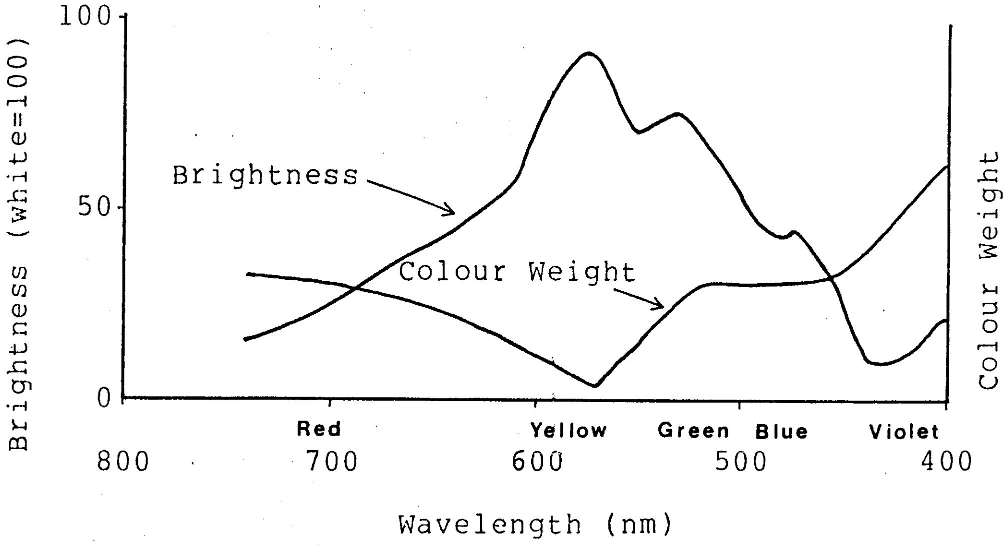
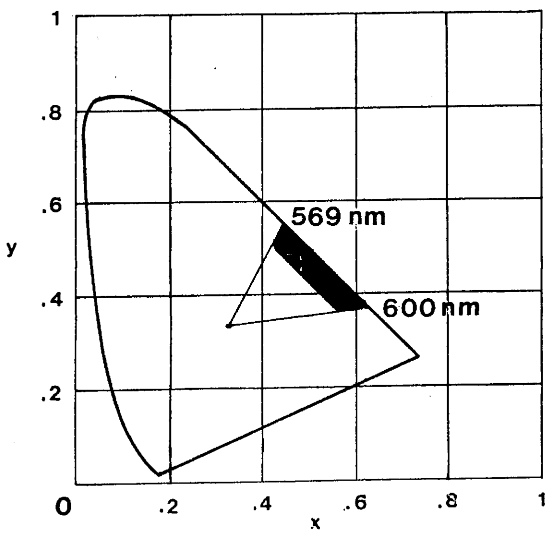


Figure 6.2 CIE CHROMATICITY DIAGRAM AREA OF HIGH VISIBILITY COLOURS



Area as defined by British Standard BS 4610, 1970

6.1 The Relative Visibility of Colours

It would appear that, both at and above the threshold of perception, visibility in object-background colour combinations is determined primarily by brightness contrast. For an object to be visible, it must meet certain minimum requirements for size (visual angular subtense) and contrast with its background (R.M.Taylor,1985; Clark,1985). 'Usually, brightness...differences between symbol and background, rather than hue and chroma differences, are regarded as necessary for legibility, although in theory it should be possible to resolve critical detail with hue or chroma differences alone.' (R.M.Taylor,1985,p.196) Luckiesh (1944,p.242) referred to brightness as the 'backbone of visibility', with colour (i.e. chromaticity) as its 'secondary aspect'.

Studies of visibility have related mainly to text legibility, which is obviously an easier testing condition than coloured lines, as it can be read back to verify correct communication. Several tests were carried out in the 1930s and 1940s using various text/background colour combinations and experimental measures such as recognition thresholds, subjective legibility assessments and reading times (e.g. Sumner,1932; Vickerstaff and Woolvin,1944). The latter have become the most common in more recent studies, which include work on CRT displays (e.g. Bouma,1980; Bruce and Foster,1982). The results of McLean (1965) are difficult to interpret because he compared brightness contrast in an achromatic situation with a 'color contrast' measure which was simply lightness differences between a set of colours. Otherwise the researchers are unanimous in the view that brightness contrast is by far the most important determinant of legibility, except

1) where chromatic contrast is very small, e.g. between

red and magenta, combinations are less legible than would be expected (Bruce and Foster,1982)

2) where brightness contrast is low, chromatic contrast can considerably improve acuity (Potash,1977; Watts,1980)

3) where specific combinations of low brightness contrast are particularly aesthetically pleasing.

It has been suggested that the CIE 1924 photopic luminosity curve (figure 5.2), as a measure of the relative sensitivity of foveal vision to energy of different wavelengths, defines the relative visibility of hues. Yellow-green, at the peak of the curve (about 555 nm), has been recommended for situations where high acuity is required (Potash,1977). Again a clear distiction must be made between lights and surface colours. British Standard 4610 specifies an acceptable range of colours for high-visibility clothing (BSI,1970), which is generally seen against relatively dull and achromatic backgrounds such as roadscapes. It defines the colours with the highest visibility to be

1) within a trapezium on the CIE chromaticity diagram (figure 6.2) covering saturated hues in the green-yellow to orange-red part of the spectrum,

2) above a minimum required luminance factor (Y value), which decreases with increasing wavelength.

These colours are consequently broadly within the higher reaches of the luminosity curve. However, given the limited purity of available dyes and pigments, these standards may not always be attainable without the use of fluorescence.

Because of the impurities introduced by pigment and paper on conventional maps, printed colours generally combine many wavelengths, so that while saturated hues do have inherent variations in brightness, the influence of spectral luminosity is considerably diluted. In any case, inherently light hues contrast least with the paper and so have little visual effect. Thus with positive contrast, lightness and emphasis are often working against each other (figure 6.1). There are however exceptions to this. Light hues can be effective when surrounded by darker colours, for example as a filling colour for cased road symbols (Arnberger,1966), where a colour such as yellow attracts whilst contrasting with the more emphatic darker casings. The 'compelling attraction' of a yellow background and the reduction it causes in the blur and glare of the image have also been cited by Birren (1941) to explain the superior legibility of black on yellow to black on white in several tests.

Differences were also found between combinations and their reversals (e.g. black on white/white on black). This suggests that visibility is more than a simple commutative contrast relationship, with absolute brightness being of significance in certain circumstances. Various perceptual phenomena might account for this. One such is irradiation, the observed effect that white lines on a black background appear wider and more visible than black lines of equivalent gauge on a white ground, probably because of a slight spreading of cone excitation (Greenberg,1971). One reason for this is that when the eye is adapted to a mainly dark image, the white lines appear to spread through being 'overexposed.' Also where the luminance of a colour is slightly above that of its surround, it appears to be self-luminous and has the fluorescent quality named 'fluorence' by Evans (1959). With higher luminance it becomes still more brilliant. These effects are more difficult to explain as they have no psychophysical correlate. Insistence (see below) may also be of relevance. Generally however the absolute levels of stimulation involved with reflecting surfaces are too low for lightnesses in themselves to have dramatic effects.

Other experiments have investigated the relative spatial penetration of small colour stimuli. Tests with signal lights (presumably against a dark background) have suggested a visibility ranking of red first, followed by green, yellow, white, blue and purple (Birren,1941 and other sources reported in Wyszecki and Stiles,1967). This confirms the poor performance of blue and purple in small visual fields noted above. Birren continues (p.309) 'The quality of visibility is largely dependent upon the attraction that any light source will have for the eye. Although white and yellow light may travel farther through distance, they will not be recognised as quickly as red or green.' This series is the source for Robinson's and Saunders' lists of the supposed relative sensitivity of the eye to different hues. On its own, it could be interpreted in several ways. It could be seen as a modified spectrum, reflecting perhaps differential contributions of the R,G and B cones and/or some specific perceptual effects of hue. Alternatively, apart from red, a colour well known to be preferable for distance viewing (Potash,1977), it correlates quite well with the CIE luminosity curve.

A study by Frenzel (1965) used patches of 10 solid (and equally poorly specified) printing colours, exposed to bright daylight, but covered over with 14 sheets of 0.3mm thick Astralon which were removed one by one until gradually all the colours appeared. This method was criticised by Arnberger (1966), who considered that the colour properties were altered by the Astralon acting as a colour filter and selectively dispersing certain wavelengths. However, Frenzel's broad conclusions were that the strongest (most saturated) colours were the most penetrating, while the darkest ones had the sharpest outlines. Thus the significance of lightness contrast in contour formation is again tentatively confirmed, while at the same time saturation is coming increasingly into the picture.

6.2 THE CONSPICUITY OF COLOURS

Experience suggests that the attraction a colour patch will have for the eye will generally increase with

1) its size (up to a point certainly not reached by lines), and

2) the more extreme and contrasting its lightness and saturation are in the visual field (Evans,1948; Graves,1952; Saunders,1961; Robinson and Sale,1969; BSI,1970).

However, a more precise definition of the causes of conspicuity in colours remains elusive, presumably because emphasis and attention are by their very nature relative, and hence complicated immensely by the multiplicity and variety of contrasts present in individual contexts. The body of evidence on the subject is severely disjointed.

6.2.1 Conspicuity and Visibility

The relationship between conspicuity and visibility is problematic. The visibility of a symbol seems to depend primarily upon contrast with its immediate environment, and obviously a symbol must be clearly visible before it can be conspicuous. Conspicuity, however, is concerned with the allocation of scarce processing resources across an often highly complex visual field which contains many competing features, each highly visible in itself. It could perhaps be said to be the icing on the visibility cake, the added extra property which makes a red line attract and hold the attention more than a black line of the same gauge. In which case, what is it that determines the relative conspicuity of colours? Is it still basically brightness contrast, or some other measure of colour difference that is most significant? Are there in addition inherent properties of certain colours which make them conspicuous in most contexts? If so, can they be defined on existing scales (e.g. of brightness, colourfulness, intensity or brilliance) or perhaps related to some physiologically-induced differences in sensitivity to particular wavelengths across different parts of the retina?

Two concepts described by the German psychologist David Katz are of interest in this respect, namely 'insistence' (Eindringlichkeit) and 'pronouncedness' (Ausgeprägtheit). He described the insistence of a colour as 'the strength with which it bores its way into consciousness.' (Katz,1935,p.280) Of a pair of colours, the more insistent is the one 'which seems to possess the power of catching the eye more readily and of holding it more steadily. In general, if two achromatic colours are seen in the same illumination and against the same background, the brighter will be of this nature.' (p.108) Insistence is not derived from brightness, but is in fact the more primitive experience, relating directly to the absolute intensity of retinal excitation. On the other hand, pronouncedness is less a measure of how inherently 'gripping' a colour patch is than of its clarity and prominence in its specific local context. In general the endpoints of the grey scale are the most pronounced, although only at the light end does this coincide with high insistence. Pronouncedness varies greatly with the level of incident illumination. Thus insistence and pronouncedness would seem to relate broadly to our respective definitions of conspicuity and relative visibility.

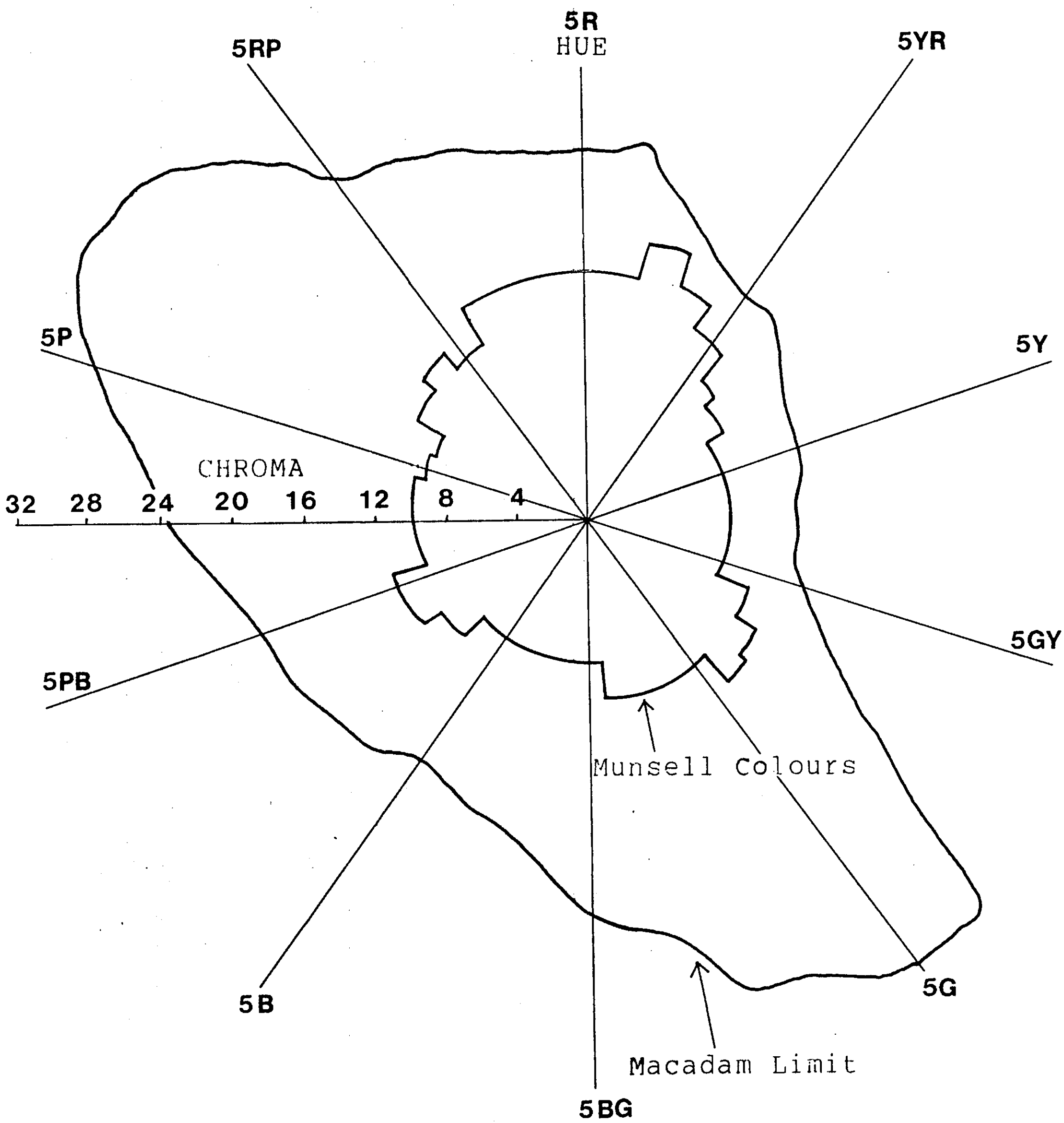
While the visibility of a colour may define its ease of detection, its conspicuity is also dependent on its ease of identification or recognition. In this respect the inverse relationship between brightness and saturation for lights and optimum surface colours correlates interestingly with the results of Birren's experiment noted above (section 6.1), where bright mid-spectral colours were superior for detection while more saturated end-spectral colours were more easily recognised. Tests of text legibility have also recommended dark/end-spectral colours for identification, while bright/mid-spectral colours are better used as a background (Huddleston,1982). If brightness and saturation are both important, how do they relate to each other in the determination of conspicuity?

It is unfortunate that the more recent literature on the conspicuity of colours does not problematise the criteria used to determine conspicuity. This would perhaps be excusable if the same criteria were used in each study, implying that these factors may be common knowledge somewhere along the line. Apparently they are not. Shellswell (1976) ranks surface colours in a 'visual order of importance' in terms of differences in reflected energy (ΔE) from a white reference surface, measured in L* u* v* space. This implies that dark colours with high chromas are visually most important. Hunt (1979) is a little more precise in his definition, but just as quiet about the basis for it, although he appears to stress the recognition aspect. For luminous colours, he defines conspicuity by saturation: distance from the subjective achromatic point on the u' v' diagram for a given luminance. Presumably the rationale behind this is that the 'average' colour in any normal field of view would tend to be achromatic. In contrast to the luminosity curve, the two ends of the spectrum (red and violet) are the most distant spectral hues and thus would be expected to be the most conspicuous. Hunt however adds that violet in light is often inconspicuous because of its inherent darkness (and potentially the 'small field' effects mentioned above). The equivalent basis for the conspicuity of non-luminous colours in L* u* v* space is perceived chroma, of which the values for both optimum and 'typical real surface' colours peak in the orange-red region, making these colours the most conspicuous. He does not mention whether perceived chroma values can legitimately be compared across different lightness planes.

At this point it is important to remember where specifications relate to optimum colours that the range of colours that is physically realisable using current papers, pigments and phosphors falls well short of optimum. This is especially true for surface colours. Munsell chroma maxima invariably fall well short of the MacAdam Limits (figure 6.3)- the extreme case being at 7.5PB where the maximum of 13.5 (at value 1.7) compares with an optimum of 40. The inherent lightnesses of the chroma maxima for each hue do still vary with spectral luminosity. Makowski (1967) however demonstrated an equalising effect of spectral impurities upon printed colours, making distinctions between them in terms of their inherent lightness potentially difficult- e.g. between standard reds, greens, blues and browns. This is particularly marked in lines, as mentioned above. Obviously effects which are either derived from or vary with hue become less marked as chroma is reduced.

Interestingly Hunt's assessment of the most conspicuous colours correlates well with the experience of map analysts and cartographers. Taylor and Belyavin (1980) noted that saturated yellows (disagreeing with their own citation of Hunt), oranges and reds are usually the most attention-getting colours on maps, while Keates (1973) recommended red, orange and violet for emphasis of small features against relatively light and unsaturated backgrounds. The superior 'attention value' of red, against both uniform and varied backgrounds, is well-established both experimentally and in artistic and cartographic convention (Graves,1952; Saunders,1961; Cuff,1972a), with black fairly close behind. Yellow is again controversial- alternatively rated poorly for its lack of contrast with the paper, or highly for its alleged

Figure 6.3 MACADAM LIMITS AND THE GAMUT OF MUNSELL COLOUR CHIPS AT VALUE 5



Information Source: Nickerson and Newhall,1943; Munsell Color,1970.

presumably chroma-based attraction. Before consideration of the experimental evidence, it would therefore seem that the conspicuity of a colour is likely to be largely a function of contrasts of brightness, the amount of overall sensory stimulation, and colourfulness, the amount of chromatic stimulation.

6.2.2 Conspicuity and Colour Differences

Some experimental investigations of the relationship between conspicuity and colours have used search tasks where the subject is seeking a target of known specification, and have consequently been interested in colour differences between separate objects in the visual field rather than those between individual objects and the background or an achromatic 'average'. L.G.Williams (1967b,1971) used relative fixation rate (RFR) as a measure of conspicuity, and found that it declines (i.e. targets become more conspicuous) with increasing difference between target and background (distractor) objects on each individual Munsell dimension (hue, value and chroma). His findings on the relative sensitivity of the dimensions correlate well with established Munsell colour difference formulae (Godlove,1951; Farmer et al,1980), with 1 value unit equivalent to 3 of chroma. Carter and Carter (1981), aiming for a quantitative expression of 'the principle that things are easier to find if they contrast with their environment', applied various colour difference measures (including the 1976 L* u* v* and L* a* b* formulae) to their experimental data. They found that over half the variance in their conspicuity indices (RFR and search time) was explained by overall colour difference between target and background objects. They suggest that this proportion could be further improved by modifying the colour difference equations to take account of perceptual quirks such as induction and small field effects. A minimum distinction between colours of 40 CIELUV units is also recommended. Farmer and Taylor (1980), using Godlove's colour difference formula, discovered that for a target to 'emerge from' its background

1) target/background contrast, and

2) similarity between background items

(in colour and form) should both be maximised. In a separate paper (R.M.Taylor and Farmer,1980) they suggest a minimum usable distinction of 4 just noticeable differences between colours in quantitative schemes, although this would probably be grossly inadequate for lines.

Because of the prespecified target however, these search experiments involve a rather different usage of 'conspicuity' (L.G.Williams,1966) from the above definition. The experimental measures are affected far less by the inherent properties of different colours than by the quality (i.e. distinctiveness) and size of the groupings they define (Smith,1963). Following preattentive colour separation, the subject directs his attention to the precued target grouping (colour), and continues to search within that group (L.G.Williams,1967a). Thus the attentional pattern is internally determined. In tests of the conspicuity of motorcyclists (Fulton et al,1980), when the task was to report their presence or absence in different situations, they were always detected irrespective of their clothing. However, in street interviews where pedestrians were not approached and asked until they had been seen looking towards the motorcyclist, the wearing and size of fluorescent garments was found to improve the perception rate significantly.

While it would be possible to enhance the conspicuity of a single grouping in a classification by maximising its colour differences from the other classes and minimising the differences between them, such qualitative distinctions cannot easily be used alone to create a smooth hierarchy of classes. Clearly the distinctiveness of a class is a prerequisite for it to be conspicuous. In a normal line series however, in order that each class be distinct, inter-class differences tend to be very even. Classes may however be clearly ordered by their differences from the background or adapted colour.

6.2.3 Conspicuity, Lightness, Chroma and the Periphery

Where attention is wholly or partially involuntary, however, what map symbol characteristics attract it? Dobson (1979b) reports several studies suggesting that attention is primarily directed at the elements in a display which contribute the most visual information- i.e. the least 'probable' and therefore most contrasting and visually dominant marks. Swezey's (1984) study confirmed the significance of stimulus intensity in influencing attention and retention, with performance on perceptual (but not necessarily cognitive) tasks deteriorating with decreasing intensity. Dobson's own work, using a statistical map, showed that the probability of fixation on a particular mark increased with its overall or local graphical dominance. In terms of colour contrast, if the adaptation level is some achromatic average, the most dominant colours will be the most contrasting- i.e. those with the greatest colour intensity.

The relationship between lightness contrasts and conspicuity in an achromatic situation has been investigated by Engel et al.(1971,1972). The measure used was Engel's (1969) 'conspicuity area'- the retinal area within which a target can be perceived in a brief presentation allowing only a single fixation. Using light background discs and a dark surround, they found that the target was most conspicuous (i.e. it subtended a larger conspicuity area) at a luminance midway on a log scale between those of the background objects and the surround. It should be noted that the background objects took up a considerable proportion of the test area. Watts (1984) found that his experimental measures (the detection distance of a cyclist approaching from a 30° eccentricity, and the amount of room given to cyclists by passing cars) were strongly associated with both the area of the 'saturn yellow' conspicuity aid worn by the cyclist and its luminous reflectance. This confirmed the results of his own previous research.

Logically the conspicuity of a stimulus due to colour ought also to be related to the amount of chromatic stimulation it causes, which is presumably a function of its saturation or chroma and its angular subtense. Chroma contrast was included in a further experiment by Engel et al.(1972). They tested the spatial thresholds of conspicuity of a coloured disc in the periphery of an otherwise achromatic visual field of white background discs with a black surround. The Munsell hue and chroma of the disc were varied. While hue-related differences were found increasingly away from the fovea (blue, green and red/orange being the most conspicuous), their most interesting finding was that the conspicuity area was determined in a given context by <u>either</u> lightness or chroma contrast, whichever would by itself subtend the larger conspicuity area. So as brightness contrast reduced, chroma contrast took over, indicating that the relative conspicuity of sets of coloured lines of similar lightness might be primarily determined by chroma, although the perceptual differences between them might consequently be quite small. In Watts' (1984) experiment, the chroma of 3 jackets (white, orange and yellow) of similar reflectance had no significant effect. However he also suggests that when lightness contrast is low, and especially against unsaturated backgrounds such as a concrete facade or a cloudy sky, chromatic contrast may aid detection. This may provide some insight into why yachtsmen find orange marker buoys to be considerably more conspicuous than yellow ones. Similar tests with more chromatic backgrounds/surrounds would be of obvious interest.

Certain consistent differences have been found in the relative sensitivity of the eye to different wavelengths with respect to the spatial distribution of cone types in the retinal periphery. The visual periphery is clearly very important in conspicuity as it is from here that features are 'acquired'. Because of the paucity of cones and corresponding bipolar cells more than 10° out from the fovea, the visual system becomes progressively less sensitive overall with increasing eccentricity. Generally features in the periphery must be highly discriminable in order to compete with those viewed foveally (Dobson,1979a). The insistence of a given item in a uniformly illuminated field decreases towards the periphery until finally colour vanishes (Katz,1935).

At the maximum eccentricities relevant to a single involuntary eye movement (10-15°), the classic symptoms of stimulus degradation (see above) begin to take effect so that stimulus size becomes all important to both feature perception and colour recognition (Barbur,1979; Gordon and Abramov,1977). Hues must also be highly saturated in order to be perceived correctly (Kinney,1979). Sensitivity to green is also gradually reduced (Weale,1951; Kinney,1979), but the periphery is relatively much more sensitive to short (blue) wavelengths than the blue-cone-deficient fovea (Weale,1953; Gordon and Abramov,1977). However blue is still far less clearly recognised than red, possibly because of the sensitivity of the achromatic rods at these wavelengths contributing to the signal (Gordon and Abramov,1977). Langran (1984) claims that orange is the most easily seen colour in peripheral vision, but apparently miscites the source of her information. Thus red and, to some extent, blue stimuli (the 'saturated end-spectral colours') appear to be more robust to peripheral degradation.

In summary, it would seem that for a line colour to be conspicuous in a given spatial configuration, it must
1) be clearly visible against its immediate background, which is largely a function of brightness contrast,
2) be distinguishable from other line colours (chromaticity contrast), and
3) stand out from the image as a whole, by its colour intensity contrast with the adaptation level.

6.3 COLOURS AND VISUAL LEVELS

The role of colour in the perceptual organisation of displays and emergence of figures has been discussed above. However, it has often been suggested that different colours can be used to create a hierarchy of separate visual planes. Within cartography, the most common attempt to exploit this has been the usage of the spectral series of hues for hypsometric tints, demonstrated by Karl Peucker in 1898, on the grounds of a stereoscopic effect caused by the differential refraction of light, which would cause long wavelengths (reds) to appear nearer to the eye than (or 'above') short wavelengths (blues) (Imhof,1982). The principle was later commended in some popular general cartographic texts (e.g. Eckert,1925; Raisz,1962). Peucker also suggested that lighter and more saturated colours advance towards the eye. Robinson et al.(1984) mention that differences in lightness can be used to create or strengthen a 3-dimensional effect (see section 4.5), while hue and chroma also carry some connotations of depth.

6.3.1 The Advance/Retreat of Hues

Because of the spectral series, and their contrasting views about its efficacy, cartographers have for long been interested in the apparent advance/retreat of hues. It has been suggested that hue might be responsible for this effect because of chromatic aberration in the eye lens. This causes short wavelength light to be brought into focus closer to the lens than long wavelengths, in the same way that more distant objects are focussed in front of nearer ones (Evans,1948). Alternatively there may be a disparity in alignment between the visual and optical axes of the eye, such that chromatic dispersion occurs and short wavelengths are imaged more nasally than long ones (Walraven,1984). Thus in conditions when the eye cannot adequately compensate for these disparities, the effect might occur either spatially or through hue-related differences in focussing quality.

The spatial effect is often referred to as colour stereoscopy, or chromostereopsis (Walraven,1984). The theory is that in accommodating the above disparity red colours appear nearer and larger while blues are pushed back and appear smaller (Saunders,1961). Walraven has however also pointed out that some people often see blue in front of red or no effect at all, and suggested that this may be due to variations in the directional sensitivity of people's photoreceptors. In fact it seems more likely that the effect occurs in particular conditions when the accommodating power of the visual system is inadequate to remove the disparity. For example, spectacle wearers viewing a stimulus composed of alternating red and blue strips, may experience an interesting effect. When viewed without spectacles the red strips may appear to be considerably longer (at each end) than the blue ones, but the illusion disappears when the glasses are put on and the whole image is brought into better focus. This would imply that the severity of any effects could vary immensely between people because of the individuality of optical systems.

Chromatic aberration, by causing defocussing and consequently blurring of short wavelengths, may be partially responsible for poor acuity in blue/violet (Myers,1967). It has consequently been suggested that colours used for the definition of detail near the threshold of visibility (e.g. fine lines) should preferably be

1) of the longer wavelengths (green to red), and
2) monochromatic (i.e. pure spectral hues combining fewer wavelengths).

This applies both for the object and its background, and for both lights (Luckiesh,1944) and surface colours (Robinson et al.,1978). Thus brown, a colour of low purity which has become used conventionally for contour lines on maps, is said to be ill-suited to this role. It is possible that differential focussing might be able to create visual planes cognitively through the learned association of features nearer the eye appearing clearer.

In practice these effects certainly are slight, occurring mainly with contiguous, spectrally pure colours on dark backgrounds, where the accommodating powers of the eye are particularly tested. It has been claimed that advance/retreat has never been found on a map surface to the extent of spatially separating visual planes (Tobler,1957; Hopkin and Taylor,1979), although Saunders (1961) disagrees. M.Wood (1968) has suggested that any effect is neutralised by the perception of paper as a flat surface. According to Imhof (1982), Peucker's ideas about stereoscopy were based upon a misunderstanding of the work of Einthoven, and any apparent effect is in fact a psychological illusion caused by the failure of depth perception in parts of the visual field which lack clearly identifiable points. Thus red images appear to stand out from photographic transparencies because their depth (distance from the eye) is resolved correctly while the spatially indeterminate background is not and retreats. Similar effects can occur in snow and apply to <u>all</u> hues. However, chromostereopsis can be apparent on self-luminous colour electronic displays, where high saturations and large brightness contrasts are used, and on other projected maps with dark backgrounds (R.M.Taylor,1985). Taylor agrees with Imhof's conclusion that these planes are separated by perceptual organisation rather than spatial segregation.

6.3.2 Colour, Depth and Apparent Size

Cartographers' disagreements about the desirability of the spectral series indicate that hue itself does not produce a clear vertical ordering of colours. Proponents of the spectral scheme chiefly cite its harmony and beauty (e.g. Gaussen,1964), while opponents have dismissed it as physiologically inconsistent because of the lightness reversal(s) inherent in it (Robinson,1967; Cuff,1972a). Experimental evidence of the existence of the varying depth of colours and its causes is mixed, and difficult to interpret because of the small numbers and often imprecise definitions of the colours used. Tests have used measures either of how distant a coloured surface appears to be, or more commonly its apparent size at a fixed distance as a surrogate of depth. Where more than one colour sample was viewed at once, the samples were invariably non-contiguous.

Early interest emerged from central Europe, with the observations of workers such as Brücke, Belajew-Exemplarsky and Matthaei (reported by Katz,1935) that

1) reds, oranges and yellows appeared to rise up from a variegated surface while blues receded,

2) colours generally appeared nearer the eye than achromatic stimuli, and

3) such effects were weakened if the object and background were made equally bright or overall illumination was reduced.

Other investigators began to look at the connection between colour and apparent size. A study by Warden and Flynn (1926), requiring the ranking of eight differently coloured cuboids in various configurations, failed to find a consistent relationship. However Gundlach and Macoubrey (1931) replicated the experiment using a grey background and obtained a positive correlation between the luminous reflectance of a colour and its apparent size of 0.86. A subsequent test by Wallis (1935) produced a ranking in apparent size of yellow (largest), white, red, green, blue, and black, implying a broadly lightness-based progression. It is notable that Wallis rated red higher than an earlier less rigorous experiment which had used a 'very dull' red. Bevan and Dukes' (1953) experiments involved correlating the apparent size of coloured cards viewed outdoors from a fixed distance with the size of one of a series of neutral grey cards given to the subject. They found that red and yellow cards seemed to be significantly larger than blue and green ones. Given the small sample (4 colours) the results were difficult to interpret confidently. However they concluded that neither hue, lightness nor saturation would individually be able to make colours advance and retreat, but that the combination of dimensions represented by Katz's 'Eindringlichkeit' (insistence) could be responsible.

R.L.Williams' (1956) experiments are particularly interesting as they involved cartographic point symbols of a variety of shapes and colours on paper. The 'colour factors' he derived for the area of a symbol judged to be equivalent in size to a black of area 1 are tabulated overleaf.

Red	(7.5R 5.5/14)	0.994
Black	-	1.0
Brown	(10YR 4/2)	1.005
Orange	(2.5YR 6.5/12)	1.01
Blue	(5B 5.5/8)	1.011
Green	(10GY 6/10)	1.026
Yellow	(5Y 8.5/10)	1.056

(Munsell specifications have been computed from the CIE figures quoted by Williams.) However, because of the narrow range of values, he considered that only three separate levels- yellow, green, and all the rest- were practically usable. The ranking however implies that the more highly-contrasting symbols (darker and more saturated colours) appeared to be larger.

Three further experiments used Howard and Dohlman depth perception apparatus. This consists of a stationary pole and a moveable pole, one of which is covered with a coloured paper and the other with a paper of a reference grey. The two poles are set against a grey background. The subject is asked to move the mobile pole so that it is perceived to be equally distant as the stationary pole. In the first of these experiments (I.L.Taylor and Sumner,1945), with grey on the stationary pole, a 0.99 correlation was obtained between the reflectances of the test colours and average pole displacement, the ranking of colours being yellow (apparently nearest), white, green, red, black and blue (apparently furthest away). The experiment was later replicated by Johns and Sumner (1948), but with grey on the mobile pole. The ranking obtained was (near to far) red, white, yellow, green, blue, and black, which was interpreted as a lightness progression as, with the exception of red, this was the descending order of the remeasured Y tristimulus values.

The interpretation given in either study was that bright colours appear nearer because the intensity of sensory excitation causes them to be brought into clearer focus, and because the eye has a focal association with perspective, a view shared by Ittelson and Kilpatrick (1951) and J.E.Williams (1953). Other aspects of stimulus intensity may also be important- e.g. chroma, which would explain the otherwise abnormally high ranking of red, which had the highest measured chroma. Another interesting point is that in the Taylor and Sumner study, the movable pole was placed behind the reference pole even when it was of a darker colour. The apparent nearness of the mobile pole was, they claimed, a function of its greater vividness to the subject, as it was receiving most of his attention. Thus foveal elaboration of the stimulus clearly promoted its apparent nearness. Such a fixation might be induced by the stimulus through figural emergence and/or an association of its strength of contrast with perspective. In this respect, they too cited 'Eindringlichkeit der Farben', as a potential link between the relative attention-getting properties of colours and depth perception. Katz (1935), too, reports other studies where the more 'insistent' colour of a pair stood out in front.

In the last of the three experiments, Edwards (1955), using larger numbers of subjects, failed to find any consistent or statistically significant evidence that colour has the quality of depth. He concluded that if an effect is seen, it is probably learned through colour associations and artistic teaching. However, his experimental set-up did differ from the previous studies in terms of illumination and, most significantly, viewing distance (20 feet as opposed to 7' 6"), such that the colour samples subtended a visual angle of only 0.54 x 0.36 degrees. Studies reported by Katz (1935) indicate that quite strong advance/retreat effects (red up to 1cm in front of blue at 80cm distance) disappeared under conditions of stimulus degradation (effective desaturation and loss of 'insistence'), so it is perhaps not surprising that no significant effects were found in this case.

From these studies it is clear that at least for isolated colours against a mid-lightness neutral background, lightness has a dominant role to play in depth perception, with higher lightnesses emerging 'on top'. This is consistent with the phenomenon of irradiation, whereby white figures on a dark background can appear to be up to 20% larger than black figures of the same size on a white background (Arnheim,1954). Chroma may also be significant, especially given the high ratings of bright reds in these studies. Interestingly, the particular spectral series demonstrated by Peucker is one of consistently declining chroma (Imhof,1982). Two related phenomena seem to summarise the apparent relationship between colour and depth perception-

1) association with everyday perspective cues, and
2) 'insistence', which may be the key to an important link between colour intensity, attention and perceived depth.

6.3.3 Applications of Depth Cueing

The nature and graphical application of perspective is of particular importance to artists. It has been suggested that 'aerial perspective' operates chromatically, as 'warm' (long wavelength) colours are diffused or filtered out more by the haze in the atmosphere, so that distant objects appear bluer and by experience redder objects therefore appear to be nearer. Raisz (1962) uses this as a justification for the spectral series. However, artists have found that much stronger perceptual gradients can be created by a more conventional application of aerial perspective, first described by Leonardo da Vinci, where objects are made paler with increasing distance from the observer (Arnheim,1954).

This is said to correlate with everyday visual experience, where distant features appear paler and more blurred, hence the use of hazy backgrounds in many landscape paintings. In practice, according to Clark (1985), distance and haze do not actually reduce the clarity of definition of objects, but merely their perceived contrast. The visual system consequently interprets less contrasting features as being more distant. Bertin (1983a), in his discussion of the 'retinal variables' defining depth perception, also considers that a reduction in the lightness contrast and/or saturation of a given object causes it to recede. Thus strong contrasts- extremes of lightness (white and black) and high saturations (pure hues and deep shades) appear nearer the eye, tending to grey, effectively the average colour of the picture, in the distance (Birren,1969).

Obviously, however, paintings involve much more complex and varied 'backgrounds' than printed road maps and can incorporate much more subtle contrasts. Also, given the large proportion of the surface of a road map covered by the background, an achromatic average may be of little relevance where the background is coloured, and chromatic adaptation effects may become important. Imhof (1982) has however noted the significance of aerial perspective in layer colour series on maps, claiming that it is the only phenomenon that can create advance/retreat. Consequently the most effective layer colouring schemes are based on 'the higher the lighter', with the lighter tints appearing nearer the eye and allowing for strong contrasts to be introduced through hill shading. The principle of the advance of light colours is however of no use for coding lines on conventional road maps with positive contrast (i.e. dark networks on light backgrounds). Other work (reported by Crawford,1971) with dark lettering on light backgrounds has also found that the strongest contrasts (i.e. <u>darker</u> symbols in this case) were perceived to be closest to the eye. This would imply that as the perceived darkness of a background decreases, there is a crossover point where darker figures begin to appear to advance.

Actual examples of the successful use of discrete visual planes in map design are occasionally documented. Saunders (1961) mentioned some research at the University of Adelaide which demonstrated the superior communicating ability of maps with visual planes, and suggested that the use of a saturated hue with white and black would create these planes better than using red and blue on the same map. Bartholomew and Kinniburgh (1973) incorporated three distinct levels into the design of the Bartholomew Edinburgh City Plan. The top level was for street names, which were considered to be the most important information on the map. Consequently they were printed in bold black type over unfilled (white) streets for maximum contrast. The rest of the background, in recessed greys and greens, formed the lowest level, from which major public buildings were 'raised' by outlining (sharper edge contrast). The use of solid yellow for bus routes also lifted them from the screened-back and weakly chromatic background, while still being light enough to be overprinted clearly by names. Andrew Holmes' London Transport bus map used very similar design techniques, including for bus routes 'broad, bright yellow-orange bands, which have the effect of lifting the bus network from the detailed road map in the background.' (Braidwood,1981,p.54)

So it can be seen that certain colour-induced depth cues can be successfully applied on printed maps, although these cases appear to rely upon psychological effects of perceptual (figure-ground) organisation more than any physiological influence of hue.

6.4 Magnitude Implications of Colours

It would seem logical to expect that the visual significance of a feature would increase with its size and its relative visibility, or the intensity of perceptual stimulation it induces. Other things being equal, a larger object will carry more weight in a composition (Arnheim,1954). However, the precise relationship of aspects of stimulus intensity to the inferred magnitude of features symbolised by coloured lines has apparently not been previously studied.

The cartographic literature tends to rely upon heuristics or 'rules of thumb' for general principles concerning colour use (R.M.Taylor,1984), and those periodically redefined by Robinson (1952,1967 and 1978) have remained essentially unchanged. With regard to the magnitude implication of marks on white paper maps, Robinson elaborated upon Eckert's (1925) principle that the darkest and strongest colours were the most significant. Thus 'the fundamental rule is that the darker the value, the greater the magnitude' (Robinson and Sale,1969,p.262-3), and although its overall contribution is somewhat less, 'conventionally and subjectively, the greater the intensity [i.e. chroma], the greater the magnitude implication.' (Robinson et al,1978,p.314) They continue: 'A significant physiologically based aspect of chroma is that the larger the map area of a colour, the more intense it will appear.' Thus large patches of 'darker and more intense colours' inherently carry the greatest magnitude implications, representing the greatest 'amounts' of ink/colour on the paper. As Cuff (1972a,p.112) noted, 'perceiving a sequence of magnitudes in a colour scheme is a matter of perceiving an analogy.' Saunders (1961,p.5) stated it thus: 'Many people can not distinguish between intellectual and visual stimuli. For example, darkness and intensity [i.e. chroma] are definitely assumed to represent 'more of' something and lightness and tints 'less', and failure to observe this correlation may well result in confusion.'

Cuff (1972a) investigated quantitative sequences on choropleth maps, whose areal units were of broadly similar size, and attempted to find colour schemes which 'give the same impression of greater-lesser magnitude to most map readers without a key or legend to explain them.'(p.6) He produced experimental maps which he tested on a wide range of schoolchildren. His results confirmed the association of greater magnitudes with darker, more chromatic colours, and suggested that chroma was more significant than was previously thought. These associations proved much stronger than any qualitative associations with particular hues or hue sequences such as the spectrum. Both (Munsell) value and chroma changes were individually able to convey magnitude messages, and although the eye is somewhat more sensitive to changes in value, they cannot be used to override chroma reversals in a series (Cuff,1972b). However, this might not be as critical when dealing with road networks, whose colours are for the most part not seen in contiguity.

Cuff (1972a) also reported unpublished work by Bartz-Petchenik which found that children do associate darker colours with 'more' and lighter colours with 'less'. Different hues definitely implied a change in <u>type</u> and not <u>amount</u>, and hue associations were not strong enough to reliably carry a quantitative message. Olson's (1981) tests on the US Bureau of the Census 'two-variable' maps, which used a wide variety of colour mixtures, confirmed that hue was not perceived quantitatively. When asked to arrange colour samples for the 16 classes (four levels on each variable) into a sensible scheme, all 27 subjects chose different arrangements, many using some vague colour associations.

6.4.1 Colour and Visual Weight

In addition to the aforementioned experiments on the effects of colour upon apparent size, various investigations have been made into its influence on the perceived weight of objects.

These experiments, generally conceived out of work on aesthetics, have involved either visual judgments of the weight of coloured blocks or discs, or have simulated weighing by determination of the 'balance point' between two stimuli. The above-mentioned preliminary study by Warden and Flynn (1926), whose interests related to the display of retail goods, also involved assessment of the cuboids' weight. They found that weight judgments varied more consistently with colour than size judgments, and that there was no correlation between the two. The overall ranking obtained was black (heaviest), red, purple, grey, blue, green, yellow and white (lightest), although the steps between red and blue were quite small.

In general, early studies had shown that red and blue were the 'heaviest' colours, and that yellow was the 'lightest'. However, these results were criticised for their lack of statistical evaluation and the possibility that they were confounding the effects of chromatic variables and brightness (Pinkerton and Humphrey,1974). Warden and Flynn's rankings would appear to form a lightness series, although they omit the appropriate specifications. The significance of lightness was however clearly borne out in work by Payne (1958), using coloured cubes against a white background, who obtained a consistent ranking of red (heaviest), blue, red-purple, yellow, turquoise, and green, with a 0.94 rank correlation with measured luminous reflectance, the darkest being the heaviest. Subjects were also asked to assess the colours used aesthetically, and interestingly enough there was no significant correlation between their colour preferences and their judgements of apparent weight.

In the German-language literature, the observations of Eckert, Imhof and Arnberger (reported in Frenzel,1965) confirm the association of weight (Farbgewicht) with darkness, and to a lesser extent, colour 'strength'. Arnberger has from his own (unreferenced) investigations, derived a curve for the visual weight of pure spectral colours on paper maps, in order to achieve a more reliable specification for the quantitative use of colour (Arnberger,1966). The curve (figure 6.1) represents an almost perfect inverse of luminosity- i.e. darker colours are heavier. Blue, green and red are very similar in weight, and violet comes out strongly, as it often appears to in the German-language literature. Arnberger suggests an overall ranking of black, blue-violet, red-violet, blue, green and red, orange, yellow, white. Dark, non-spectral colours are more difficult to place precisely- brown can be either heavier or lighter than red/green. Less chromatic colours also carry less weight.

Overall it would certainly appear that on a white background the darkness of objects carries the connotations of 'more of' mentioned above, while chroma has a similar but less powerful effect. Cuff (1972a) however was concerned at what seemed to be contradictions between these results and the findings on apparent size where lighter colours generally appeared to be larger. However, while the size studies were looking at colours as the cause of actual sensory differences in apparent size, it was made clear to the subjects in some of the weight studies that the test objects were in fact the same size. Consequently any perceived differences were by association with density, which is clearly of relevance to the magnitude implication of a colour patch of a given size.

A further important difference is that while the backgrounds utilised in many of the 'size' experiments were of mid-lightness (grey), in Payne's work on 'weight' reported below they were white. This would clearly affect the relative 'pronouncedness' of differently coloured objects set against them. In fact the experience of colour weight by artists, whose compositions generally have an 'average' lightness near the middle of the scale, backs up the 'size' rankings, with red appearing heavier than blue, and bright colours heavier than dark ones, perhaps partially because of irradiation (Arnheim,1954).

6.4.2 Learned Associations?

Other studies have controlled brightness and saturation and still found consistent effects of colour upon visual weight. For example, Payne (1961) used Munsell papers of equal chroma (except grey) and value in order to attempt to assess the role of hue. Overall red, blue and purple (the end-spectral colours) were perceived to be significantly heavier than yellow, green and grey (mid-spectral or neutral), although the differences were much less marked than those due to brightness. Pinkerton and Humphrey's own experiment investigated chromaticity and brightness separately using comparisons against a white of constant brightness. They found no consistent differences in weight with brightness, but as they were using transilluminated stimuli, this does not mean that the lightness of surface colours has no effect. At constant brightness, all the colours used were regarded to be heavier than the standard, with yellow significantly lighter than the rest (blue, green, orange and red), and red significantly heavier than all the others except blue. The responses of men and women were essentially similar.

The reasons for this effect of colour are a mystery- 'No plausible explanation has yet been offered for why people should see any equivalence between colour and weight, nor can we offer one.'(Pinkerton and Humphrey,1974,p.165) They suggest that the two most likely sources of this covariation are colour preferences and indirect associations of hues. However, their results do not entirely correspond with findings on colour preferences presented in section 6.5 below, and the link between weight and aesthetics has already been refuted by Payne. Payne himself suggested that a colour-weight illusion might be derived from a colour-size illusion such that objects appeared heavier because they seemed to be larger, but there are notable differences between Payne's ranking and that of the most comparable colour/size experiment (R.L.Williams,1956). Connotations of hues are more likely to be important. For example red, generally considered to be a striking colour, is often used to denote importance, which implies heaviness. This convention is also established in cartography (Robinson et al,1978).

These results are particularly interesting in their implication that colour associations derived from general experience can create perceptual differences or magnify small existing ones. The colours which seemed to be heavier were no darker or more saturated than the others, but were of hues which normally would be so but for the enforced artificiality of the situation. For example, reds are normally darker and more saturated than yellows. In fact many of the magnitude-related colour schemes which occur other than on maps, such as the points values of snooker balls, rely on learned associations for their comprehension. There is no consistent lightness progression through the colours, as viewers of snooker on monochrome televisions are well aware. Other schemes use a direct visual analogy, such as the colour coding of judo belts, where the developing competence of the performer is marked by a series of belt colours of increasing contrast with the white garment. The progression is from white through yellow, orange, green, blue and brown to black, with red or red-and-white for the most exceptional performers on ceremonial occasions (Reay and Hobbs,1979). The situation is however clearly less exacting than in snooker where the colours may all be seen, and require to be distinguished, together.

6.5 The Affective Value of Colours

The connection between the beauty of a map, in terms of an aesthetically pleasing design and a harmonious use of colour, and the clarity of its communication has been noted (Makowski,1967; Imhof,1982). This link has however rarely been demonstrated in tests of mapreader performance (Langran,1984), although perceived pleasantness has been cited as the cause of the enhanced legibility of certain colour combinations (Vickerstaff and Woolvin,1944). A preferred colour may be attended to preferentially (Woodworth and Schlosberg (1955) named emotional appeal as an internal determiner of attention) or carry a greater quantitative implication, especially for children (Cuff,1972a). In a colour series too it is possible that the use of too many widespread hues could create disharmony and hinder the perception of hierarchy.

Clearly colour harmony and preferences are subjective, contextual and complex considerations, which are poorly understood with respect to their psychological causes and point of entry into the eye/brain system. However, artists in particular have developed some ground rules as to which colour combinations and compositions are generally likely to appear harmonious. One such is that rich, saturated colours are preferred for small areas in an image (such as lines) to tints (Graves,1952; Saunders,1961; Imhof,1982). It has also been noted for a long time (cf.Katz,1935; Birren,1969) that certain human biological and psychological reactions to colour vary with hue from red (exciting, stimulating, active) to blue (relaxing, calming, passive). Red light apparently raises both the pulse rate and blood pressure (Graves,1952); blue has the opposite effect. According to Arnheim (1954), strong brightness, high saturation and long wavelengths all produce excitement in a manner too direct and spontaneous to be the product of learned association. Conversely, Imhof (1982) has argued that no colour in itself is either beautiful or ugly, and human reactions to colours are entirely dependent upon their contextual meaning (e.g. red roses and blood).

Writers from several disciplines have searched for consistent variations in the affective value of individual colours. Unfortunately many of the studies again attempt a ranking of the major hues using poorly specified colour terms (e.g.'blue','purple'), with lightness or chroma rarely even mentioned. Thus apparently contradictory results may simply be caused by the confounding of variables. McManus et al.(1981) also found the literature on colour preferences to be 'bewildering, confused and contradictory' (p.651) and generally worthless. A summary of the more rigorous investigations (several have used Munsell specifications) would suggest that, for Britain and North America at least, red (by women) or blue (by men) are generally considered to be the most pleasant fully-saturated hues; green and purple are intermediate, and yellow and yellow-green are relatively disliked (Cuff,1972a; McManus et al,1981). Saunders (1961) has suggested that this represents a lightness-based series, the preferred colours contrasting most with the paper.

The study by Guilford and Smith (1959) was particularly thorough. From an experiment involving 40 observers (20 men and 20 women), who each rated 316 Munsell colour specimens on a 10-point scale, they were able to derive isohedonic charts with a correlation between observed and expected values of 0.93 for men and

0.88 for women. Guilford and Smith were struck by the considerable consistency shown in what was generally considered to be a very subjective matter of taste. Their findings for hues fit in well with the relationship suggested above. They also showed that, with few exceptions, affective values were positively correlated with lightness and chroma, with the precise shape of these relationships differing for each hue. Affective value would seem to be one way in which lightness and chroma levels (not just contrasts) and even hue itself have some inherent significance.

The work of McManus et al.(1981) involved 54 subjects (27 male, 27 female) who performed paired comparisons of colour samples. They found considerable consistency of judgments, especially at higher chromas, and generally agreed with Guilford and Smith about preferred hues, lightnesses and chromas. With respect to hue preferences, however, subjects could be categorised into four types. The largest group (30 subjects) responded similarly to the overall average- preferring blue and disliking red (except at high chromas) and yellow. However, a separate and predominantly female group responded in a near opposite manner, preferring red and yellow, while a third (predominantly male) group preferred blue and yellow to red, purple and green. The remaining four subjects were relatively idiosyncratic. The overall picture is thus one of far less variation between <u>individuals</u> than might be expected.

Preferences for colour combinations are certainly even more complex, and predictive equations on the basis of Munsell specifications have been less successful. Some general patterns do however emerge: small and large hue differences are normally preferred, especially by women, to medium ones. (For large hue differences, an absolute maximum of 4 hues (90° separation) can be used.) Men in particular tend to prefer large lightness differences, while a slight leaning towards small chroma differences has also been detected (Allen and Guilford,1935). It is interesting that in A.Morrison's (1974) test of speed maps, the preferred scheme (section 2.4.2) involved, unusually for a road map, gradual changes of hue, lightness and chroma, and induced slightly better performance than schemes with more contrasting colours. It is possible that with the unusual locations of class changes on the speed map, colour harmony becomes more perceptually important. On occasion, however, it may be desirable to contravene these preferences deliberately and use disharmony to create centres of attention (Arnheim,1954). This is a possible way of highlighting one particular class, but at the cost of a smooth perceptual gradation of the whole series.

6.6 Salient Colours

Perhaps an investigation of those colours known by experience to be salient would be useful. The special role of red for attention-getting is well-established both in light (for warning lights, danger signals etc.) and surface colour. On maps red seems to form the highest visual plane of non-fluorescent colours (M.Wood,1968) and is often used on road maps for the major road network or as a sparingly-used code for special information emphasis (e.g. RAC Navigator atlases). An overview of the wide range of tests of performance and judgement mentioned above (sections 6.1-6.5) reveals that red colours (and particularly so-called 'bright reds') were normally the most salient, in terms of attention-getting, colour recognition, advance, magnitude implication and possibly affectation, often disrupting otherwise good lightness-based progressions. There is evidence too that red is the first chromatic colour to be named and distinguished not only by primitive languages but also by children (Berlin and Kay,1969). This significance of red may be partially derived from its infrequent and highly contrasting occurences in nature. However the clear perceptual advantage it seems to have over other colours is that it can be (and normally is) reproduced at high saturations, and for surface colours, high chromas at medium lightnesses, contrasting well with white paper.

A group of colours that are evidently very conspicuous are those which display considerable fluorescence. They fool conventional spectrophotometers and lie outside defined colour spaces. This is because ultraviolet light incident upon them is reradiated at longer (visible) wavelengths, leading to much greater brightnesses- they appear to reflect more light than they receive (Boynton,1978). Consequently they resemble shining lights more than reflecting surfaces. This brightness boost occurs in the area of maximum reflectance of visible rays (dominant wavelength) so that hue is maintained and chroma is also boosted (Chaloogin,1967). The relative conspicuity of different fluorescent colours can change with varying amounts of UV yield (i.e. the brightness boost), and strong advance/retreat effects are noticeable (R.M.Taylor,1981).

Fluorescent colours are used a great deal in life for attention-getting- e.g. for highlighting text, and on posters, safety equipment, and small objects such as marker buoys which need to be picked out against monotonous backgrounds. They are also especially useful in negative contrast surface colour displays, such as the information boards used at some major railway stations and athletics meetings. However, until recently the use of fluorescence on maps has mainly been restricted to complete military maps for nighttime use under UV light (Crandall,1973), and in whitening paper. However, their potential use for highlighting linear information under daylight conditions is considerable, as daylight includes plenty of UV radiation. In Taylor's (1981) experiment using fluorescent inks for coding linear (and other) information, extra brightness was perceived under tungsten illumination as well. Royal Australian Survey Corps photomaps use fluorescent colours to make lines emerge from the dark background. In perceptual terms, the energy boost given may be as significant for the extra chroma (Chaloogin,1967), accentuating existing colour contrasts, as the additional brightness, especially on white paper.

Thus both the experimental and experiential evidence presented so far points to some combination of lightness and chroma contrast (i.e. colour intensity contrast) from the stimulus adaptation level as the fundamental determinant of the saliency (or 'insistence') of surface colours. Inherent brightness clearly becomes increasingly important in conditions of apparent or physical fluorescence. However, from the five categories mentioned at the start of the chapter, there remains one which may not operate in a correlated way.

6.7 Colour Associations and Conventions

Connotations not directly related to any visual properties of colour may also have a strong influence on the ordered perception of line symbol series on road maps. These connotations may derive from

1) association with concepts related to magnitude or hierarchy

2) learned association with the symbolised objects themselves (roads), and particular classes of them.

A certain amount of work on the conceptual associations of colour has been undertaken in Germany using scaling techniques based on semantic differentials. Wright and Rainwater (1962) established a connection between forcefulness, darkness and chroma. Work by

Schiede (reported in Frenzel,1965) revealed the following associations:

Heavy	: violet	Light	: yellow
Elevated	: purple	Submissive	: yellow, green
Prominent	: purple	Inferior	: yellow, green
Active	: red	Passive	: green

The causes of such associations are not at all clear. They may well vary between cultures- purple for example is considered to be particularly prominent in the German literature as a whole. In Britain, associations of colours with size were investigated by Poulton (1975), who asked people to create a 6-step series. Over half of those questioned chose red for the largest size, while white was the most popular for the smallest although it only attracted about 20 per cent of the vote. Coding systems in common use and natural sequences, such as the colour code for resistors and capacitors, based on a combination of the spectrum and the colour sequence of a heating-up furnace, were either not known or completely ignored, except by a handful of people.

No reports have been found of an association of colour with speed, but it is interesting that a colour coding convention has evolved for orienteering maps, which are effectively maps of likely travel speed judged by the vegetation and microtopography of an area. The convention is a two-hue series with white as the central, intermediate category (open woodland) separating various degrees of openness (yellows of increasing chroma) from various levels of traverse difficulty (greens of increasing chroma). The choice of colours however has little relevance to a linear situation: green is appropriate because of its association with vegetation, while yellow is excellent for areas as it is not too dominant and provides for good legibility of overprinted black detail. It does however provide an effective example of depicting 'more of' something by chroma.

Particular colour names (i.e. adequately-saturated hues) may also be associated with mapped features and consequently suitable for their representation. The use of green tints for vegetation is an obvious example of this so-called iconic colour coding. Van der Weiden and Ormeling (1972) studied people's colour name associations with particular phenomena, making no specific reference to maps. The majority of people they asked associated 'highway' with grey (clearly iconic), and 'transportation' with either grey, red or black. Interestingly they found no significant differences between males' and females' associations, or between those aged under/over 28 (which was expected from other studies). Unfortunately no more detailed questions were asked about associations with particular classes of highway.

Conventions for colour use, which may have no innate associations, also develop historically. Some have become so well-established that their interpretation is often simply assumed (e.g. blue areas for lakes or sea) irrespective of their specific visual characteristics. In certain circumstances therefore the connotative power of colour may clearly exceed its quantitative power. On maps, the convention of using 'bold' reds to depict items of special importance is very well established (Robinson et al.,1978; Imhof,1982), and with main roads it dates back at least to the 10th century Arab geographer Maqdisi, who used it so 'that the descriptions may be readily understood by everyone.' (Keates,1962,p.21) There is also evidence that the use of brown or red for roads had become an established convention in European mapping by the sixteenth century (Skelton,1964), and in 1807 the German geographer J.Schemerl suggested using red and yellow for the most important roads (Schiede,1968). As colour lithography developed in the late 19th century, red was often the first colour to be introduced onto road maps and was commonly used to highlight the best roads in contrast to the remaining black detail, while yellow gradually became associated with secondary roads (Nicholson,1983).

The situation has unfortunately been confused by new conventions which have emerged in specific cultural contexts and, in military mapping, through standardisation agreements. Over the last 20 years, almost all road maps of Britain have incorporated in some way at least the background colour of signposts for motorways (white on blue) and primary routes (yellow on green). These sign colours were not chosen for their attention-getting properties, but rather to provide a background against which light lettering would be reasonably legible <u>without</u> being visually intrusive in the landscape (Kinneir,1984). Most maps now depict motorways by a blue line symbol and primary routes are commonly shown in green. Red is consequently often relegated to the 'other A roads' category. However many map users may still not be aware of this symbolisation, especially

1) in the case of primary routes where the concept may also be unfamiliar, and

2) because of the familiarity of the sequence used on Ordnance Survey one inch/1:50,000 maps of red, brown/orange, yellow and white- a part-spectral and increasing lightness progression with good perceptual order. R.M.Taylor (1985) mentions that the introduction by the Ordnance Survey of blue motorways at the top of this series has created potential confusion between motorways and hydrology and a visual discontinuity. He suggests that colour differences are more important to map design than the particular colours employed and that conventional hues should not be used if they disturb 'the integrity of the structural design logic.' (R.M.Taylor,1985,p.196)

There is clearly a trade-off here in terms of task performance, because while conventional colours may facilitate navigation by aiding map/ground correlation, disturbance of the graphical hierarchy is likely to reduce the efficiency of route planning. Clearly the lack of standardisation of colour usage on road maps can cause learning difficulties to users encountering more than one type of map. Overall it is difficult to assess the areas of influence and significance of colour associations and conventions on the perception of coloured line symbols, because no studies are known which have examined these effects and their variation between people and in different contexts.

6.8 Coloured Lines on CRT Displays

The use of colour to organise map images on cathode ray tube (CRT) displays is an area where very few sound principles exist (Robinson et al,1984). Very little specifically cartographic research has as yet been undertaken, while the human factors literature offers few relevant guidelines (Dobson,1983b). In general terms, however, CRT maps differ fundamentally from printed maps in three main ways which will be considered in turn:

1) the greater amounts and contrasts of luminous energy involved

2) the limitations of the display medium and display devices

3) the common usage of a negative contrast format.

1) At 'normal' illumination levels, the adaptation and accommodation mechanisms of the eye are clearly stretched far more by the range of contrasts on self-luminous displays than on the reflecting surfaces of paper maps. The fatigue and after-images caused by the greater brightnesses of self-luminous displays are clear evidence of this. CRT images do not possess the qualities of lightness and chroma, but they do have measurable brightness and saturation. It would seem that (in broad

equivalence to the situation for surface colours) symbol legibility on these displays is basically a function of size and colour contrast (Dobson,1983a), with brightness contrast being the most important contributor to this (Menu and Santucci,1982). However because of its actual luminance, the relative brightness of a coloured symbol is likely to be less degraded by background colour than in print (Bruce and Foster,1982). Other effects such as irradiation and chromostereopsis may also become more important. Creation of figure-ground by chromaticity differences has also been recommended in order to offset the dominating effect of brightness (Huddleston,1982).

2) These considerations all have to be related to the capabilities of the display devices used. For example, linear degradation is less of a problem because with the limited resolution of displays, fine lines are not generally realisable. A typical high-resolution display has of the order of 3-4 pixels/mm horizontally. Finer lines can however be produced by illusion using very low contrasts on displays with a fine control over electron beam intensity levels (voltages). At 8 bits/pixel monochrome (i.e. 256 grey levels), smooth lines with an apparent gauge of 1/8 pixel width can be drawn (Forrest,1985), but obviously at the expense of any control over their contrast level. Misconvergence of the red, green and blue beams on a coloured display can also blur edges and fine detail and effectively desaturate colours.

The range of colour available, and consequently the subtlety with which it can be applied, is entirely device-dependent: the most basic machines only generate 7 colours (on/off only for each gun plus combinations), which are not easily ordered (McGranaghan,1985), whereas a pallette of 4096 colours is not uncommon on current personal computers. Colour specifications chosen from a relevant uniform colour space such as CIELUV can be translated into gun intensities by device modelling (Robertson and O'Callaghan,1986). The gamuts available on a CRT monitor are however limited in brightness and saturation by the phosphors involved- the theoretical relative luminances being green 59, red 30 and blue 11 (white=100 by additive mixing). Thus the brightest pure colours are in the mid-spectral region of high luminosity. There is also limited evidence that greens and yellows are the most accurately identified colours on CRTs (Dobson,1983a). Given the limited saturations attainable, and the additivity of luminance by mixing, brightness contrasts may appear to be perceptually dominant, especially on cursive displays with no area filling. In fact the contribution of saturation to prominence may well be much weaker than in print where the perceived 'dirtyness' or greyishness of less saturated colours presents a problem. The blue primary comes out very poorly, and it is one of the few consistently repeated recommendations in the literature that short wavelength colours should be avoided on CRTs, especially for small symbols on dark backgrounds, because their low brightness contrast and acuity and accommodation problems have been shown to impair performance on certain tasks (Huddleston,1982; Silverstein,1982; Dobson,1983a). A further problem with phosphors is that they deteriorate, so that displays gradually lose their brightness.

3) Given the nature of road maps, it is not surprising that those which have been produced on CRT displays- such as the Etak navigator, CARTRIPS (Robb,1985) and the CARIN prototype (Royce,1986)- are in cursive formats and, apart from CARIN, in negative contrast. R.M.Taylor considers that it is difficult to provide good perceptual organisation and a strong visual hierarchy with this format, because 'only relatively small areas of colour are visible, they are usually uniformly bright and highly saturated so that the hues can be identified, and contrasts cannot be varied by background differences.'

(R.M.Taylor,1985,p.199) Extra weight is given to these contrasts by the fact that on self-luminous displays, saturation increases with brightness (as opposed to darkness with surface colours). While on many displays the lines need not be uniformly bright, the lack of effective control over line width presents a further problem. The apparent widths of differently coloured lines will however vary considerably in such circumstances owing to irradiation effects, which are strongest with such self-luminous images not 'toned down' by background colour, because of both the inherent intensity of the image and its strength of contrast and contour (Greenberg,1971). This can be demonstrated quite simply by rapid adjustment of the brightness control on a conventional television, and clearly adds to the conspicuity of bright lines. Problems of induction are also reduced by a dark neutral background.

A further problem area with negative contrast maps (including printed ones such as the town plans in Geographia's Travellers' Britain atlas) is the operation of colour associations. Robinson et al.(1984) consider than the association of dark=more, light=less would be reversed in such situations, but experiments by McGranaghan (1985) with choropleth maps on CRTs revealed that only about 30% of subjects changed to light=more judgments on a black background, and dark=more and particularly more saturated=more were used most commonly and most consistently. However, it is not yet possible to estimate how much this is due to the unfamiliarity of the CRT format, or whether there is some general dark=more association based on everyday visual experience which holds in most circumstances. The unconventional mode of appearance may also reduce the influence of colour conventions, especially as the conventional colours for the less important roads (yellow and white) have the highest brightness contrast.

6.9 Colour Series Design

In applying these findings to the design of colour series, the resolution of the tension between order and distinction becomes all important. Cuff (1975) referred to the 'conflicting goals' of making each class colour sufficiently distinct to be individually named and identified with reference to the key (or memory) whilst maintaining an unambiguous sense of quantitative change. Many schemes have been proposed and evaluated for area colours, either for layer tinting (e.g.Imhof,1982) or statistical mapping (e.g.Cuff,1972a). Of these the strongest order is conveyed by a 'tonal series' of increasing lightness and decreasing chroma within one hue (and thus a diagonal cut across colour space) such as would be produced by a standard lithographic tint progression (Cuff,1973; Langran,1984). An alternative single-hue series suggested by Bertin (1983a) is from black to white via any saturated mid-lightness hue, relying on the assumption of lightness conveying stronger order than chroma. With lines, sequences involving such subtle contrasts are unusable because of the problems of stimulus degradation, and because the colours are not directly juxtaposed, so that the steps are not perceptually widened by induction effects along boundaries. A slight compensation is that where the background to the line network is uniform, relative differences caused by induction in different contexts are less of a problem.

Obviously to increase the separation between each step a wider hue range is required. Cuff (1973) also investigated two-hue series for temperature maps, where the end points were pure hues associated with a relevant semantic differential (red=hot, blue=cold) which were progressively desaturated towards the middle of the series. It was however perceived more like two separate single-hue schemes, indicating that the ordered effect of the tonal sequences is far stronger than the indirect associations involved, and that darker=more is more relevant than darker=more extreme. With lines, the need to use high saturations means that several hues are generally involved, so there are clearly great difficulties in integrating them into a usable progression.

Series of more widely spaced hues have attracted less research- the aforementioned spectral sequence for layer tints has however been tested at a variety of saturation levels. The inherent lightness ordering is not random, but is similar to the two-hue situation in rising to a single peak in mid-series (yellow). J.K.Wright (1942), whilst generally recommending the use of tonal series, noted that the spectrum had been found by experiment and experience to give a graphic impression of relative altitudes, and this was confirmed by Patton and Crawford (1977) in an experiment involving a prominent key. However, Phillips (1982) compared spectral and single-hue tonal schemes with keys, imposing fairly strict time constraints on his subjects. He found that absolute height judgements were more accurate on the spectral maps (with more distinct classes) but relative height was perceived more clearly on the tonal maps. Before they were shown the maps subjects were also asked to place spectral colour samples in height order. As only 3 out of 99 chose a spectral order, Phillips concluded that the spectrum did not appear to be a natural progression. This confirmed the work of Cuff (1972a) whose spectral series, including a special increasing lightness/decreasing chroma spectrum, were not perceived quantitatively overall but only in parts. This implies that series perception can only be extended across a hue range which can be considered to be a related family of colours, beyond which differences appear to be qualitative (Cuff,1975). Bertin's (1983a) scheme for rearranging the fully-saturated spectral hues into the order of their inherent lightness would also fall foul of this.

Part spectral series may offer a partial solution. The two most obvious schemes link yellow with the other artists' primaries: red-orange-yellow (the 'warm side') and blue-green-yellow (the 'cool side'). In Cuff's (1972a) tests, these two schemes conveyed a better impression of magnitude gradation than any other series except single-hue ones. The range of hue is clearly restricted, but there is some evidence that small hue steps are considered to be more harmonious than medium ones (Allen and Guilford,1936). Newhall (1939) has suggested that over short ranges between unitary colours the hue scale is virtually quantitative, with intermediate hues perceived as mixes. Bertin (1983a) agrees, considering these to be the only ordered scales in chromaticity. However the main perceptual asset of these part-spectra is that they avoid the lightness reversal by not crossing yellow, and in fact both form natural lightness series even with fully saturated hues. A.Morrison (1971) considers red-orange-yellow, used on several British maps including the Ordnance Survey 1:50,000 series, to be the only part-spectrum which is in both conspicuity (for a light background) and lightness order, although a third sequence comprising the remainder of the hue circle (red-violet-light blue) is nearly acceptable. It would therefore seem that 120° of hue is about the maximum limit for a 'family of colours'. Use of the whole of this range, if suitably adjusted for chroma, plus achromatic extreme(s) provides a longer series such as black, purple, brown, red, orange, yellow, (white), which is often used on maps of population density.

It is clearly possible to design series as tracks through a uniform colour space in order to obtain perceptually smooth gradations. Lightness progression is obviously a vital factor in the design of any colour series, whatever variations of hue are employed. The type of scheme often recommended by cartographers of a systematic lightness scaling across a range of hues (Robinson et al.,1978; Phillips,1982) requires a careful balance between quantitative and qualitative distinctions. While lightness is perhaps the strongest ordering dimension, it is especially important to check the chroma progression, because of its most significant and undersung contribution to the perception of magnitude. A chroma reversal too can completely disrupt the perception of a series (Cuff,1972b). Shellswell's (1976) scheme however stresses equal qualitative intervals in the series. Standard process colour tints are related to L* u* v* space, and tracks with equal colour difference steps are effectively chosen across the uniform chromaticity plane. The minimum recommended step (20 CIELUV units) is sufficient to be consistently printable and to maintain the perceived individuality of each class given induction problems. At each step, the Y tristimulus value is merely checked to ensure that the series is moving in a consistent direction on the lightness scale, to enable visual ordering. Thus quantitative differences are not controlled.

Alternatively tracks of different shapes can be defined using all three dimensions of colour space. Curved tracks lack the equal-appearing intervals definable with straight tracks, but enable better colour space utilisation. A three-dimensional spiral in Munsell space was described by Graves (1952), and Robertson and O'Callaghan (1986) have defined spirals in L* a* b* and L* u* v* spaces by regular numerical incrementation of the hue, saturation and brightness parameters of the spaces. Thus explicit control is exercised over the perceptually significant variables. There are however two caveats to this: to preserve intuitive order and reasonable spacing the pitch of the curve should not be too small, and care must be taken to keep the spirals within the range of colours realisable by the relevant reproduction technique.

Designers of colour series for lines might well consider that range to be inadequate. It is extendable by the use of fluorescent colours, but under tungsten or sodium lighting, for example, fluorescence may be barely visible if at all. It may be considered preferable to use a set of colours with almost maximal qualitative differences and consistent, though inevitably small, quantitative differences. As long as the sequence includes no actual visual reversals, it should be relatively simple to comprehend and learn, in which case cognitive processes can effectively enhance the hierarchical perception in the medium term. Care must however be taken that the qualitative differences are relatively uniform, or else one class may appear to be so deviant that it is dissociated from the others: for example a blue motorway in a network of red and yellow roads might be interpreted as a river.

Clearly with the quantitative-qualitative trade-off there is no obvious single optimum approach to the design of colour series for lines, and various options need to be tested. In practice, where redundant width and/or casing variations can also be used, they may greatly assist the hierarchical perception of a series.

7. EXPERIMENTAL METHODOLOGY

Board (1975) has stressed that any objective map evaluation must bear in mind the purpose of the map, the type(s) of reader for whom it is intended and the conditions in which it is used. It is especially important that any experimental tasks used are realistic and appropriate to the stated function of the map if they are to assess how well this function is fulfilled (Board,1975; DeLucia,1976; Hopkin and Taylor,1979). Maps with a particularly well-defined purpose such as road maps are thus more easy to assess (A.Morrison,1975), especially where they have been designed with a specific task, such as fastest-route planning, and a specific population of users in mind.

Consequently it is important to consider the particular nature of the present research problem in order to derive an appropriate methodology. Firstly, it must be reiterated that the basic concern of this study is not what information should be included on the map but rather how it should be displayed, the objective being to improve the communicating abilities of the map by changing the map marks themselves rather than by educating map readers. The particular task involved is planning an efficient route, especially in terms of time minimisation, which at least plays a part in the vast majority of route choices. In reading the map to achieve this goal, various subsidiary tasks are undertaken, including for example the acquisition of stimuli from the visual periphery and the assessment of the prominence of those viewed foveally. The target population is all users or potential users of road maps, namely all drivers and non-driving navigators, in fact anybody who plans routes over U.K. roads. This includes some foreign motorists (A.Morrison,1975). Conditions of use have been discussed above (section 5.6), but use under amber street lighting, which would severely restrict the usable range of colours, is a minority occurrence too constraining to be generally considered.

This raises the problem of the number of variables relating to both subjects and map use conditions that are involved in reading the map. Many of these have been outlined in the model in section 3.1. The quality and quantity of illumination, the individual's visual acuity and colour vision, his criteria in route choice, cognitive strategy, general spatial ability, knowledge of the area and experience and familiarity with maps in general and the particular one in question, may all vary between different map use situations. People's familiarity with graphical conventions is a particular problem, as it is not really possible to simulate experimentally the long-term process of becoming accustomed to particular symbology. The external circumstances of route selection also range widely from relaxed, lengthy, pre-travel contemplation to a sudden contingency diversion in intensely frustrating traffic conditions. The amount of time the user is willing to spend with the map may also have a significant effect on the way it is read. All of these variables must consequently be considered in addition to the targetted variation in the design of the map stimulus itself.

7.1 Test Validity

One of the most important aspects of test selection in an experimental programme is the balance between internal and external validity. For an experiment to be internally valid, it must maximise the probability of the attribution of an observed experimental response to the correct independent variable, while to be externally valid it must be possible to generalise the results to different people or contexts (Campbell and Stanley,1966). Often in map design research internal validity is at a premium, and it is difficult to obtain statistically significant differences in performance between two differently designed fully complex maps because of the multitude of confounding influences involved, many of which may be working in different directions.

For example, in an experiment by Rhind et al.(1973) involving the interpretation of wind-rose symbols on geochemical maps, no consistently significant differences were found which could be attributed to the cartographic variables under test, largely because of the difficulty of the task (revealed by the very low response accuracy) which caused large differences between different map users and between different performances by any one user. Phillips (1984), while agreeing with Board about the desirability of using real map reading tasks, has suggested that where the target is the difference between maps rather than between their readers, whose abilities with respect to complex tasks vary immensely, easier activities are preferable to extended map interpretation tasks. He too stressed the importance of statistical significance in the generally accepted validation of a design-related difference. Board (1975) also mentioned the 'content validity' of a test, stating that test content must be balanced with respect to its objectives and not biased towards aspects that are easy to measure. Consequently he suggested that tasks related to more complex aspects of the perceptual organisation of symbol groups (section 8.3) should not be avoided.

However, while it might be tempting to use easier tasks and/or homogenous samples, this can only be done at the cost of external validity, which is required if general design principles, defining the area within which the map designer can exercise his creativity without detriment to communication, are to be established. Robinson especially has repeatedly (e.g. 1952,1977,1982) called for research to enable design rules based upon the least practical perceptual differences on graphic magnitude scales to be discovered, especially in the less understood area of the colour variables. If sources of variation are excluded in experimentation, any theory established by such tests may be valueless if it cannot be shown to be general (Hopkin and Taylor,1979). One aspect of generality is that the test chosen should measure a general theoretical construct such as legibility or conspicuity (Board,1975).

However, as Rhind et al.(1973,p.115) stated, 'even the simplest map is an extremely complex form of information display: not only are there a large number of different graphic factors involved...but these separate factors are interrelated with one another...This ensures that it is difficult to alter one factor without either having to alter, or having an effect on, all the other factors. Previous research in this area has usually attempted to isolate the effect of one particular factor by taking it out of a real map context and testing it in a more controlled situation. The obvious problem with this approach is that it is difficult to extrapolate the results to a real situation and to practical applications.' External validity may however be established by replication- conducting several similar independent studies such as those which have compared contours and layer tints in the depiction of relief (Phillips,1984)- although this is clearly often impractical. As far as the present study is concerned, the main need for external validity is not so much in the extension of results to other map-using tasks, as the relevant ones can be targetted fairly realistically, but to other people, as obviously only a very small proportion of motorists/ road map users can be tested.

Clearly any individual experiment involves a necessary trade-off between the control of variables required to establish internal validity, and the realism of complex environments within which relationships must hold in order to be externally valid. This trade-off is evident in three related tensions:

1) between experimentation in laboratory conditions and the 'real world',

2) between simple and complex stimuli and contexts, and

3) between perceptual and cognitive levels of processing.

In the formulation of this collaborative study, it was envisaged that laboratory experiments would be undertaken at the R.A.F. Institute of Aviation Medicine (I.A.M.). The advantages of laboratory testing, compared to interviews in private households or public places, lie in the amount of control the experimenter is able to exert over variables such as illumination and subjects' colour vision, given that the length of trial that can reasonably be undertaken would allow for the inclusion of rigorous colour vision tests. Clearly, too, experiments involving certain types of non-portable test equipment must necessarily be carried out in the laboratory. However, the laboratory situation is inevitably artificial in some ways, such that subjects' response patterns may change in such environments (Hopkin and Taylor,1979). As people's performance and utilsation of memory is highly dependent on motivation, it is clearly advisable not to use tasks which become especially meaningless in the laboratory context (Wetherell,1984) such as perhaps the route planning task itself. A further implication of this motivation dependency, according to Petchenik (1983), is that where the subject finds little of use or interest in the map, he tends to focus on the graphic characteristics alone, so that they play a far more important role than they would in a realistic situation.

Where the experimental goal (n.b. not the task) is not obvious to the subject, however, the laboratory may be a more appropriate environment where the subject is more likely to expect and accept apparently trivial or meaningless activities. However, tests of subsidiary tasks in the laboratory need to be validated against 'real world' tests to assess the relevance of the measured indices and determine how 'artificial' the results are. For example, in the perception of road signs (Dewar and Ells,1984), which measures are more important amongst legibility distance, glance legibility and attention value? A further limitation of laboratory studies in the present context is the necessarily restricted range of subjects, in terms of age and education levels, at the I.A.M., where for security reasons testing is limited to establishment personnel. However, the perfect 'real world' experiment is also elusive. Resource constraints make it effectively impossible to replicate more realistic motivation by actually asking subjects to follow their chosen routes, nor is it easy to see how an experimenter could attempt to observe the route choice process, either in or out of the vehicle, as a 'fly on the wall' where his presence would not influence the relevant behaviour.

The tension between simple and complex stimuli is of special importance to map design research (cf. section 4.4) because of the inherent complexity of map images. Clearly the simplification of the stimuli allows for much more control over stimulus variables, and consequently improves the internal validity of the test. As Noyes (1979) has stated, the problem of interpreting a significant performance difference between two maps is determining, through the confusing and diluting effect of 'visual clutter', which of the relevant stimulus differences is/are responsible. However, there are always particular problems with maps in attempting to isolate one target variable amongst their interrelated symbology (Petchenik,1983), and to vary it while all else is held constant (Potash,1977). Even a semi-realistic map has the potential to provide conflicting cues.

While Hopkin and Taylor (1979) have suggested that examining how people learn to see symbols may be more practicable than studying how they see whole maps, they have noted, along with Board and Buchanan (1974) the failure of psychological experiments based on highly simplified stimuli (e.g. the spatial frequency grating, a pattern of alternating light and dark lines with gradual contrast) to be relatable to more complex cartographic contexts, whilst violating Board's principle of content validity. Simplified maps too may be of little use, as it may be difficult to relate findings from them 'to a realistic cartographic context where other detail sometimes helps (through supporting associations)', especially in cases of coding redundancy, 'or hinders (because of visual clutter or 'noise') the communication process (Board and Buchanan,1974,p.128). Consequently it is not surprising that cartographers have often called for experiments dealing with real maps. 'Only by experimenting with complete and finished maps can we learn anything.' (Keates,1962) Unfortunately the results of such work have revealed few significant differences.

While to a certain extent it is impossible to isolate completely either perceptual or cognitive processes in a task, there is a fundamental difference in the methodology appropriate to the investigation of each set. Perceptual processes are stimulus-driven and performed without conscious effort, and thus are within the realm of the positivist epistemology of the natural sciences, because, as noted above, their operation varies little between people, and consequently this is the level where map design differences are likely to have the greatest effects. When cognitive processes and therefore 'meaning' are involved, stimulus characteristics still have some influence through their affective and connotative aspects, but the importance of interpersonal differences is greatly increased (M.Wood,1972). The cognitive aspects of a complex map reading task such as route planning also become more significant the longer the subject is prepared to spend in viewing (Castner,1978), and this too will clearly vary between people. Consequently epistemology and methodology must move into the more complicated realm of the social sciences in order to approach the problem of how meaning is assigned by individuals to features and images.

In fact the central problem for the map designer concerned with efficient cartographic communication is in assessing which aspects of map reading performance are due to innate characteristics of the map user that should be 'respected, utilised and deferred to in design', and which are based on acquired conventions that might be altered where desirable and eventually relearned (Petchenik,1983, p.60-61).

Consequently one of the main implications of these tensions for experimental design is in sampling. In simple perceptual tasks with simple stimuli, no relationship has been found between test scores and measures of the subject's experience (Castner,1983; Eastman and Castner,1983), suggesting that these reactions are to some extent innate, and that maps should be designed to accommodate them. The precise nature of the 'experience factor' has not been defined, but it appears to relate more to the development of cognitive abilities rather than a store of factual knowledge (Eastman and Castner,1983). However, it seems to have emerged in experiments involving route planning, where A.Morrison (1974) found significant differences in performance between people of different social class/ education levels. Thus in the present context it would be possible to carry out such tests, focussing upon internal validity, on the unrepresentative I.A.M. subjects, while experiments involving real route planning tasks and complex map stimuli, and having more external validity, would necessitate the use of a representative sample.

7.2 Techniques for Map Design Evaluation

According to Robinson (1977), the speed and accuracy of task performance, the minimisation of visual confusion and an appropriate calibration of qualitative and quantitative graphical magnitude scales are all important aspects of the efficient communication of a map. For the researcher wishing to assess the efficiency of a given map design in such respects, Robinson has distinguished between two sets of available techniques. 'Indirect' techniques include the assessment of 'good design' by the level of sales of a map or by the trial-and-error of the map designer. 'Direct' techniques consist of both subjective assessments and more objective measures of the performance of the map or of particular map symbols.

Subjective evaluations of map design might come in the form of a map review, or the preferences of a group of 'experts' (e.g. Van Meeteren,1974) or the map customer, possibly by a questionnaire enclosed with the map. While questionnaire surveys are more easily adapted to finding out what information is thought to be required on a map rather than how it should be shown, they have been used to provide rankings of map users' preferences for different coding dimensions (e.g. R.M.Taylor,1974). Adams (1967) worked backwards, by choosing maps he considered to be well-designed and analysing their specifications. Subjects may be asked to categorise their preferences in terms of specific semantic or numerical scales (often 5- or 7-point scales). Studies of preferences are likely to provide more definite recommendations for map design than performance-based experiments (Board and Buchanan,1974) but these may be of limited practical use as, however they are measured, there is no necessary link between people's preferences for existing maps and their performance with them. In fact they generally appear to prefer maps and map designs with which they are more familiar (Sheppard and Adams,1971; A.Morrison,1975). Information on preferences may however provide a useful background to more objective tests of map performance, especially with unfamiliar designs.

Task-orientated design research techniques fall into two main categories- the magnitude estimation of symbols and the evaluation of a map user's performance. The former class is occupied by psychophysical experiments which aim to determine a definite relationship between stimulus magnitude and subjective (perceived) magnitude for any given graphical dimension, by asking subjects to provide a numerical description of their sensations. Such studies are necessary if graphic magnitude scales are to be determined, and have a long history in map design research. The first experiments by Flannery (1956) and R.L.Williams (1956) on the perception of graduated proportional circles date from a period when certain psychologists were taking renewed interest in psychophysics and seeking scaling coefficients for a whole variety of perceptual continua. 'Direct' methods of magnitude estimation were used to create quantitative scales of graphic magnitude for the depiction of interval/ratio or ordered polychotomous data, while 'indirect' methods involving difference estimates in paired comparisons and perceptual thresholds were (less frequently) applied to the determination of 'just noticeable differences' and scales of qualitative discrimination, from which in the case of quantitative variables, magnitude scales could be inferred (Gilmartin,1981; Castner,1983). Psychophysical tests are predominantly concerned with low-level processes, and measures of subject experience have not been found to influence their results (Castner,1983).

However, the tide has since turned against psychophysical study in cartographic research for a variety of reasons based largely upon the unfulfilled expectations of this work. Clearly where the scaling of a single stimulus dimension was being assessed, the stimuli themselves were very simple, and difficulties were experienced in relating these findings to realistic map contexts (Board and Buchanan,1974). Other criticisms have focussed upon the limited role of quantitative comparisons in map reading: according to J.L.Morrison (1976), estimation is just one of the four main map reading tasks. Board (1975) and Petchenik (1983) suggested that many of the studies involved tasks which were probably very different from those that subjects performed spontaneously in map reading, and that by treating the subject as a 'black box' no light was shed upon task validity. The cognitive process of communicating the nature of the required scale via the test instructions may also present complications, as slight variations in the question asked may have a considerable influence on the results (Teghtsoonian,1965 cf. section 4.6; Shortridge and Welch,1980). Dobson (1980) complained that these studies also ignored the deterioration of response quality through time in memory, relevant in the performance of complex tasks, and called for work on 'memory psychophysics' in preference to tests based on an immediate response to the stimulus.

Notwithstanding all these limitations, there is no need to throw the psychophysical baby out with the bath water. Gilmartin (1981) has emphasised the need in cartography for all three types of experiment- the primarily psychophysical, the primarily cognitive, and the integrated psychophysical/ cognitive. Magnitude estimation clearly plays a part in map reading, and a scale is required for the perception of multidimensional line symbols, but in doing so three things must be realised. Firstly, magnitude estimation is but one subsidiary task in the route planning process, and only in real map contexts can its true significance be assessed. Secondly, great care must be taken to use an appropriate task and question formulation- the assessment of 'prominence' being the most commonly-used relevant term by subjects in the pilot study. Thirdly, subjects must also be encouraged to comment on those influences on their assessments that they consider to be important (Petchenik,1983). This is necessary to ensure that the final results are not a 'uniform' average with a very high variance which actually reflects no individual's criteria of assessment.

The most objective evaluation techniques are those based upon the quality of task performance, most typically the speed and accuracy of response. Reaction times, used to measure speed, are well-established in experimental psychology as an index of task difficulty, without which accuracy data is fairly meaningless (Dobson,1980), and are commonly used to investigate bottlenecks in the information processing system such as the allocation of focussed attention (often in connection with timing devices linked to tachistoscopes). In cartographic studies, search is the most commonly used task (e.g. Bartz,1970; Phillips and Noyes,1980) as most map reading activities include a large search element. Search experiments obviously involve peripheral as well as central vision, allow for the use of more complex and meaningful stimuli, and may serve as a relatively simple guide in determining those aspects of cartographic communication over which the map designer may have some influence (Castner,1983). Dobson (1980) has also suggested that rewards and debits could be used to provide greater motivation for the subject and encourage him to optimise his performance. Many of the more complex tasks cannot be treated in this way: in route planning, for example, there is not necessarily a single right answer against which the accuracy of every route can be judged, but the quality of the <u>outcome</u> of task performance, namely the route itself, can be assessed.

Performance tests are often fairly simple to administer but much more difficult to analyse. Audley et al.(1974) consider tham to be useful for quick comparisons of map pairs (or presumably for a series of studies focussed on different individual target variables), but are dubious about their value for more fundamental assessments. The basic analytical problem, in addition to the usual intersubject variability, is that the test outcome may be just a single statistic, such as a difference in mean reaction times for correct responses between two maps or stimulus sets, which yields no idea of which stimulus variables might be responsible (Robinson,1977; Noyes,1979). Where the tasks involved are performed with cognitive effort, verbal information supplied by the subject may again be an invaluable aid to analysis.

Occasionally other evaluation techniques have been suggested which might target the more fundamental processes of map reading, providing a slower and less certain payoff for the experimenter but perhaps making a greater long-term contribution (Audley et al.,1974). Ideas of how maps are generally read are in fact very vague (Hopkin and Taylor,1979). The difficulties of accessing this underlying level experimentally have however been discussed above. The two main techniques which have been used for such endeavours are the 'fly on the wall' approach and eye movement analysis. As mentioned above, the former is not feasible for the current project, and while the latter might be expected to relate more to the attraction and holding of the viewer's attention, the establishment of visual hierarchies and the ease of map reading (Dobson,1979b; Castner,1983), interpretation of the movement records presents substantial difficulties (Hill,1975; cf.section 8.3).

7.3 The Experimental Programme

The above analysis has suggested that any individual experiment necessarily involves a trade-off between control and realism in terms of the sample tested, the experimental conditions involved and the stimuli used. One way of attempting to navigate through this potential minefield, used by Peterson (1979), is to set up a linked series of experiments which progress from the psychophysical to the cognitive, and from the simpler tasks and contexts to the more difficult and complex. Consequently the following programme of three experiments was devised, with the intention that the results from the first two experiments could be used in the design of the latter one(s), and that the final route planning experiment might assess the influence of line prominence and conspicuity, targetted specifically in the first two experiments, upon route choices.

Thus experiment 1 involved a psychophysical estimation of the relative magnitude of pairs of line symbols, varying multidimensionally (in width, colour and casing state), based on their apparent prominence, with no mention of maps in the experimental instructions. Given the use of simple stimuli and the predominance of low-level processing in the task, this could be carried out in controlled laboratory conditions on an expedient sample, with the emphasis on internal validity. Experiment 2 differed in consisting of a visual search task, assessed by speed and accuracy measures, on a somewhat more complex stimulus array but still involving much the same level of processing in the same laboratory context with an equivalent sample, and with the emphasis still primarily upon internal validity. However, experiment 3 involved complete route planning tasks on experimental maps of realistic complexity, incorporating both low- and high-level processes. Routes were evaluated for their quality and composition, and the testing was carried out in public places (and some households) on a sample of road map users that was as representative as possible in terms of their education and experience levels. Thus more emphasis was placed upon external validity. It was hoped that together these experiments could provide specific guidelines for design which could improve the overall efficiency of routes chosen from road maps.

8. THE LABORATORY EXPERIMENTS (EXPERIMENTS 1 AND 2)

The aim of these experiments was to determine the relative saliency and attention-getting properties of lines of different colour, thickness and casing state employed on a generally light background (in positive contrast). As mentioned above, the symbol for the highest classification should induce the largest perceptual response and be the most attention-getting. Because of the effects of context, separate consideration was required for:

1) the magnitude implication of individual (out-of-context) line segments, or their prominence in central vision (experiment 1)

2) the conspicuity (attention-getting properties) of these line segments within specified networks (experiment 2).

8.1 Design of the Experimental Stimuli

It was necessary to design and construct an array of experimental stimuli for use in the magnitude implication and conspicuity experiments. The array (rear pocket) consisted of a rectangular grid, with each cell traversed diagonally by a line of a type in common cartographic use as a road symbol. The stimuli were to vary in terms of

1) line width
2) line colour
3) presence/absence, thickness and colour of casing
4) cell background colour.

An investigation of published road maps revealed the range of each of these variables in use as follows:

-line widths from 0.15 to 2.1mm

-13 nameable colour groups (red, green, blue, yellow, magenta, cyan, orange, purple, brown, pink, white, grey and black)

-black casings up to 0.4mm wide. A few symbols involved coloured casings (red, magenta or green) or the use of two spaced parallel coloured lines (red or blue).

-white and various tints of yellow, brown and green for large areas of background, with built-up areas represented in a wide variety of colours.

However, the range which can be utilised in the array is clearly constrained by two main factors. Firstly there is a need to prevent a combinatorial explosion of testing conditions. For example, using 13 colours, 6 widths and 4 casing states would require 312 cells, creating in paired comparison experiments a total of (312 x 311/2) or 48,516 testable permutations for each background colour. This is clearly beyond the patience threshold of the average experimental subject. Secondly, there are practical constraints of array size and production costs. It was decided to work to an A3-size 6-colour format. Use of process colours would have extended the range of colours available for a given cost, but they were considered to be unsuitable because of the unrealistically precise registration that would be required, for the fine lines especially, to avoid significant alteration of their width and perceived colour from the designed values.

In order that a 4x3 grid of cells could be viewed on a tachistoscope, cell size was set at 2.8 by 2.1cm, allowing a 14x14 grid of 196 cells to be constructed on an A3 sheet. Thus the length of each line segment (the diagonal of the cell) was 3.5cm. Given certain constraints (e.g. lines which could not be cased on account of being either too thin or black) a design of 4 line widths, 6 colours and 3 casing states (plus a few individual '2-colour' lines) on each of three background colours, filled 194 of the cells. With the maximum number of lines on any one background being 66, the number of permutations required in a full paired comparison test even then was 2145.

The width/colour/casing specifications were basically chosen from the ranges mentioned above. Few maps use lines thicker than 1.4mm, so this was the largest gauge used, with the first two steps down in the series (to 1.0 and 0.7mm) involving the same width ratio. The narrowest line (0.2mm) was chosen to see whether colour magnitude perception could still operate in the same way at the level of the most degraded linear stimulus found on road maps. The printing colours themselves are the six (5 plus black) most commonly found on road maps- consequently brown is included rather than purple, although the latter would have provided a fuller hue range. They represent the 'perceptual primaries', according to the Robinson et al. (1978) definition (see section 5.5). The Munsell system was used to select the colours, so that visual comparisons could easily be made (between chips), and to correlate with other relevant work which had used Munsell coordinates. The CIE specifications of the printed colours were also measured, on an ICS Colour Difference Meter (table B.1). Equations and graphs are now available to facilitate translations between some of the colour measurement systems, so the choice of which one to use is less critical. Munsell designations can be converted to process colour tint screen selections via CIE coordinates (Kimerling,1980).

The particular inks used were carefully selected to vary considerably in terms of Munsell value and chroma without being able to form a straightforward increasing value/decreasing chroma series. Such a series would be unproblematic in that it should yield a straightforward colour ranking (Cuff,1972a). This would not help to isolate any general principles which might be operating in more normal situations, where the highly distinct colours used, such as yellow (high lightness, high chroma) and brown (low lightness, low chroma) cannot be organised in this way. The chosen casing widths, 0.1 and 0.3mm, are also the most commonly used, with the thicker one enabling certain '2-colour' lines (i.e. those with coloured casings) to be included.

The uncased lines and the casings were produced by scribing, while hand-cut masks were used for the fillings of cased lines. Photographic combinations were undertaken within the Glasgow University Geography Department. The printing inks were chosen from the Munsell colours using the Pantone matching system, with the two background colours being 10% screens of the yellow and brown line colours. The stimuli were then printed at the expense of the RAF Institute of Aviation Medicine. The printing process inevitably caused some slight deviations from the specifications: the colours generally came out slightly stronger and darker than the selected values, making brown less distinguishable from black than was intended. The actual figures are shown in appendix B, table B.1.

8.2 THE MAGNITUDE ESTIMATION EXPERIMENT (EXPERIMENT 1)

The choice of techniques open to the investigator attempting to analyse the magnitude of visual responses is very limited. Objective measurement is very difficult: probing the brain is clearly not feasible. Clark (1985) has however, since the present experiment, demonstrated a practical 'scattering visibility gauge', which works on the principle expounded in Frenzel's experiment (section 6.1) of controlled stimulus degradation. It consists of a perspex strip graded monotonically in opacity by controlled 'sandblasting'. As the surface becomes progressively more pitted, the scattering of incoming light rays, and thus opacity, is increased. The visibility of an object is measured by the point along the strip at which it can first be seen.

Otherwise some form of subjective scaling technique is required, where the observer is asked to make a numerical description of his sensations. As Newhall (1939) remarked, sensations cannot be added or multiplied, but numbers representing them can. While this is likely to be less precise than an objectively measured threshold, it does cover all the stages of perceptual processing, taking connotations of magnitude into account as well as purely visual considerations. This particular situation was however constrained by two main factors: the unusual nature of the information required and potential subject fatigue. Problems a subject might have in a simple psychophysical experiment in reporting the loudness or brightness of a stimulus, which are everyday and frequently verbalised experiences, are likely to be multiplied here for two reasons. Firstly, people rarely encounter problems where they are <u>consciously</u> assessing the magnitude of a line. Secondly, in doing so there is no obvious single scale of magnitude given the multidimensionality of the stimuli, with both size and colour variables expected to contribute to the magnitude effect. Moreover, numerical scaling of colour in itself is unusual to most people. However, as mentioned above (section 5.3), multidimensional scaling techniques have been used to analyse overall colour difference estimates in terms of established perceptually linear scales (i.e. Munsell dimensions). Yet the relationships between perceptual magnitude and width and colour measures are not so well established and may well be non-linear.

Subject fatigue is certainly a factor with potential to confound test results (C.H.Wood,1976). Given the difficulties of the task, how long can the subject spend on the experiment before the quality of his judgements begins to deteriorate significantly? Unfortunately he certainly could not be expected to make the 1891 comparisons which would be necessary to test each of the 62 lines on the white background against the remaining 61. This would be a particular problem in this case, where neither the task nor the method used would be likely to induce much interest from the subject. Thus 30 minutes seemed to be a reasonable maximum, and certainly in retrospect it would have been difficult to procure voluntary subjects or ensure the maintenance of response quality using a longer experimental session.

Consequently at the outset there were no obvious precedents to this particular problem to imply that consistent results might be obtainable at all. The main requirements in experimental design were to make the task as simple and clearly understandable as possible, and to extract the maximum amount of valuable information from each short experimental session. The initial aim was to be able to produce for each subject an interval or at least ordinal scale for the perceived magnitudes of each of the 62 lines on the white background.

Experimentally, estimated magnitudes can be obtained in two different ways. One method is to ask the subject to select or vary the stimuli to represent a given ratio- e.g. adjusting the volume of a tone to make it half as loud (the method of fractionation) or twice as loud (the method of multiple stimuli) as a reference tone (Guilford,1954). In the current situation the large set of invariant stimuli militates against this approach, although a method was considered in which the subject would be presented with several lines at once and asked to arrange them in rank order of perceived magnitude. It was however thought that handling several stimuli at once would make this a relatively awkward and time-consuming task for a low information return. Alternatively, the subject is asked to assign numerical estimates to stimuli presented before him. Experimental methods involving paired comparisons of stimuli, keeping one of each pair as a fixed standard, had long been used for their speed and consistency of results with simple greater/lesser judgments. Higher-level data have subsequently been obtained from them by what Stevens (1957) called 'ratio estimation' methods, where the subject is asked to assess the ratio between the two stimuli. Witzel et al (1973) reported several studies which had found that 'meaningful colour difference configurations could be obtained from ratio judgements of visual colour differences.'(p.615) Newhall (1939) also found ratio judgements to be sufficiently reliable to use them for the 3 million colour difference estimates that were made at the time to improve the specification of Munsell colours. He distinguishes between two types of 'ratio method':

1) the ratio of the sense magnitude (i.e. for quantitative scales) of the test stimulus to the standard, where every pair consists of one of each. Taking 3:1 as an example, it is either reported directly, as in the original experiment by Richardson and Ross (1930), or by the 'constant sum' method, where a fixed number of 'points', usually 100, is divided between the stimuli (75:25) by the subject (Metfessel,1947; Comrey,1950).

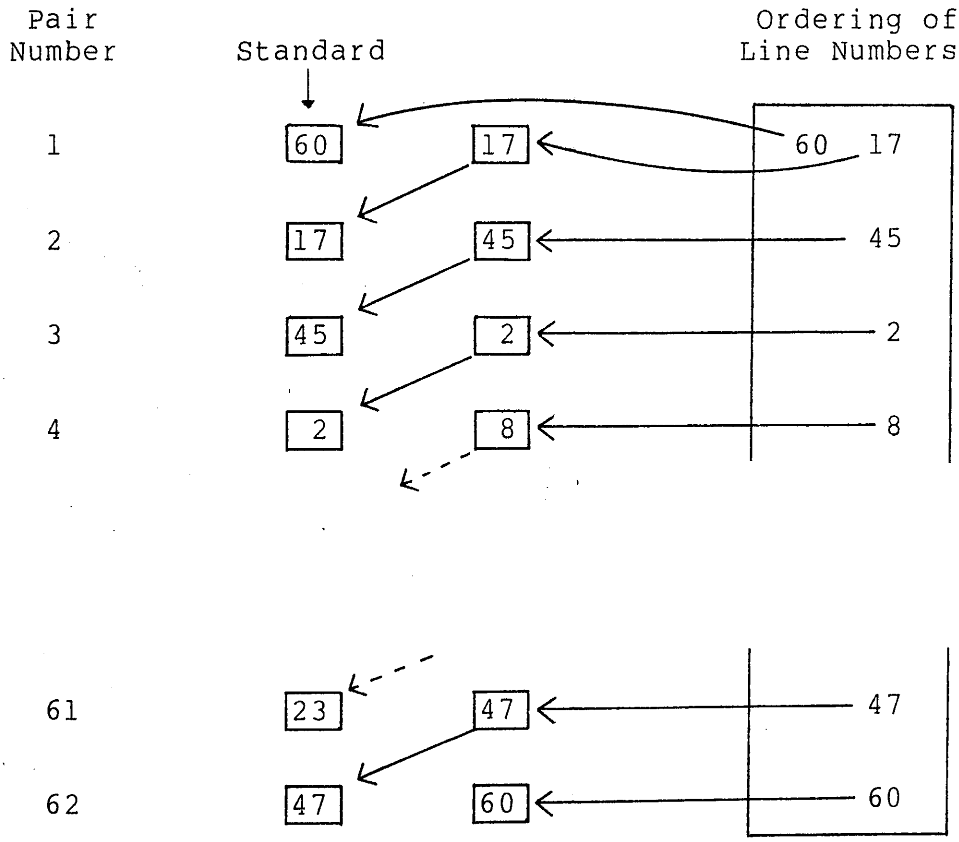
2) the 'difference ratio method' where the difference between the pair is compared to a standard difference (e.g. between the endpoints of an interval scale). The first application of this method (Richardson,1929) was in fact to colour- marking the perceived position of various pinks on a line representing a continuum from white to scarlet. A variation of this used by Pinkerton and Humphrey (1974) made the two stimuli the endpoints, and asked the subject to move a pointer to the visual 'balancing point' between them. Direct reporting of numbers can also be used.

The former method is appropriate to the present experiment where a quantitative scale of stimulus magnitude is sought. Guidance in experimental design was taken from Stevens (1956), who claimed to have used the technique with very consistent results over a wide range of subjective magnitudes. He noted the potential fragility of the process of quantifying one's sensations, and recommended that as few constraints as possible be imposed upon the subject. Consequently direct reporting was used rather than the 'constant sum' method, which it was considered would be slower and more awkward for the subject, particularly when attempting to specify large ratios. Also the advantages of 'constant sums' in analysis by allowing the computation of the average score of a stimulus in all its comparisons (Comrey,1950) require all possible pairs to be assessed.

In Stevens' experience, subjects had found 10 to be the easiest number to work with as a designation for the standard (or denominator of each ratio) so it was adopted for this experiment. Given the fundamental problem in experimental design of extracting as much quality information as possible from the subject before fatigue began to impair judgements, keeping a fixed standard stimulus in each pair was clearly to be avoided if possible. If the stimuli on both sides of the pair could be changed (one at a time) and compared to each other, twice as many estimates would be available to enable the relative positioning of each line on the magnitude scale. With the problems of accurately remembering a standard, even for short periods, in the midst of a barrage of different stimuli (Newhall,1939), one of the lines in each pair was chosen to be the standard. Consequently, as these lines changed with each pairing, so did the standard. A subjectively defined rolling standard had been used by Severud (1968) in tests of the relative visual importance of line pairs, in that the subject called the more important line in each pair 100, and provided a relative estimate for the less important line. For simplicity, the standard in the current experiment was defined spatially: throughout one cycle of the experiment the line on the left of each pair was the standard, while for the other cycle the line on the right was used. Each cycle was a random ordering of the 62 stimuli. The first two were presented as a pair, and subsequent pairs were made up by replacing the first line of each pair by the next on the list (figure 8.1), until finally the very

Figure 8.1 STIMULUS PRESENTATION (Experiment 1)

Cycle 1.



first stimulus returned to form the final pair and the cycle was closed. Thus in each cycle every line segment was compared with two others, so that over the two cycles four out of the possible 61 comparisons were investigated, a sampling fraction of 6.56%, involving a total of 124 observations per subject. This experimental design met Kendall's (1955) criteria for 'incomplete balanced blocks', to be used when the testing of all possible pairs is not feasible.

The experiments were carried out in two different locations- the RAF Institute of Aviation Medicine and Glasgow University Geography Department- in an attempt to get a reasonable number of subjects, and conditions were made as comparable as possible. The experimental stimuli were two sets of the 62 lines mounted individually on 6x4" white cards of the same lightness as the line background. One set had the lines to the left of centre while on the other they were right of centre. The equipment used at Farnborough was an Instrumental Colour Systems viewing cabinet fitted with two 20 watt Atlas artificial daylight tubes. (The Munsell matches were made in the same quality of illumination.) This provided a neutral grey background against which the card pairs were viewed. No viewing cabinet was available in Glasgow, and here the tests were illuminated by a ceiling-mounted Omega 65/80 watt daylight tube. The stimuli were viewed from the subjects' normal reading distance. All subjects were first tested for normal colour vision using the Ishihara colour blindness tests.

The following instructions were given verbally to each subject:

"The aim of this experiment is to examine the impressions of quantity, or magnitude, conveyed by different line symbols. You will be shown pairs of lines of equal length. One of each pair is assigned the value 10. Please estimate according to the relative prominence of the lines the value you would consequently give to the other line in each pair. For example, if it is twice as prominent as the reference line, give it a value of 20, or if it is half as prominent... (Here the subject was invited to supply the value in order to check that he understood the scoring system.) Please respond as quickly as you can." The subject was then asked if he understood the instructions and if he had any questions. The experimenter then presented the first stimulus pair, explaining which line was the standard, and the experiment began. The substitution and moving of the stimuli was performed by the experimenter as soon as the subject had verbally reported his estimate for the non-standard line in each pair. At the end of the experiment the subject was asked what factors he had taken into account in assessing the magnitudes of the lines.

34 volunteers with normal colour vision acted as subjects- 17 employees of the Institute of Aviation Medicine and 17 students/research staff from the Glasgow University Geography Department. The tests were carried out during September and October 1984. The results from three of the subjects had to be discarded because they had either misunderstood the experimental instructions, or had been insufficiently discriminating to provide worthwhile results. This left 31 usable subjects, 18 males and 13 females.

8.21 Obtaining the Magnitude Scores

The main problems involved in the statistical analysis of ratio estimation experiments are invariably caused by inconsistencies both within and between subjects in their use of numerical scales. In this experiment, with each cycle beginning and ending with the same stimulus, by starting with a value of 10 and multiplying through by each successive estimate divided by 10, a final partial product of 10 would be obtained for a perfectly consistent subject. In practice this did not occur, and the final 'misclosure' was often very large. Attempts were made to distribute this error using the 'equal shifts' method employed by surveyors for adjusting traverses, but these were unsuccessful because discrepancies in subjects' scaling were not gradually and equally incremented throughout the entire cycle, but were clearly caused mainly by estimates of large differences, such as most comparisons with the 0.15/0.2 mm lines. The problem of extreme differences is well-known in both subjective estimation and difference matching techniques (Witzel et al,1973) where there is a general tendency to underestimate them (e.g. Bevan and Dukes,1953). Large differences are clearly more difficult to assess quantitatively with accuracy, but a further cause of this in estimation experiments is the subject's often unconscious inclination towards treating the estimates additively rather than multiplicatively. For example, in a totally additive treatment, 2 (10-8 and 10/5) would be the inverse equivalent of 18 (10+8) rather than 50 (10x5). Such disparities clearly have the greatest effect on the estimates at the extremities of the scales, the extreme negative differences being overestimated relative to the extreme positive ones, a tendency also found by Castner (1983). In such a situation, the frequency distribution of the logarithms of the estimates, which would be symmetrical with consistent ratio estimation, would be negatively skewed. This was the case for every subject in the current experiment.

An investigation of each subject's scores revealed big disparities between the ratios of the smallest and largest figures used by all but four of the subjects, the respective modes being 1 (1:10) and 40 (only 4:1). The apparent causes of this are twofold. Firstly, with line width not unexpectedly proving to be the major determinant of magnitude, the inclusion of the 0.15/0.2 mm lines acted as a major discontinuity in the stimulus set, inducing extreme reactions from the subject which were not compensated for in return comparisons with the remaining (0.7 to 1.4mm wide) lines. Secondly, all of the subjects used non-integers very rarely if at all. These were clearly not necessary for the majority of line pairs in the set at expectable levels of subject resolution. However in using 10 as the standard, imprecision clearly arose in the assessment of the smallest ratios where there were very few integers to choose from. For some people 1 clearly became a default value as soon as one of the 0.15/0.2 mm lines was introduced. Even if this was the best integer that they could use to describe the sensation, it could still represent any ratio between 0.5:10 and 1.5:10, or in other words an implied magnitude of anything between 1/20 and 1/6.67 of the reference stimulus. Thus any attempt to fix lines involved in such comparisons on a hierarchical scale of line magnitudes was fraught with difficulty, especially with only four comparisons being made for each line. In retrospect it would seem that 100 would have been a better reference value for the range of stimuli involved here.

The aim of this experiment was to construct a single interval scale for the perceived magnitude 'scores' of each line. In order that this could be done, each subject's own scale had first to be transformed into a common interval scale of perceived differences. To this end it was necessary to compensate for two separate aspects of subjects' different use of scales, namely the skewness problem mentioned above, and wide variations in the overall range of numbers used by an individual (from 5-20 to 0.1-1000). The skewness does not necessarily mean that the data are not consistent relative to some other scale. In this case, while an estimate of 40 does not necessarily represent the equivalent inverse subjective ratio to a 2.5, it may still consistently indicate a larger subjective difference than, say, a 35 or a 30. If so, a suitable monotonic transformation could be applied to each subject's own scale to remove the skew and make the distribution more 'normal'. The problem of range is primarily a consequence of some subjects being more conservative than others in the use of numbers. This situation was encountered by Pinkerton and Humphrey (1974), who found that although subjects varied considerably in terms of absolute pointer displacements, their rank orderings of the stimuli were very consistent. Once the data have been 'normalised', this problem could be overcome by the use of standard scores (z-scores).

Firstly a check was made to assess whether the data were consistent in terms of the <u>directions</u> of the estimated differences. Consequently the estimates were aggregated into two categories

1) all those over 10 (assessed line > reference line)
2) all those under 10 (assessed line < reference line).

Inconsistencies could arise in two ways. Firstly, in cases where the random stimulus ordering had presented a subject with the same pair in each cycle, and a different decision was made each time (A>B,B>A). The second type, with the potential to occur far more frequently, includes what McManus et al.(1981) called 'circular (or intransitive) triads' (A>B,B>C,C>A), and equivalent higher-order incompatibilities (e.g. A>B,B>C,C>D,D>E,E>A). A computer program was written to check for these incompatibilities and only four were found, two of each type, leaving 27 (87%) of the subjects who were completely consistent in this respect.

Each subject's scores were investigated individually in an attempt to find an appropriate transformation for each, and a computer program was written to enable a rapid assessment of skewness for a given transformation. The logarithms of five subjects' estimates were found not to be significantly skewed (using the stricter of Pearson's (1930) bounding values) without transformation, while for the others the negative skewness was removed adequately by raising the figures to a power varying between 1.5 and 2.5. Computational problems were however encountered for subjects who included estimates of less than 1 which produced negative logarithms. While these transforms were having statistically desirable effects in removing skewness and balancing extreme values, their use could only be justified behaviourally if the differences between each subject's scale and the 'common scale' varied in a systematic way across the whole range of the data, such that each estimate could legitimately be transformed by the same function. It became increasingly obvious that this assumption was untenable as certain clear behavioural relationships in several subjects' estimates (e.g. the use of 20 to represent a ratio of 2:1) became distorted. What was required was a transform which examined the whole range of estimates used by each subject and treated each figure separately.

At this point it was necessary to call upon the statistical expertise of Mr.M.Spencer of the RAF Institute of Aviation Medicine who selected and applied an appropriate transform using a NAG library subroutine (E04JAF). In order to obtain consistent difference measures which could be used to space out the perceived magnitudes of the 62 lines on a single linear scale, the transforms were selected by the subroutine for each subject such that each transformed estimate equalled the required difference measure plus an error term, with the difference measures estimated by least squares (minimising the sum of squares of the errors over the sum of the transformed estimates squared). The only constraints imposed on the transforming functions were behaviourly justifiable, namely that

1) f(10)=0, i.e. the estimate 10 should be taken to be a reliable indicator of no perceived magnitude difference between the pair of lines involved, and

2) f should be monotonic: i.e. if one estimate was

greater than another before transformation it should still be so after transformation, such that the transformed value of 5 will be greater than that of 1, although the actual interval distance between them would in most cases have been considerably narrowed.

The transformed differences were then converted to a linear magnitude scale, creating a base by setting line 1 in the stimulus set to 0, and were standardised to mean 0 and unit standard deviation for each subject.

8.2.2 The Causes of Perceived Magnitude

An interesting overview of the likely causes of the perceived magnitude of line symbols is provided by a summary of the comments made by the subjects about the factors they were aware of in making their decisions. Overall, line width was most frequently cited (13 mentions), followed by darkness (9), colour strength (5) and colour pleasantness (3). Eight subjects also commented upon the relative significance of width and darkness with all but one considering width to be the more influential, and three also considering colour strength to be less important than either. However, two people remarked that colourfulness was the most important element in the attention-getting ability of each line. When specific colours were referred to, red (11 mentions) and black (7) were always (in) the top category, with yellow (5) at the bottom. Brown was also considered to be fairly prominent, while views on blue and green were mixed and often related to affective considerations. With regard to casings, the most frequent comments were that they boosted the prominence of the line segments (12), although 7 of these felt that the boost only occurred, or was much more marked, with light fillings. Three people also mentioned the effect of casings in altering the apparent colour of the fillings.

The line scores themselves (table B.2) were analysed in two main ways. An analysis of variance, treating colour and designed width as categorical variables, demonstrated the significance of their overall effects, while multiple regressions were run in order to

1) provide a breakdown of the influence of colour variables, and

2) derive a more general formula for perceived magnitude which might apply to stimuli with different specifications. These techniques are particularly suitable here as the effects of width and colour are conceptually independent and the variables are uncorrelated (table 8.1). Moreover, the same could also be said of the breakdown of colour into lightness and chromatic effects. As mentioned above colour responses can be considered, both physiologically and psychophysically (Evans,1974), to have separate achromatic and chromatic components. Luckiesh (1944,p.242) talked about colour as being 'superposed upon the more basic world of brightness', and clearly people can have totally normal lightness perception whilst being totally colour blind.

TABLE 8.1 Correlations of Regression Variables (Uncased Equation)

Log. width and lightness contrast	$r = -0.024$
Log. width and saturation	$r = 0.013$
Lightness contrast and saturation	$r = -0.118$

The computational speed and useful regression diagnostics of the interactive statistical package MINITAB allow regression to be used as an exploratory technique in the search for the best derivable model. Each variable could be examined individually to determine how it should be specified in an overall linear equation. For the purposes of analysis, cased and uncased lines were treated separately.

8.2.3 Uncased Lines

The analysis of variance considered the effects on the line scores of colour, width, subject group and the interactions between them. Of these, the effects of width ($F(3,87)=672.18$, $p<0.0001$) and colour ($F(5,145)=21.84$, $p<0.0001$) were highly significant. There was also a significant colour/width interaction ($F(15,435)=1.98$, $p=0.016$). This interaction was due mainly to slight differences between the designed and actual widths of certain individual lines, and a different ordering of the colours in the scores for the narrowest lines (black and green being particularly affected, see table 8.2). For the three wider categories the relative positions of the colours are generally very consistent, and confirm the comments made by the subjects.

TABLE 8.2 LINE SCORES: COLOUR RANKINGS BY WIDTH CATEGORY

Rank	Actual Widths (mm)			
	0.15	0.7	1.05	1.4
1	red*	red	black	red
2	green	black*	red	black
3	blue	brown	brown	brown
4	brown	blue	blue*	blue
5	black	green	green	green
6	yellow	yellow	yellow	yellow

* Adjusted for slight width differences

The Newman-Keuls shrinking range test and Dunn's multiple t-test (where Newman-Keuls was inappropriate because of the different sizes of the error terms) were used to assess the significance of differences between all pairs of means within the one-way analysis of variance model for colour. Overall the red lines were perceived to be significantly ($p<0.001$) more prominent than the brown, blue and green lines, which were in turn more prominent ($p<0.01$) than the yellow at all widths. Additionally, for the two widest categories only, the black lines were of significantly higher magnitude than the greens. With the narrower lines, there is a marked bunching together of the values for the 'middle four' colours (black, brown, blue and green), especially at 0.15 mm. Clearly subjects had difficulty in discriminating quantitatively between these stimuli. Clearly the effects of stimulus degradation mentioned in section 5.4 operate at a width of 0.15 mm, with only the red and yellow lines clearly maintaining their distinctiveness, as M.Wood (1968) had suggested.

A series of multiple regressions were undertaken on the mean scores for each line (table B.2) in order to find the best quantitative descriptions of the effects of width and colour variables. Preliminary analysis was undertaken on the data from the 13 Farnborough subjects before the Glasgow scores became available, and the findings were confirmed by the full sample, from which the results presented below are derived. Generally very high r-squared values were obtained, but care had to be taken in interpreting these because of the dominating effect of width overall (explaining up to 95% of the variance on its own), and the fair amount of leverage exerted by extreme values in width (the 0.15 mm lines) and colour (red and particularly yellow). Several of which were around the critical point to be diagnosed as 'X-points' (values with a large influence upon the regression line) by MINITAB.

Analysis proceeded by respecification of the independent variables until any autocorrelation in the residuals which they had caused was removed. The effect of width was investigated first. For the full set of 24 lines, it was discovered that the shape of its relationship with perceived magnitude was not explained by either straightforward width or its log, which both left autocorrelation in the residuals, and that the best fit was obtained using (width/6)+log(width). The significance of this is examined below. However, the main problem in analysis was in seeking an adequate specification of the effect of colour. Initially, overall colour difference equations derived from colour discrimination experiments (Godlove,1951, Farmer et al.,1980) were tried with various weighting factors between lightness and chroma contrasts. Obviously no hue contrasts were involved as each line had an achromatic white background. However these models were poor predictors- either the predicted values for black were higher than those for red (indicating that lightness was weighted too heavily) or yellow was higher than blue (indicating that chroma was weighted too heavily) and there was no intermediate balance point. Consequently lightness contrast (in Munsell steps against the background) and chroma contrast (equal to actual chroma on a white background) were subsequently treated as separate variables. In the final equation, the best specification of lightness contrast was as a simple linear relationship. It became clear that the problem lay in the chromatic measures.

Firstly it was found that the effects of the more chromatic colours were underestimated using the straightforward chroma measure, and a far better description was obtained by squaring the chroma. Although Munsell specifications of high chromas are known to be relatively unreliable and generally underestimates (Evans,1959), this clearly represents a domination of extreme chromas. Secondly, the effect of colour was clearly not solely one of chroma, or else the scores for yellow lines with their high chroma contrast would have been considerably higher. Because of their low lightness contrast, their colourfulness was clearly not having much impact. One subject commented 'You could've made them as wide as the wall, but they still wouldn't hit you.' In terms of visual impact it seemed that a more relevant description of colour effect would be given by a saturation measure, where darker colours of the same chroma would have higher values- i.e. some combination of colourfulness and lightness contrast. An operational measure of saturation as Munsell chroma contrast squared times lightness contrast was found to be the most appropriate. This is not 'double counting' lightness contrast, but indicates that aside from its achromatic effects, it also has a separate and additional contribution to make to the visual impact of coloured lines. For example, if these lines were reproduced on a colour television and the colour control were to be turned up gradually from minimum (achromatic), the brightness contrasts would not be altered as the added dimension, colour, is superimposed. Yet they also influence the amount of <u>extra</u> visual impact that colour induces.

With the width, lightness contrast and saturation measures in the equation, despite a multiple r-squared approaching 99%, there was still some systematic effect of colour remaining in the residuals: the values for red and black were considerably underpredicted by the model, while brown was overpredicted to the extent of having higher predicted values than black. There were various possibilities for the extra ingredient. The literature review had suggested three in particular which, because they were highly intercorrelated, had to be investigated separately:

1) physical fluorescence
2) 'brilliance'
3) affective values

It had been found in tests on the ICS Colour Difference Meter that some of the colours displayed considerable fluorescence, although the significance of this could not be gauged as no fluorescent standard was available. The fluorescent yield was highest for the blue, the colour that was most difficult to match precisely to the Munsell chips. When the fluorescence figures were entered into the overall equation, there was a significant increase in explained variance. However, from an intuitive point of view it was difficult to see how a small amount of fluorescence could affect magnitude estimates other than by boosting chroma contrast, unless it translates into slightly higher affective values. However attempts to relate fluorescence to chroma contrast were unsuccessful.

The two other variables which might account for the remaining systematic variance, 'brilliance' and affective value, have both been related graphically to the Munsell coordinates and could therefore be operationalised in this experiment. Brilliance was measured on Evans' (1959) 'gray-fluorent' scale, while affective values were read off the charts supplied by Guilford and Smith (1959), or interpolated between them as necessary. Separate charts are provided for men's and women's preferences, so an average, weighted according to the composition of the sample, was derived from the figures for the two sexes. It should be noted that these charts relate to pre-1943 Munsell specifications, although the differences involved are very slight. The figure for green was however suspect as it required interpolation across a hue range where the affective values were changing most sharply.

The significance of 'brilliance' was likely to be affected by the limited context of a pair of short line symbols seen against a homogenous light background (Evans,1959). However, it was also significant in the equation, although when inserted with just width and lightness measures, its effect was absolutely identical to that of chroma. Of the three measures, affective value was clearly the most successful in the equation. This was in part due to the balancing of brown and black, with the latter having a higher affective value as 'pure' black is generally perceived to be more pleasant than weakly chromatic mid-to-low value colours. This is the only colour measure to yield the significant value for black necessary for the explanation of the current experimental results. However, the scores for green lines were overpredicted by this model. Given the interpolational difficulties mentioned above and the dislike for green expressed by some of the subjects, a lower value based on non-linear interpolation was substituted. This not only removed the overprediction but also significantly improved the overall model, which was becoming more sensitive to changes in affective values as their contribution to it increased. Ultimately this still leaves questions unanswered. As Pinkerton and Humphrey (1974,p.165) noted 'a correlation between colour preference and apparent weight...if it exists has little explanatory power. The reasons for colour preferences are themselves unclear.' However, lightness contrast, saturation and affective value together accounted for more than 99% of the systematically explainable variance due to colour (i.e. the residuals for each line from the width regression meaned for each colour, removing the effects of random variations and colour/width interaction).

The overall equation with the measures of width and colour mentioned above produced an r-squared of 99.54%, with nearly 95% due to width alone. However, the leverage exerted by the narrowest (0.15/0.2 mm) lines was clearly having an undue effect upon the regression line, so they were dropped from further analysis. This had the added advantage of making the rest of the data set more comparable by removing the more degraded stimuli. Without these lines the best descriptor for width was simply its log., confirming the implications of Wright's experiment (section 4.4). Also, as expected, the overall contribution of width decreased (to 73% of overall variance) whilst the proportion of the remaining variance covered by the colour measures increased (to 96.7%), leading to an overall r-squared of 99.1% without residual autocorrelation. A certain amount of random error was clearly present in the line specifications owing to the difficulty of precise measurement of the actual line widths and the sensitivity of the regression equation to small width differences. There was also potential for slight errors in the Munsell matching and in the interpolation and wholesale application to these subjects of the affective value figures. Given these problems this is clearly a very good description of the overall data.

The final equation derived (for lines wide enough for 'normal' colour discrimination) was:

Perceived Magnitude = -1.56
+ 3.44 x log. line width (mm)
+ 0.178 x lightness contrast (Munsell value steps)
+ 0.00035 x saturation (Munsell chroma^2 x value contrast)
+ 0.133 x affective value (Guilford and Smith's index)

(Alternatively, saturation and affective value can be combined in a 'colour' measure (saturation + 379 x affective value) with the same coefficient as saturation.)

There remains however the problem of intersubject variability. The analysis of variance showed that overall there were no significant differences in performance between the Farnborough and Glasgow subject groups. However, there were clearly considerable differences between individuals. Regressing each subject's individual scores rather than means is a relatively conservative test as on their own the individual figures tend to lack precision. This is because, with the necessarily limited sampling fraction, they were each computed on the basis of relatively few comparisons. However, over half the variance in them (52.4%) is still accounted for by the above regression equation. The relative diversity of subject's comments on the subject and the above-mentioned problems of variable specification imply that much of the variability between subjects was in the chromatic measures, with width and lightness contrast judgments being particularly consistent.

Returning to the regression of mean scores, the proportions of the total variance explained by the different factors were:

Log. line width		73.0 %
Lightness contrast		15.1 %
Colour measures		11.0 %
Unexplained		0.9 %
	TOTAL	100.0 %

Another way of expressing the relative significance of these variables is to look at the width of line of a given colour that would be of the same perceived magnitude as, for example, a 1 mm wide black line:

Test Colour	Munsell Specification	Equivalent Width
Red	7.5R 5/14	0.93 mm
Black	N 2.0	1.00 mm
Brown	5R 3/6	1.13 mm
Blue	7.5B 6/8	1.17 mm
Green	10GY 5/10	1.20 mm
Yellow	7.5Y 8.5/12	1.60 mm

This correlates interestingly with the aforementioned results of R.L.Williams (section 6.3.2) on the areas of coloured point symbols judged to be equivalent in size to a black symbol of unit area. Although the colour specifications he provides are tristimulus values, conversion to Munsell coordinates reveals the similarity of the two sets of colours: except for the brown (5.73), the largest Farmer difference between equivalent colour names is 2.64. The equivalent colour names yield the following comparison:

Colour Name	Width factors (current expmt.)	Area Factors (Williams' expmt.)
Red	0.93	0.994
Black	1.00	1.000
Brown	1.13	1.005
Blue	1.17	1.011
Green	1.20	1.026
Yellow	1.60	1.056

Williams suggests that the only practically usable differences are between yellow and green, and between these two and any of the others, which are somewhat more tightly bunched than in the present experiment. However the correlation between the columns of 0.975 is remarkable. It is difficult to see why there should be such a strong relationship other than that the effect of a prominent colour on line perception is psychophysically equivalent to enlarging the line (or placing it nearer to the eye), but it confirms the interrelationships between various quantitative effects of colour suggested by the literature (cf. chapter 6).

8.2.4 Cased Lines

The analysis of the scores for the cased lines presented greater difficulties. An overall comparison of the black-cased lines with the equivalent uncased ones shows that in most instances the casings increased the perceived magnitudes fairly slightly, and the overall effect of casing ($F(2,58)=4.24$, $p<0.05$) was far less dramatic than that of width or colour. The breakdown of this effect by line width shows that the only significant differences (Newman-Keuls, $p<0.05$) were at 1.4 mm (thin and thick > uncased), indicating that the width of the filling may sometimes be important. With regard to individual colours, yellow was affected the most (thick >

thin > uncased, $p<0.05$), and the only other significant differences were in green (thin > uncased) and brown (thick > uncased). Some casings even had a negative effect on magnitudes, notably on certain red lines, and with thick casings on green (table 8.3).

TABLE 8.3 DIFFERENCES IN THE PERCEIVED MAGNITUDES OF EQUIVALENT LINES OF DIFFERENT CASING STATES (BLACK CASINGS)

	Changes in Scores from				
	Uncased to Thin Casing			Thin to Thick Casing	
Width:	0.7	1.0	1.4	1.0	1.4 mm
Red	-0.19	+0.03	-0.07	-0.07	+0.08
Green	+0.07	+0.16	+0.27	-0.05	-0.21
Yellow	+0.14	+0.25	+0.22	+0.11	+0.35
Blue	+0.11	+0.05	+0.14	+0.03	+0.05
Brown	-0.05	+0.09	+0.14	+0.14	+0.05

Scores have been corrected for slight differences in actual widths. Thin casings are each 0.1mm wide, thick casings 0.3mm.

Clearly the strengthened edge contrast provided by a dark casing considerably boosts the prominence of a light-coloured line such as yellow, but with some other colours the assimilation or spread effect (section 5.4) emanating from an achromatic casing may have more of a 'dirtying' effect by reducing the apparent chroma of a line. Unfortunately, as table 8.3 shows, these effects do not increment steadily for each colour as the absolute or relative width of the casing increases. This is where the main problems of analysis lie: the overall interaction between colour and the three casing states is very significant ($F(8,232)=5.3$, $p<0.0001$), especially between thin and thick casings ($F(5,145)=17.86$, $p<0.0001$). There is also interaction between casing state and width ($F(2,58)=3.575$, $p<0.05$), as there is a tendency for increases in prominence due to casing to be more marked in the wider lines.

In regression of the scores for the black-cased lines, the deviancy of the unfilled (white) lines became immediately obvious, both in terms of large residuals (gross underprediction for thick-cased lines and overprediction for thins) and leverage (with scores considerably lower than those for equivalent filled lines). Consequently they were removed from the data set and treated separately. The scores for these particular lines had the highest standard deviations of all, as subjects were evidently not entirely sure whether to treat such hollow (or 'outline') symbols primarily on their overall width or their meagre contrast. However, a satisfactory regression equation (r-squared 99.2%) was derived for the mean scores. It can be expressed in either of two ways:

1) Perceived Magnitude = 0.167
 + 1.896 x log. line width (mm)
 + 1.401 x log. casing width (mm)

2) Perceived Magnitude = 0.167
 + 3.297 x log. line width (mm)
 + 1.401 x log. casing proportion

where 'casing width' is the combined width of the two casings, and 'casing proportion' is the combined casing width divided by the total line width (and therefore when logged is always negative). Thus the rise with width is somewhat less steep than for uncased lines, and both the casing width and the filling width contribute to the perceived magnitude of the line.

For the remaining (filled) lines, the best specification of total width was again found to be its log. For the colour measures, various different approaches were tried.

1) The casing and filling were treated entirely separately as if no interaction was occurring. This might be justified intuitively, by considering that the effects of assimilation are equivalent to spreading the jam more thinly without altering the actual amount of jam used. Different ways of generating the predicted scores were tried:

a) treating the lines as cased white lines (using the above formula) to which a coloured filling (predicted by the 'uncased' formula) had been added. The two figures were summed. This method does however involve double counting of width to a certain extent.

b) using the colour specification of the 'uncased' formula (for lightness contrast, saturation and affective value) to compute the separate colour magnitudes of the casing and the filling. From these an 'average' weighted by the proportions of the overall width they represent was derived and inserted into the regression equation with the log. of overall width. Alternatively, the separate colour magnitudes of casing and filling were predicted by the 'uncased' formula, multiplied by their respective widths, added together and subsequently logged. Using this measure an r-squared of 97.2% was obtained. Because the colour weight of black is the highest apart from red, this predicts an increase from the uncased scores for every colour except red.

2) A simulation was made of the assimiliation effects that occur when black casings are added to a line. Assuming that both lightness and chroma are affected, the result would always be a reduction in both (i.e. an <u>increase</u> in lightness contrast with the background), as black is darker than any of the filling colours and is achromatic. Overall lightness and chroma figures for the

lines were calculated by weighted averaging of the casing and filling values, with the weighting of the casing varied from its proportional width to find its best specification. Similar manipulations were independently made to the specification of the lightness contrast measure used to calculate saturation, and the appropriate lightness/chroma levels at which the affective value measure should be read off.

Many permutations were tried, and it was discovered that the best specifications of lightness and chroma were obtained by weighting the casings at half the casing proportion, i.e. as if half of each casing spreads into the filling. The lightness element in the saturation term should be treated in the same way, but for affective values this only applies to chroma, and the lightness values should remain unchanged. Substituting these revised values, with log. width, into the 'uncased' regression equation yielded an r-squared of 97.7%. This was better than those achieved by the other methods and specifications. However, as before, there were still considerable colour/casing interactions present in the residuals. The scores for red, yellow and brown lines were consistently overpredicted for lines with thin casings and underpredicted for those with thick casings. For the green and blue lines, the opposite situation prevailed. This is because, as casing and filling widths increase, the effect of casing tends to become more positive for red, yellow and brown, but more negative for green and blue. These anomalies were still present if the lines with thin and thick casings were regressed separately. Even within yellow, the colour with the most straightforward effect of casing, it can be seen (table 8.3) that for each casing level, the casing boost generally <u>increases</u> with overall width (i.e. with filling width), whilst this equation predicts a decrease.

Assimilation was also simulated in the alternative model in 1b above (the log. of the weighted sum of the casing and filling colour scores). Here the best specification was obtained with the same affective value measure as above, but no lightness spread and a chroma measure obtained by weighting the casings at 30% of their proportional width. Its explanatory power was marginally less (97.6%), but the predicted scores for yellow do increase slightly with filling width as the casing proportion (causing the chroma reduction) decreases. However, the effect is very slight as chroma makes little overall contribution to this particular equation.

Another possibility is that a measure of the effect of the casing (i.e. difference in perceived magnitude from the uncased equivalent) in each line could be derived from absolute casing and filling widths rather than proportions. This would seem to be logical as with an increase in casing size (at least), spread effects firstly weaken until eventually contrast is enhanced (Helson and Rohles,1959). The crossover point is unknown: subjectively, however, chromatic assimilation is still very clear on the red-cased yellow line 61, with a pair of 0.3mm casings and a 0.8mm filling. The significant overall width/ casing interaction is largely due to an increasingly positive effect of casings on prominence as filling widths increase, mainly for yellow (as noted above), red and brown lines. Work by Ginsburg (1981) suggests a possible connection between filling width and the visual significance of casings. He noted that the sensitivity of vision to sine-wave gratings at the threshold of perceivable contrast peaks at a spatial frequency of 2 cycles per degree, equivalent to a 30 minute spacing between casings, or 2.6 mm at 30 cm viewing distance. In other words, the visual system becomes increasingly sensitive to casings of a given contrast level as their separation (i.e. largely filling width) increases up to 2.6 mm.

However, the most basic problem is still that of the colour/ casing interaction which cannot apparently be explained simply by various specifications of lightness, saturation, affective value and spread. Perhaps separate equations are required for different colours/ colour groups. Intuitively, it might be considered that people's assessment of a cased line is dependent on whether it is 'chunked', and seen as an effectively homogenous whole, or whether the casing and the filling appear to be two separate and distinct items. Subjectively, the wider (less degraded) the casing and the filling are and the greater the qualitative colour contrast between the two, the clearer the separation will appear, but if there is a threshold on these scales (or, more likely, a fuzzier threshold zone) which divides the two modes of perception, it has apparently not been measured or defined. In order of decreasing subjective and measured colour contrast (Farmer) with the black, the current filling colours are

Yellow	(27.23)
Red	(17.98)
Blue	(17.03)
Green	(15.07)
Brown	(7.08)

Thus yellow, red and brown are not grouped together in this respect, but lie at the extremes, with green and blue in the middle of the scale. Alternatively, it might be considered that some non-linearity exists in the differential effect on each colour of the balance between lightness contrast boost and chroma loss. In terms of the ratio of chroma contrast to lightness contrast (with the black) the colours are ranked as follows:

Brown	6/1	= 6.00
Red	14/3	= 4.67
Green	10/3	= 3.33
Blue	8/4	= 2.00
Yellow	12/6.5	= 1.85

Again yellow, red and brown are at the extremes, where one of the contrasts is perhaps dominant. Clearly there are here potential problems of equifinality in attempting to uncover a complex relationship from a very limited sampling of colour space.

Whatever the causes of these behaviourally different colour groupings, the data set could be split for separate investigation. For yellow, red and brown fillings, the specification providing the best overall explanation was the same as for the full data set above (i.e. the variables from the 'uncased' equation, with lightness and chroma reduced by half the casing proportion and affective values read off from the adjusted chroma value and the original filling lightness). The equation, with an r-squared of 99.2% and negligible casing/width interaction in the residuals, is

Perceived Magnitude (yellow, red, brown fillings) =

$$
\begin{aligned}
-4.69 &+ 4.11 \times \text{log. width} \\
&+ 0.462 \times \text{lightness contrast} \\
&+ 0.000709 \times \text{colour}
\end{aligned}
$$

where colour = (-saturation + 728 x affective value).

The gradients are noticeably much steeper than in the uncased equation, implying in the width measure the added involvement of width in the casing effect. Overall this specification appears to be a 'best compromise' solution.

For green and blue, log. width alone explains 98% of the variance, but the best overall specification (r-squared 99.2%, no residual interaction) was still obtained from the same variables but with different coefficients.

Perceived Magnitude (green, blue fillings) =

$$
\begin{aligned}
-2.92 &+ 3.41 \times \text{log. width} \\
&+ 0.292 \times \text{lightness contrast} \\
&+ 0.00136 \times \text{colour}
\end{aligned}
$$

where colour = (saturation + 177 x affective value).

This implies that the increase in prominence due to width and lightness contrast is less steep than for the other filling colours, the width coefficient being similar to the one for uncased lines. However, the higher weighting of colour and especially saturation is necessary to predict the chroma/ prominence drop with thicker casings.

With both white and yellow infills, casing colour also has a significant influence. The relative prominence of lines of the same width and casing thickness varies with the colour magnitude of the casing, such that the prominence order of the casing colours was red, black, blue for the unfilled lines, and red, black, green for those with yellow fillings. As with the black-cased lines alone, overall width had significantly less effect on the scores of the unfilled lines. With the yellow fillings, there was a significant difference between the Farnborough and Glasgow subjects, with the latter showing a much less marked effect of casing colour which particularly affected the scores of the red-cased lines. In fact there was a slight tendency throughout the experiment for red lines to receive lower ratings from the Glasgow sample. As there is no evidence of a Scottish cultural bias against red, this could imply that, although lighting conditions were made as comparable as possible between the venues, slight differences may be critical with colours whose prominence relies largely upon chromatic insistence, such as red and red/yellow. If sufficient numbers of these lines had been used to allow meaningful regressions, a different factor- the affective value of colour combinations (Allen and Guilford,1936)- would enter into the equation.

In summary, it would seem that casings tend to boost the prominence of a line and that this boost increases

1) where the perceived colour magnitude of the casing greatly exceeds that of the filling

2) with increasing casing width

3) with increasing filling width.
The effect does however seem to vary with different hues such that prominence drops can occur in certain circumstances where the magnitude of the casing colour is still greater than that of the filling. This can be summarised by varying specifications of a regression model involving a weighting of casings in overall line colour measures at half their actual width. In real map contexts, however, casings may derive much of their significance from the continuity of strong contour, which cannot be apparent in a short line segment.

8.2.5 The Effects of Background Colour

In order to investigate the effects of the pale yellow and brown backgrounds upon the relative prominence of the lines it was necessary to run a continuation of the experiment. Given the results already obtained for the white background, the 0.15 mm lines were omitted and, because the appropriate width measure (its log.) appeared to be consistent and predictable, only one width of line was selected for each colour in each casing state (uncased, thin black, thick black and coloured). The particular width used was selected at random. Consequently a total of 23 lines were tested on the brown background, but only 22 on the yellow as the uncased white was almost imperceptible. Each subject saw two cycles on each background with the order of the lines randomised as in the experiment with the white background.

Background Colour	Number of poss. pairs	Number of pairs shown	Sampling fraction
Yellow	231	22 x 2 = 44	19.05%
Brown	253	23 x 2 = 46	18.18%
		TOTAL 90	

The specifications of the lines are presented at table

B.3. The only differences in experimental method from the previous experiment were the addition of a short practice session (6 pairs) at the start, and the mounting of the lines on thin clear acetate sheets backed up by Pantone paper (about A1 size) of the appropriate background colour to ensure appropriate adaptation. 15 people volunteered as subjects. Experimental sessions were conducted in daylight and were generally of 20-25 minute duration.

The addition of background colour obviously alters the achromatic and chromatic contrasts of the line. Further complications are also introduced with respect to interaction between line and background colours. Cartographic theory (cf. section 5.4) suggests the following points.

1) Dark and highly saturated line colours are altered less in appearance by their backgrounds.

2) With uncased lines, chromatic induction should reduce the apparent chroma of line colours of similar hue to the background and enhance the chroma of complementary hues. Therefore a chroma boost would be expected in blue lines on the yellow background and green lines on the brown.

3) Casings reduce interaction with the background and therefore maintain the distinctiveness of line colours against it.

4) Large areas of colour, such as backgrounds, tend to appear to be more saturated than their specifications would suggest.

The effect of these four would be to vary the contribution of the chromatic variables to the regression equations.

5) By reducing line/ background lightness contrasts, lightness induction, and consequently the lightness contrast coefficient, should be lessened.

6) By introducing hue contrasts, coloured backgrounds may enhance apparent distinctions between line colours of similar lightness and saturation.

7) Through the darkening of the background, more light

colours are likely to appear fluorent.

Background colours add further complications to colour measures as some line colours may have negative (lightness) contrasts with the background (e.g. yellow and white on the brown background here). It is also possible for a line to contrast with its background in chroma but not in lightness (e.g. a colourful yellow line on a cream background) and therefore have zero saturation contrast by our present definition. This may not however be too significant a problem as the added chroma is usually reflected in higher affective values.

Analysis of the comments made by the experimental subjects suggests that the factors contributing to the prominence of lines on coloured backgrounds are more complex and vary between individuals more than on the white background. Overall the influence of line width appeared to be considerably diminished: while 6 people still mentioned it as a major factor in their assessments, a further 6 considered it to be less of an influence than colour or darkness. Darkness itself was quite commonly cited (5 mentions), while the presence of casings appeared to be a very important secondary factor (10 mentions, including 4 for light fillings only). Colour strength received only 2 plaudits, but the 'vividness' of colours was seen to be important. Among specific colours considered to be prominent, red was again the most frequently mentioned (6), but shared this position with, surprisingly, yellow, which was popular when outlined by black and red casings and on the brown background (against which it is in negative contrast). Black and blue (preferably cased) also featured significantly in the commendations. However, uncased yellows were considered to be particularly poor by some of the subjects, their blurred outlines being specifically mentioned, and white was also commonly disliked and found to be difficult to assess. Overall, width would appear to be less important and 'insistence' more so than on the white background.

Perceived magnitude scores were calculated in the same way as for the white background. The mean scores for the sample are displayed at table B.4. On the brown background, the scores for lines 12 and 23 (thinly cased and uncased white) are clearly much lower than the rest, including the yellows which actually contrast less in lightness than the whites. Overall the results appear to be less consistent than for the white background, with differences between subjects being significant on both backgrounds, and individual lines behaving in unexpected ways. For example, while the expected chroma boosts for uncased blue (yellow background) and green (brown background) appear to be reflected in the line rankings (table B.5), the uncased yellow was far more prominent on the yellow background (ranked 9/22) than on the white (ranked 15) or brown (ranked 20), suggesting a chroma boost, or particular fluorence, on the same-hued background.

This lack of consistency is again apparent in the regression equations. For uncased lines on the yellow background, log. width and lightness contrast together accounted for 79.8% of the variance in the mean scores, but none of the possible specifications of chromatic contrast/ saturation added significantly to this, except affective value, which raised the level of explanation to 99.9%. The equation is: Perceived Magnitude= -3.3 + 4.66 x log. width + 0.281 x lightness contrast + 0.356 x affective value. (N.B. These coefficients are not directly comparable with those for the white background, as the scores are standardised across the range of the variables present in each case.) The best obtainable equation for filled cased lines on the same (yellow) ground involved just log. width and lightness contrast (r-squared 80.8%). No other variables contributed significantly to the explanation, and there was no systematically explainable variance remaining in the residuals. This implies that the mean scores for these lines either reflect considerable random variations between individuals, or are the product of decisions made from very different criteria.

On the brown background width was of less overall significance, explaining only 16% of the variance with the uncased lines (white excluded). The best specification of lightness contrast retained the sign with negative contrasts, implying that a colour with a small negative contrast with the background is less prominent than one with none at all. This is also suggested by the exceptionally low ratings of the white lines. The best model by far for the uncased lines is: Perceived Magnitude= -0.839 + 1.22 x log. width + 0.154 x lightness contrast + 0.044 x chromaticity contrast (r-squared 99.8%). Chromaticity contrast is the straight line distance between line and background colours on the Munsell chromaticity plane and, as the background hues are of weak chroma, is consequently largely a function of chroma contrast. With the filled cased lines, the same variables also compose the best equation, except that the chromaticity measure is reduced by half the casing proportion. The equation is: Perceived Magnitude= -0.536 + 2.18 x log. width + 0.12 x <u>filling</u> lightness contrast + 0.057 x chromaticity contrast (r-squared 92.5%). But for the steeper rise with width, this model is not dissimilar to the uncased version, demonstrating greater consistency of judgments than on the yellow ground. The lower coefficient for chromaticity which might be expected on the grounds of reduced chromatic induction was however not apparent.

Overall the comments made by the subjects are confirmed by this analysis- namely the reduced significance of width (especially on the brown background) and colour strength (i.e. a saturation measure involving lightness contrast), the latter being replaced by simpler chromaticity contrast measures on the brown background, where the problem of lighter colours merging into the background (M.Wood,1968; section 5.4) is less apparent. However, despite their prevalence in the comments, casings appear to be less significant in terms of lightness. No intuitive reason can be supplied by the author for the different/ absent chromatic measures in the yellow background equations. More observations would be required to fix the coefficients precisely enough for them to be demonstrably reliable, and given the amount of inter-subject variance both in the comments and in the scores, a better specification would not necessarily result.

8.3 THE CONSPICUITY EXPERIMENT (EXPERIMENT 2)

Having examined pairs of line segments against a featureless background, the next stage was to investigate their attention-getting properties in a more realistically complex, but still schematised and controlled, context. Reference has been made (section 6.2.2 and 6.2.3) to various procedures and measures which have been used to determine the relative conspicuity of objects in laboratory experiments. Woodworth and Schlosberg (1955) summarise the methods as follows:

1) verbal report, where the subject is asked to state which of a small number of objects equidistant from a controlled fixation point stands out the most clearly. Obviously each object must be uniquely identifiable by a simple verbal concept (e.g. hue, relative position).

2) recall, where for example the subject, unaware of the purpose of the experiment, is asked to thumb through a magazine and is then asked what he noticed- i.e. what items had caught his attention and lodged themselves in his short-term memory.

3) direct observation of the subject's locus of attention by eye movement analysis, from which the aforementioned index of relative fixation rate can be derived. With a complex display, however, eye movement records present many interpretational difficulties (Yarbus,1967; Engel,1971), especially as covert attention shifts may also be present (Kinchla et al.,1983), and, as Hill (1975) stated, looking is not seeing. In any case, the requisite equipment was not available for the current experiment.

4) recognition, where subjects are presented with an array of objects and then asked whether a specific object was present or absent in the array. In some cases, as in Engel's (1969 etc.) determination of 'conspicuity areas', the target object or structure is prespecified. A more commonly-used measure in this scenario is the time taken by the subject (or the number of fixed-length flicker exposures required) to find the target in visual search.

In studies involving the direct measurement of search times (e.g. Farmer and Taylor,1980; Fulton et al.,1980; Carter and Carter,1982) the task becomes one of comparing current sensory information with an image or verbal concept (e.g. a green square) held in short-term memory, which may involve internal determination of attention through spatially-sequential scanning. This does not correspond with the process of search in the route planning task as described in chapter 3. The requirement for the current experiment was to determine what line characteristics would be likely to attract the map reader's attention, where it is externally determined, in his search for contenders for possible routings, which cannot instantly be identified as 'correct' or 'incorrect'. A strategy to eliminate internally determined scanning was required.

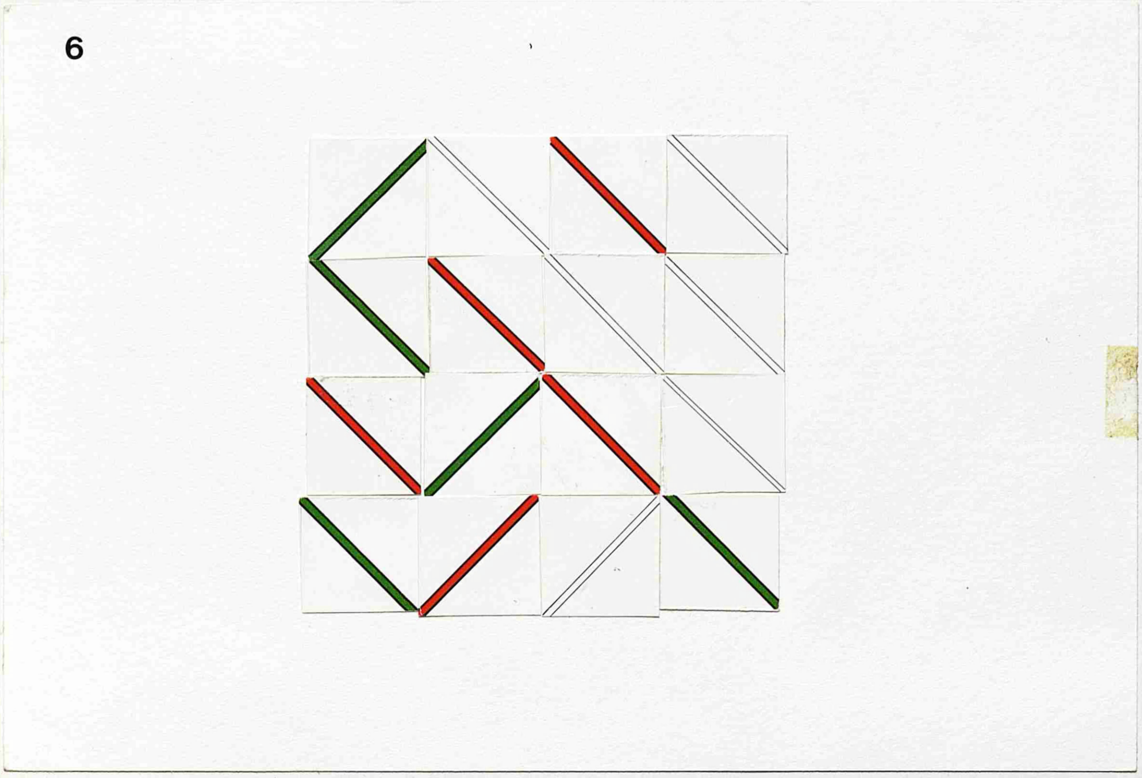
Other than making the subject unaware of the purpose of the experiment (Woodworth and Schlosberg,1955; Fulton et al.,1980), which may seriously reduce the amount of information obtainable from it, two strategies suggested themselves:

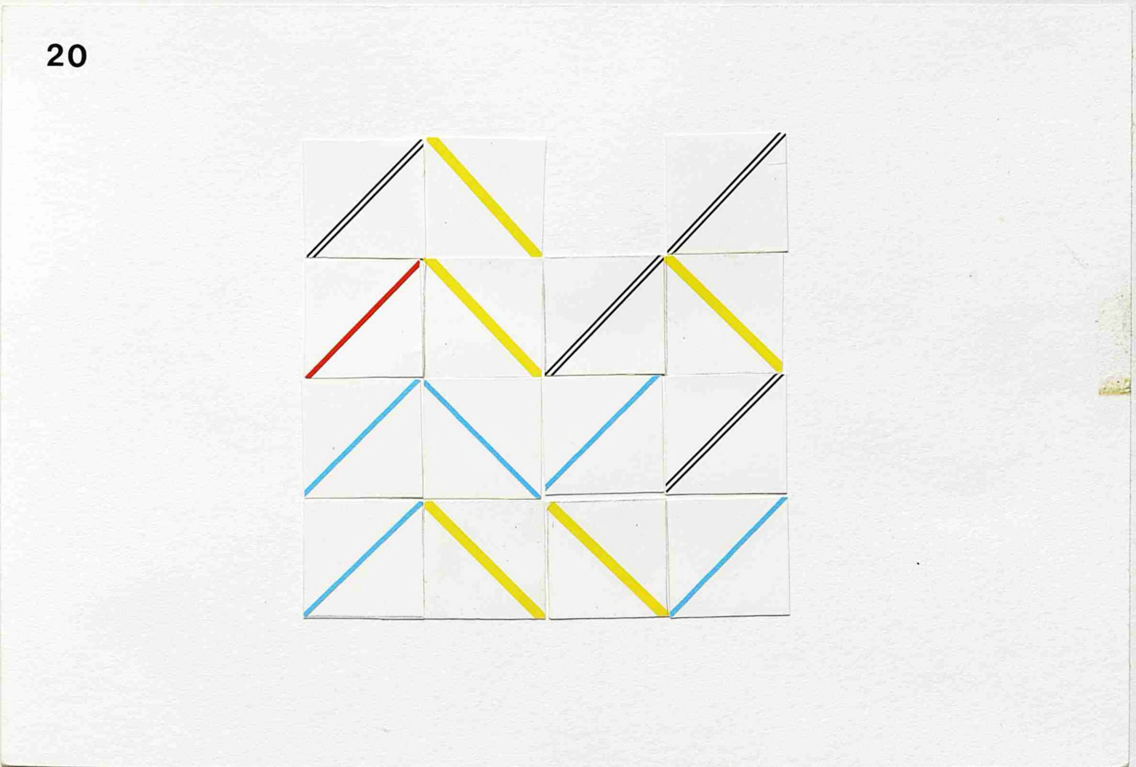
1) The use of tachistoscopically presented arrays of line symbols in single exposures short enough to suppress scanning schemata by preventing eye movements. Fixations on graphical displays generally last between 150 and 600ms (mode: 230ms), so exposures of 74-200ms have been used (Woodworth and Schlosberg,1955; Engel,1969; Dobson,1979a; Kahneman and Henik,1981; Phillips,1981). Subjects could consequently be legitimately requested to report the presence or absence of a pre- or post-presented target, or asked to describe the general or most immediate colour sensation they obtained from the arrays.

2) The use of a visual search task within an array of coloured lines, where the target is defined not by its own form, but by its relationship to the rest of the array. Thus the task necessitates the perceptual organisation of the display by parallel global processing. A measure involving search time was attractive as it is a well-tried index providing high-order data, and this was consequently adopted.

The structure used was a 4x4 array of line segments containing four different classes of line. Three of these each occurred in five cells of the array, and the remaining cell was filled by a single line of the remaining class (figure 8.2). The subject's task was to locate the line which only occurred once. Search time should be dependent on how much the target emerges spontaneously by preattentive perceptual organisation before serial search has to be initiated. In theory the emergence of candidates for subsequent scrutiny is determined by stimulus dimensions. Treisman (1982) talks about a global configuration at an early stage of processing to which all the objects' dimensions contribute according to their saliency. At this stage, she says, the target should emerge spontaneously when it is locally

Figure 8.2 EXAMPLE STIMULUS ARRAYS, EXPERIMENT 2





salient as figure against ground, with local colour contrasts being of particular importance in this context. Increasing homogeneity of distractor (non-target) items should also enhance emergence, more by enabling their rapid rejection, through reducing the number of discriminations required, than by direct guidance of search (Farmer and Taylor,1980). It might also be considered that the grouping containing the lines with the highest perceived magnitude would be the most conspicuous, on which attention would alight first, including the target itself if it is sufficiently excess in magnitude.

The aim of the experiment was consequently to discover which dimensions are the most salient in preattentive processing and how they operate, and whether emergence is purely a matter of qualitative distinctions, or whether the perceived magnitude or the 'attention value' of the stimuli have any separate significance.

Relevant stimulus variables were the width, colour and casing state and orientation of the lines and the position of the target on the array. Lines of three widths (0.7, 1.0 and 1.4 mm), all six colours and two casing states (uncased and thick) were used, plus thin- and thick-cased white. All of the lines used were on the white background. Maintaining a uniform orientation of the lines across the array would have caused distracting patterns, so orientation was controlled by cutting the line segment boxes down to squares and using diagonals, enabling either diagonal to be selected by a 90° rotation of the box. Thus potential problems of differential sensitivity to to horizontal and vertical planes were not introduced. Target position was controlled by making up four different arrays for each design (i.e. composition of line types) each presenting the target in a different row (in a randomised column). 20 different designs were used for a pilot study to test the methodology, and therefore 80 arrays were constructed. Various combinations of

1) target/ background distinction or confusability
2) target/ background differences in perceived magnitude
3) background homogeneity or heterogeneity

were used to create arrays with the best expected emergence (nos.1-5), the worst (6-8), and to examine variations in individual dimensions and specific interesting combinations (9-20). The arrays each measured 7cm square.

The pilot study was carried out at the RAF Institute of Aviation Medicine. The arrays were mounted on 6x4" white cards and displayed in a Behavioural Research and Development Ltd. 2-field 'Cambridge' tachistoscope with a field luminance of 5 ft-L. and a viewing distance of 37cm such that the array subtended a visual angle of about 11°. The rest field contained an array with a random configuration of line types, the numbering system for identifying the cell containing the target, and an orange cross at the centre. The subject fixated the orange cross and controlled the onset of each trial by pressing a button. After 300ms the trial array appeared, and the subject pressed another button as soon as she had found the target. This signalled the return of the rest field, which acted as a backward mask suppressing the image of the test array in order to prevent premature button-pressing. The subject was then asked to report verbally the position of the target and supply a confidence rating on a scale from 0 to 10 for this judgement. The experimental session included 10 practice trials followed by 80 assessed trials. Each array therefore appeared once, the order of presentation being randomised.

11 volunteers (7 female and 4 male) acted as subjects. Each had normal colour vision as assessed by Ishihara tests.

8.3.1 Results

The distribution of search times of correct answers for each array shows that for most of the arrays there are one or two outlying observations with considerably higher search times than the remainder. These were not traceable to consistent slowness by particular subjects. Because of the influence these had on the mean times, it was not really feasible to use the t-test for two paired samples to assess the significance of the differences, so the non-parametric Mann-Whitney test of medians was used. The ranking list of median times for each array is presented at table C.2 with the array specifications at table C.1. The expected extremes were clearly grouped at opposite ends of the list. Testing the differences between arrays adjacent in the list revealed 3 major breakpoints. Array no.8, where all the lines were blue and the target had to be picked out by width alone, was by far the slowest. Next came a group of four, namely numbers 6, where amidst high magnitude distractors the target was distinguishable from one of the background classes by width alone, 7, which had the smallest target/ background class colour difference (brown/black, Farmer difference 7.08), and 11 and 18, which were similar set-ups to 6 but with successively higher magnitude targets. The four slowest arrays also shared 22 errors (all made on low-confidence judgements) of the 26 made in the entire experiment. Clearly the colour difference between the target and the most similarly coloured background class is critical. Below 18 the breakpoint in search times would appear to mark the gradual beginnings of the spontaneous emergence of the target.

Overall it is clear that the most important impediment to emergence was the camouflage of the target by same- or similarly-coloured items. Over this range of each dimension, width alone (D/WW=0.29-0.3) appears to be inadequate to provide perceptual segregation, with 4 of the slowest 5 arrays requiring width discriminations for task solution. A similar situation applies for the smaller colour differences. For example, arrays 19 and 20, with a red target, are identical except that the former includes a class of brown distractors (Farmer difference 11.07) of the same width as the target, resulting in much slower ($p<0.001$) search times. (Bracketed figures below represent the significance of the difference between median search times for the array numbers given.) The stimuli are however significantly degraded at greater widths than in experiment 1, where subjects had the luxury of relatively prolonged and concentrated scrutiny, such that black and brown were more distinguishable when each was 1.4mm wide (despite the black casing on the brown) than at 0.7mm (17 < 7, $p.<0.001$). The next smallest colour difference between uncased lines (blue-green, 14.25) does not appear to present a discrimination problem (9=10), although there are no arrays containing 0.7mm lines of both colours. A minimum difference of 14 Farmer units would appear to be a reasonable design guide.

Where there is no real target/ non-target confusability, the homogeneity of background items in terms of colour can significantly reduce search times: for example, 1, with a background of uniform width but different colours, was significantly slower than 4 ($p<0.01$) and 5 ($p<0.001$) which share the same target but have more uniformly-coloured backgrounds. Distractors of homogenous width but a wide variety of colours do not tend to reduce search times as they leave little scope for a target highly distinct in colour from each of the background classes (13 > 3, $p<0.001$). Homogenous background lines are however of little use on road maps, where the usual need is not to highlight one particular class at the expense of all others, but to depict a hierarchy of several classes.

The magnitude of the target is also important, but again with respect to colour rather than width. No significant differences were recorded between the two pairs of arrays which were each identical apart from target width, on relatively homogenous (2=3) and heterogenous (12=13) backgrounds respectively. On four arrays the target was the highest magnitude line largely because of its colour (red or black)- on 1, 4 and 5 where the red emerged most strongly, and on 14, where the black target performed better than a lower magnitude green on a similarly heterogenous background (14 < 12, p<0.01). The red target, however, outperformed them both (1 < 12 and 14, p<0.001) emerging clearly on 20 (=1) despite the higher width-based magnitude of one of the distractors.

Unfortunately, however, the pilot study did not produce sufficient consistency of results to create significant differences between some of the more specific individual variations which had been built in to test the sensitivity of the methodology. An example of this is on 15, where a red filling was added to one of the distractor classes (to lure attention away from the target) while the array was otherwise identical to 14 (15=14). Similarly the effect of casings per se was not apparent. This is also particularly difficult to quantify, as where the cased line is not 'chunked', the contrast due to character is clearly greater than is revealed by some colour contrast figure 'averaged' across the whole line. Nor was it possible to derive a quantitative estimate for the relative sensitivity of the task to colour difference and width difference measures. Perhaps too much was expected from the search time measure given so many potentially relevant variables. It is possible that task sensitivity was reduced by the cognitive load imposed on subjects who made sure they could supply a number for the target position before pressing the button.

A further complication was caused by an extra source of variance. The effect of orientation (which was randomised between its two states rather than controlled) was unexpectedly significant, with times for orientation 2 being quicker ($p<0.05$). Given these added constraints, and as time was running short, it was felt to be desirable to concentrate on experiments involving the actual making of route choices from maps (experiment 3) rather than just one specific influence among many upon those choices (i.e. conspicuity).

In summary, the main conclusions which can be drawn from this study are that over this range of dimension, the effects of colour are much more important in perceptual organisation than those of width. With uncased lines of equal width, a minimum colour difference of about 14 Farmer units between each class is desirable to prevent inter-class confusion, perhaps more if lines much narrower than 0.7mm are used. Colour magnitude also appears to be an important influence upon the conspicuity of lines upon a white background.

9. THE ROUTE CHOICE EXPERIMENT (EXPERIMENT 3)

The final stage in the sequence of experiments was to test the results of the laboratory work in realistic road map contexts using realistic route planning tasks, and on a representative sample of road map users. Compared to the previous experiments with line segments, four major additional sets of factors are introduced to the relative perception of line symbols by the map context:

1) configural effects. In networks composed of several different line types, the layout of each type can affect its prominence through continuity, closure and overlay at cross-roads. Higher road classes tend to occur in longer uninterrupted stretches, link up better, and in cartographic convention should take precedence at cross-roads.

2) bendiness. In general on equivalent terrain lower class roads tend to be more sinuous, but this is by no means universal.

3) symbolic associations and cartographic conventions (such as the use of blue lines to depict motorways). These may guide the map user's search for a route and affect his interpretation of what he sees in the process.

4) the presence of other information apart from the road symbol, such as road numbers/ class letters, and other indications of road suitability such as gradients and the extent of built-up areas.

Together these factors could seriously weaken the utility of the results of the experiments on perceived magnitudes (subsequently referred to as 'magnitudes'). For example, casings may well contribute far more to the apparent prominence of more continuous lines than short segments, by emphasing the continuity of contour. Consequently an experiment was designed to attempt to assess the relative significance of these factors in inter-urban route choices made on maps. Additionally, various symbols for the most common types of restricted access junction were tested.

There is a choice in such circumstances between conducting personal interviews or circulating questionnaires (by post, or perhaps by depositing them in places such as map shops, libraries and tourist centres (McGrath and Kirby,1970)) or a combination of the two. Choosing the former option made an expensive long print run for the experimental maps unnecessary, and allowed the experimenter to procure a certain amount of standardisation of the conditions (e.g.illumination) of testing, time taken over route choice, and full completion of the questionnaire. It was also likely to be a more reliable method of obtaining a representative sample, given the inevitable bias introduced by non-respondents to circulated surveys. Also, thanks mainly to the Automobile Association, suitable interviewing locations were made available.

9.1 EXPERIMENTAL MAP DESIGN

Two sets (A and B) of three map extracts (1,2 and 3) were designed and produced as colour proofs using the Cromalin process for optimum faithfulness of colour reproduction (rear pocket). Each map was reproduced identically on both sets, except that the symbolisation of each road class was different. Half of the experimental subjects were shown set A, while the other half saw set B. In order to get a realistic network shape, an existing network was used- broadly the major (M,A and B) roads of the East Midlands of England, as found in the Department of Transport Present Year Network File (PYNF). PYNF information used in these experiments was corrected for obvious errors. On the maps, however, the classification of the roads and the names of the towns and villages were changed to prevent recognition, in order that people were entirely dependent upon the map information for their knowledge of the area. Consequently the experiment was not complicated by varying levels of pre-existing local knowledge.

In terms of the line symbol series, the main problem was to examine the performance of familiar, 'conventional' designs versus unfamiliar but 'improved' ones, created by using the results of Experiment 1 (for a white background). Each scheme contained five road classes. Overall five different design concepts were used:

1) a 'conventional' scheme, as used partially or in entirety on almost all current British road maps:

Class 1	Blue,	thick casing	('motorway')
Class 2	Green,	thin casing	('primary route')
Class 3	Red,	thin casing	('other A roads')
Class 4	Yellow,	thin casing	('B roads')
Class 5	White,	thin casing	('unclassified')

As mentioned above, this scheme has been criticised because of the relegation of red.

2) an 'improved' scheme, utilising width and colour differences to create a smooth decline in predicted magnitudes through the series, whilst maintaining distinctiveness with widespread colours. However, in order that its appearance should not be too unfamiliar, some colour conventions would be incorporated.

3) an 'improved' scheme, as above, but using an increasing lightness series, combined with the natural harmony of a part-spectrum (purple, brown, pink, orange, yellow), to see whether changes in colour alone (not width) carry sufficient quantitative connotations to make the series effective. This would be particularly useful in certain maps, such as those showing road travel speed, where the classification of a road may change

a) along its length, while some indication of the road's continuity is still desirable, and

b) from time to time, as with conventional techniques revision of colour is much easier than amending width. Cased lines cannot be used for these colours, as there is no adequate formula for predicting their magnitudes.

The 'conventional' scheme would be compared with each 'improved' scheme in turn.

The remaining two designs were concerned primarily with the effects of different depictions of changes in road quality along a link, such as from single to dual carriageway, and certain effects of casings in a map context.

4) Dual carriageways are represented by a change in colour, the symbol used being a 2-colour line (yellow with red casings) similar to that used on the OS Routeplanner and by foreign map producers especially. As this is slightly narrower than the brown symbol for single carriageway class 1 roads, does it still appear to be more important? The classes are in order of declining magnitude, such that a cased red road (whose magnitude was reduced by the 'dirtying' effect of the casing) occurs below an uncased one of the same width and colour. Does the casing however enhance the prominence of the line in a map context despite the superior continuity of the uncased version?

5) Dual carriageways are shown by a thick casing (and concomitant increase in width for the lower class roads), on colours whose magnitudes were affected in very different ways by thick casings in the previous experiment (red, green and yellow). Because of the continuity of filling colour at single-dual joins, the steep magnitude gradient, and the highly distinctive colours used, the map has a more structured appearance than design 4 above. The green lines are uncased to see if character contrast in itself has any influence.

It was then necessary to relate these series to variations in the representation of road numbers and settlements, for which there were two alternative formats:

1) hierarchical. In this case, road numbers include their class letters, and hierarchy is enhanced by a

progression from coloured boxes through outline boxes to unboxed numbers, with the lowest class unnumbered. Settlements were divided into three categories:

-upper case names, built-up area shown

-lower case names, built-up area shown

-small lower case names, point symbols only.

2) uniform, where all road classes have road numbers without class letters in outline boxes (in order not to be mistaken for distance figures). All settlements are marked by uniform point symbols and lower case names.

The following arrangement was adopted:

Map	Set A		Set B		Road nos. etc.
1	'Improved'	(2)	Conventional	(1)	Hierarchical
2	Conventional	(1)	'Improved'	(3)	Uniform
3	Dual- colour	(4)	Dual- casing	(5)	Uniform

(The precise specifications for each map design are given in Appendix D.)

Thus the effects of the road symbolisation can be assessed by comparison across sets (A1/B1,A2/B2 etc.), while the influence of road number and settlement representation can be investigated directly by comparing B1 to A2. This is possible because the network and classifications used in maps 1 and 2 are identical. Map 2 is simply map 1 upside-down and with different place names and road numbers, and, obviously, different symbolisation. Interestingly enough, during the experiment only one person noticed this. However the two maps were not generally seen together as they were viewed one at a time with the others masked off.

A further advantage of this arrangement is that interviewees are not confronted with the most unfamiliar designs (A3 and B2), and unordered road numbers and settlements, when they are first shown the maps.

9.1.1 Allocation of Road Classes

It was intended that the attractiveness of a road class on any map could be measured simply by the total length of road of that class used in the interviewees' chosen routes. In order to do this, it was necessary to allocate the classes to links in the network so that, symbolisation apart, every road class had an equal chance of being chosen for an equal distance. This was (clearly) a complicated process, which involved several stages.

Subjects were to be asked to choose one route off each of the three maps on the set they were shown. It was necessary firstly to locate the starting and finishing points of the routes. At least two point pairs were required for each map, in order that people were not effectively asked to choose the same route on maps 1 and 2 which use the same network. More than two pairs would lead to undesirable subdivision of the interviewed sample. In order to make maximum use of the rectangular map area with minimum overlap between the two routes, points near to each pair of opposite corners were selected. Their specific locations were chosen in order to maximise the number of realistic route choice options available in the network.

The next stage was to test people's visualisation of the networks to see what routes they would choose if there were no graphical differentiation between the roads. Plots were consequently prepared of each of the two networks (maps 1/2 and map 3) using monochrome lines of uniform gauge, and 20 subjects were asked to trace the line that they considered to be the shortest distance on the network between the two endpoints. In each case, 20 subjects visualised the route, 10 in each direction to allow for asymmetric influences on choice such as the perceived directional continuity of the line.

From the results of this test, each link on the network could be allocated to one of three categories.

Category 1 includes those links chosen as part of their minimum distance routes by the vast majority of subjects. For the most part, their choices were very consistent, so that for three out of the four routes, at most two basic lines covered 90-100% of the choices, and these were placed in category 1. For the fourth route (map 3, top left to bottom right), one line was chosen by 15 subjects (75%), but the remaining choices were too diverse to be incorporated into category 1.

Category 2 contains all other links that, by virtue of their position and/or orientation (<45° divergence from the straight line route), were contenders for selection in either of the routes on each map.

Category 3 was formed by the remaining links on each map.

The following stage was the most difficult. It involved allocating equal lengths of roads of each of the three categories to each of the five road classes, while maintaining a realistic choice between different road classes at every point on every route, and at the same time producing a map that looked fairly realistic in terms of the relative continuity (length of stretches) of the road classes. Map 3 involved a further allocation process as the length of dual road was divided between the remaining four 'administrative' classes in declining amounts, as would be the case for Primary, other A, and B roads in Great Britain. For each map there is consequently a balance between the classes in each category- over the perceived minimum distance routes, other realistically usable routes, and across the map as a whole. So, for example, if everybody chose the perceived minimum distance routes, the aggregate lengths of each road class used would be equal. Any differences between these lengths could be confidently attributed to the symbolisation of the roads, road numbers and settlements.

Somehow this improbable objective was achieved: in the final design the largest discrepancy between the actual and optimal lengths of any road class in any category is 0.6 Km, or less than 1.5 mm at map scale. There is also a realistic choice between different road classes all the way along each route, except sometimes at the very ends (e.g. where the terminal point has had to be placed half way along a link), but this has been compensated for in the calculations. The maps are also reasonably realistic in terms of the relative closure and continuity of the different classes.

9.2 CONDUCTING THE TESTS

9.2.1 Sampling

In order to establish the external validity of the experiments, it was necessary to seek a representative sample of British motorists/navigators, i.e. all users or potential users of road maps. Previous research by A.Morrison (1974) has suggested which measurable personal factors should be taken into account in order to make the sample representative. While age, sex and colour blindness had no significant influence on the quality of routes chosen from his experimental maps, level of education, frequency of map use and social class did.

Unfortunately, no data are available on the personal characteristics of non-driving navigators nor for the educational level of drivers. However by combining data from the 1980 General Household Survey (reported by Department of Transport (1985)) and the 1981 Census, it is possible to develop a profile of the driving population with respect to age, sex and social class. Consequently, a stratification of the sample on the grounds of social class appears to be both desirable and feasible. While to a certain extent social class is in itself only a surrogate for differences in intelligence, education and driving experience, it appears to be quite a useful one (A.Morrison,1974), and has a very strong relationship with educational level (OPCS,1982). The proportions of drivers in each social class are shown in Appendix D. The aim of the interviewing was to obtain for each map set a sample of 100 people representative of this range of social classes.

In all, 248 interviews were completed between November 1985 and February 1986 at the following locations:

AA Centre, St.Ann's Place, Manchester	(97)
AA Centre, The Headrow, Leeds	(45)
Smiley's (tyre/exhaust centre), Partick, Glasgow	(57)

It should be noted that many of those interviewed in Manchester and Leeds were not actually AA members. Illumination at each place was provided by fluorescent strip lighting. The remaining interviews were carried out in generally more relaxed circumstances at various locations:

Aberdeen area	(26)
Glasgow area	(13)
South East England	(7)
Swansea	(3)

Illumination for these was either daylight or fluorescent strip lighting.

These sub-samples varied considerably in terms of their social composition. Manchester, the largest of them, was biased towards social classes I and II and retired people, while half of the Aberdeen subjects were of class I (national average 4.4%). The remaining sub-samples were broadly representative, although this was due partially to selective interviewing in order to compensate for the imbalance caused by the Manchester tests. After 248 interviews it became possible to make up the quotas required for the two representative samples of 100 each. Thus 48 interviews had to be discarded, 24 from each map set, and these were chosen at random from the overrepresented social classes. The resulting sample of 200 is also remarkably representative of drivers in terms of age and sex. These two samples will be referred to below as:

-the full sample (248 subjects)

-the representative sample (200 subjects).

9.2.2 Experimental Procedure

Subjects were interviewed as they were leaving the AA Centres having completed their business. At Smiley's the tests were undertaken in the waiting room. In each case, the questions were asked verbally and the interviewer completed the questionnaire. Questions were asked in the order they appear on the questionnaire (overleaf), although sex and age were determined visually. For each subject, the journey to be attempted on each of the three maps was selected from one of the four possibilities: A to B, B to A, C to D, or D to C. This selection was determined by a balanced experimental design constructed subject to the constraints that

1) AB (A to B) or BA (B to A) on map 1 must be followed by CD or DC on map 2 and vice versa, so that subjects were not asked to choose the same route twice, and

2) the direction for map 3 was either of the two not previously selected.

Consequently the 124 route choices in the full sample for each map consist of 31 in each direction for each map

NUMBER: Location/date:

Sex:
Age: <25 25-44 45-64 >64

1. Do you drive?
2. Do you ever plan routes or navigate for other drivers?
(If answer to 1 and 2 both NO, terminate interview.)

3. If you were planning a journey over unfamiliar roads, would you normally use a map? (If NO, then jump to 6.)
4. Roughly how many times a year do you use maps to plan a route?
5. Are there any particular maps that you tend to use?

6. What is your occupation?
(If u-e, last occupation)
7. Are you at all colour blind?

Here are 3 road maps. On each of these I will ask you to choose a route between two marked places. In each case please choose the route that you would consider to be the quickest (by car) to your destination. Please indicate the route by reading out in order the names of the places it goes through.

MAP SET: Junc. cycle:

ROUTE 1 option: Key ?:

8. Are there any particular reasons why you chose that route?

ROUTE 2 option:

9. Reasons:

ROUTE 3 option:

10. Reasons:

11. If, when following your chosen route, you came across a signpost directing you to your destination along a different road, would you generally follow the signpost or stick to your own route?

12. These are symbols for various types of road junction where not all turns are possible. Please say whether or not it appears that you can make the indicated turn, or if it is not clear.

1 2 3 4 5

(100 and 25 respectively for the representative sample).

Subjects were asked to choose the route they considered to be the quickest, and were encouraged to give reasons for their choice. Those who asked for a key to the road symbols were told that the roads were 'classified in the order of their apparent prominence.' Settlements had been positioned on the map so that each possible route could be identified by the names of the places it passes through, while these names could further be uniquely identified by their initial letter(s) for quick and accurate coding. This was necessary as some subjects chose their routes very quickly. The average duration of the whole interview was about 10 minutes.

9.3 MAP USE CHARACTERISTICS

Before evaluation of the chosen routes, some consideration will be given to the other information that the interviewees supplied about their route planning practices. The most recent profile of the habits of the map user in the available literature is that of A.Morrison (1979a), who draws upon surveys undertaken mainly during the 1960s and early 1970s. Consequently the representative sample of 200 from the current experiment should provide a reasonable update to this information.

Of the 200 subjects, only four were non-driving navigators. This should not be taken to imply that only 2% of road map users are non-drivers, as the interview locations used were likely to bias the sample towards drivers. However, of the 196 drivers, 187 (95.4%) said they used maps to plan journeys over unfamiliar roads, although 8 of these said it was not their normal practice, at least in this country. Other route planning methods used were prepared AA/RAC routes (8), asking people (6) and setting off in the general direction and looking for the signs (4). 100 of the drivers (51%) said they did also navigate for other drivers, but many did so only rarely. The proportion was higher among women drivers (69%) than men (44%).

The mean number of map uses per year for route planning by map-using drivers was 16.1 (15.4 for all drivers). This represents a slight increase from Morrison's figures. The distribution was obviously highly skewed, with a mode of only 2 uses per year and a median of 5. 40 motorists (21%) used maps more than once a month, including four who would normally use them every working day. There is no significant relationship between frequency of map use and sex, age or social class (overall), although drivers of classes I,II and III (combined) do tend to use maps more often than the rest ($t=2.736$, $p<0.01$).

Only 54 respondents (27%) asked for an explanation of the symbols at any stage, 23 on set A and 31 on set B, of which 22 (71%) enquired on being shown the unusually coloured map B2. The level of key usage was strongly influenced by both social class (chi-squared 15.37,4 df, $p<0.01$) and frequency of map use (chi-squared 22.94,5 df, $p<0.001$), with peak usage in high social classes and medium map use levels. This last point is especially interesting given the strength of the relationship. Perhaps the most irregular map users tend to view map reading as self-explanatory, while the most regular users are more accustomed to different map designs and have less need of explanation.

If improved maps are to improve the routes people take, then they need to have confidence in the routes they plan with maps so that they will stick to them rather than diverging at the sight of an incongruent signpost. In cases of apparent discrepancy between route and signs, 102 of the 191 map users (53%) said they would normally stick to their own route, 49 (26%) would follow the sign and 40 (21%) said that their reaction would depend on the specific situation encountered. Of these, 13 would stop to check the map if possible. For the others, the most common factors considered were whether the posted road appears to be major and/or new (10), and their confidence in their own planning of that particular route and the map used to plan it. People's responses to this question do not appear to be linked in any way to the noted personal characteristics- perhaps it is more a question of temperament.

9.4 ROUTE ANALYSIS

The actual routes chosen by the representative sample (n=200) are shown on flow-line maps at figures 9.1-9.6. Various details are tabulated below, for which some explanations are required.

Travel times were calculated for each network link using speeds within the PYNF ranges of 24-hour average speeds for each road type based on 1976 data. The following were used:

Motorways	100 Km/h
Other dual carriageways	85-95
Rural 2-lane roads	56-82
Urban/ suburban roads	32 upwards

The variations are due to adjustments to take account of bendiness, road junctions and partly built-up links.

The major experimental measure used is the sum of the route lengths covered on each road class by each subject. Links used where there was no choice of road class available (i.e. near certain endpoints) have been omitted from these figures. A 'class ratio' is a class route length divided by that of the class immediately below it. 'Gradients' represent the average decline in route length with each downward step in the classification, and are therefore a useful summary measure of the overall success of the maps in concentrating routes onto the higher class roads. By omitting the Category 1 links (the perceived minimum distance routes) a measure is obtained of the 'pulling power' of the class to attract people away from them. The significance of differences is assessed by the two-sample t-test with pooled standard deviation. n.s.= not significant, * significant at p<0.05, ** at p<0.01 and *** at p<0.001.

MAP 1

MEAN CHARACTERISTICS OF CHOSEN ROUTES

	Set A	Set B	Differences
Distance (Km)	76.4	76.19	n.s.
Time (mins)	59.99	60.11	n.s.
Class changes	4.68	5.08	n.s.

AGGREGATE ROUTE LENGTHS (Km) BY ROAD CLASS
All links; class ratios in brackets

		Set A		Set B		Differences
Class	1	2805.7	(1.08)	2741.5	(1.12)	n.s.
	2	2589.8	(3.32)	2456.9	(2.37)	n.s.
	3	780.9	(1.04)	1038.0	(1.36)	*
	4	753.0	(1.97)	764.4	(3.36)	n.s.
	5	382.4		227.8		n.s.
Gradient		605.8	Km/class	628.4	Km/class	

(continued overleaf)

Map 1 (cont'd)

Category 1 links omitted; % of class total in brackets

	Set A		Set B		Differences
Class 1	2370.0	(84.5%)	2216.8	(80.9%)	n.s.
2	1727.4	(66.7%)	1660.3	(67.6%)	n.s.
3	66.2	(8.5%)	135.2	(13.0%)	n.s.
4	188.3	(25.0%)	275.4	(36.0%)	n.s.
5	28.0	(7.3%)	25.2	(11.1%)	n.s.

Overall between-map difference in class usage:
$F(4,990)=1.16$ n.s.

Clearly the 'improved' design of A1 had a negligible effect in reducing travel times. In fact the overall performance of the two maps was very similar, despite the differences in symbolisation and appearance between the uncased, highly chromatic look of A1 and the duller cased lines on B1. This was also the only one of the map pairs where the maps did not differ significantly in terms of the overall distribution of route lengths between the classes. This may be largely due to the large influence of the representation of towns discussed with respect to maps B1 and A2 below.

The only statistically significant difference to emerge was in class 3, where not surprisingly the red line of B1 was the more popular. However on B1, despite its lower perceived magnitude (see Appendix D), the green of class 2 still overshadowed the red by a factor of 2.37. This was mainly due to the complete dominance of the A42 on the AB/BA routes- nobody chose the red A609 on B1 (1.5 Km further from A to Marsh but by-passing Eye). The continuity gap caused by the B2243 was clearly a deterrent, but overall the massive popularity of the green class 2s needs more explanation, as they were a close

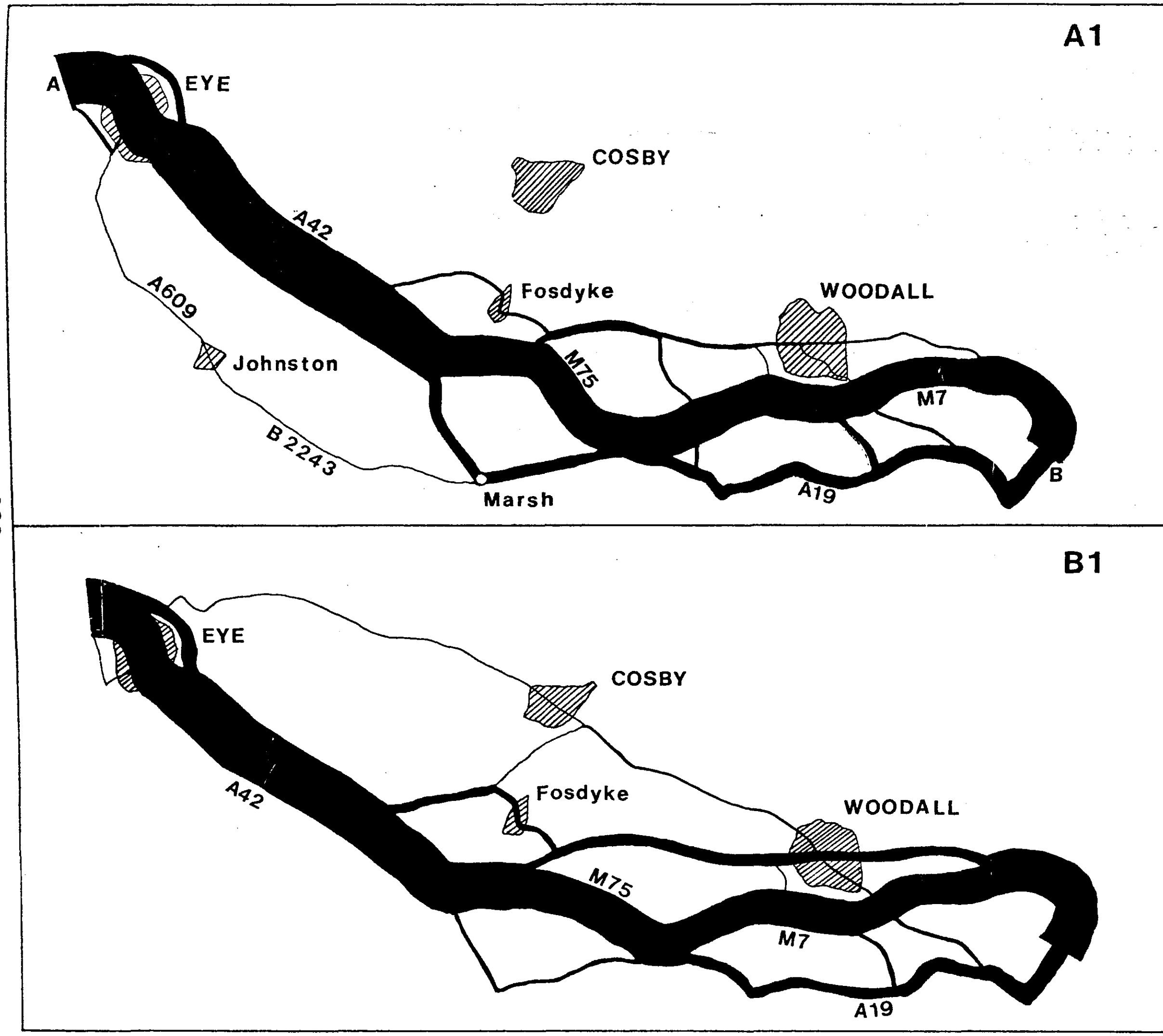


FIGURE 9.1

CHOSEN ROUTES

MAP 1

AB/BA

1mm line width
= 6 choices

50 choices per design

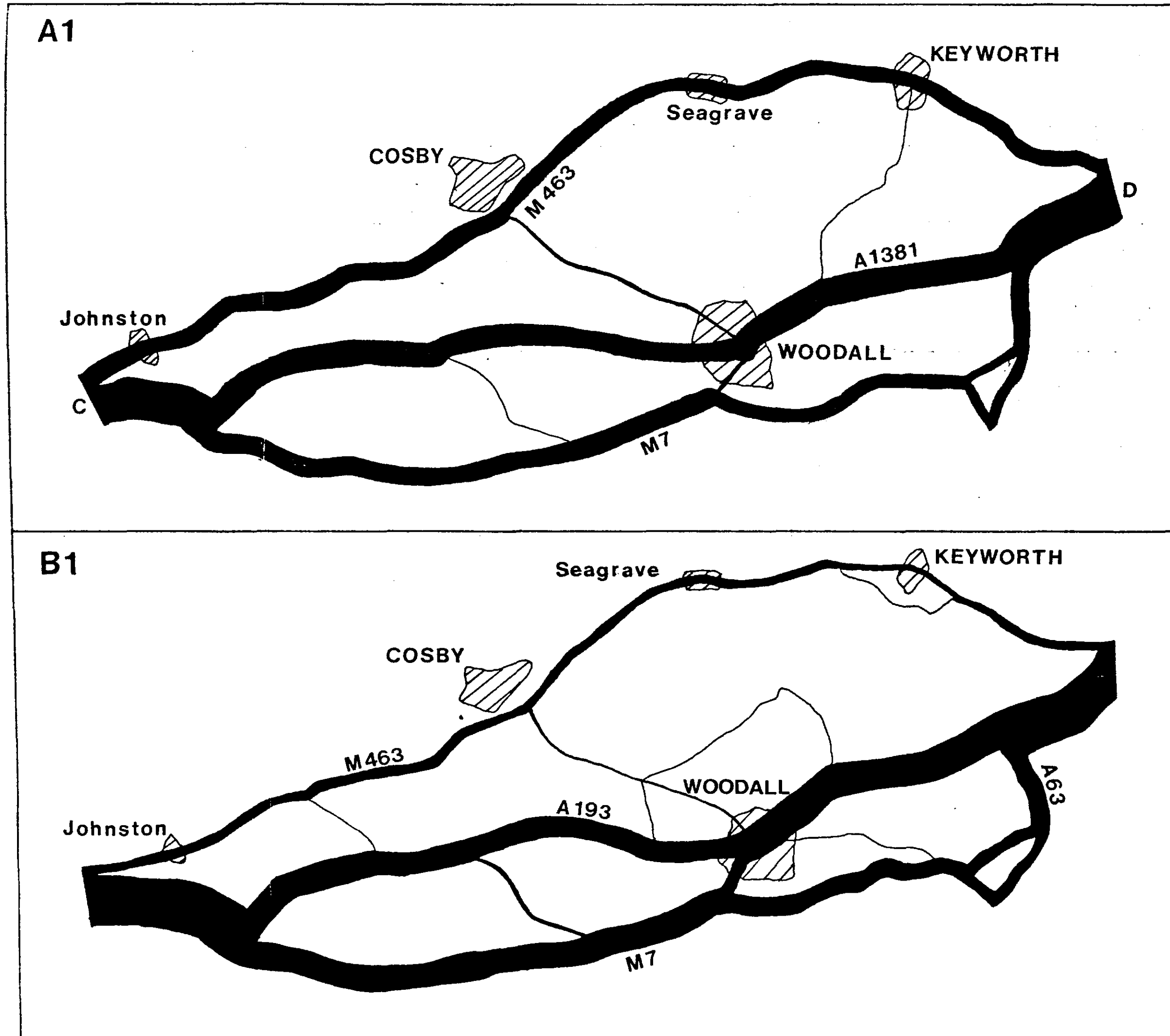


FIGURE 9.2

CHOSEN ROUTES

MAP 1

CD/DC

1mm line width
= 6 choices

50 choices per design

second to the motorways on both maps. This was despite apparent ignorance of the green = primary route convention, or ignorance of what a 'primary route' is. Although 65 of the 137 people who were able to name their own maps used designs with green primary routes, only one person in the entire experiment interpreted them as 'primary routes', and few more were aware of green roads being 'good' on their maps. However, many subjects referred to them as 'major', 'good' or 'good A' roads, while some specifically mentioned their continuity.

This would seem to be the key to the issue. On map B1 continuity is reinforced by the green road number boxes and the relatively small colour contrast with the blue motorways (due partially to the effect of the casings). In contrast on map A1 the line continuity of the motorways is interrupted by the blue number boxes. Thus, added to the obscuring effect of the point symbol for Garth on the relative classification of A19 and M58, the A19 was seen to be more attractive than expected as an approach/departure to and from point B. The combination of width and high chroma certainly helps the green to draw the eye, so that the top two classes on map A1 tower massively over the narrow and largely achromatic residue, the big discontinuity (class 2:3, ratio 3.32) being at the first width decline in the series.

A further point of interest is the usage of the northern CD/DC route via Johnston, Seagrave and Keyworth. This appeared to stand out more on map A1 because of the red central motorway section (M463), and because of its angular continuity, it appeared to some to be more direct than it actually is (5.5% longer than the shortest route chosen). However the most common justification for this route was that it would be much easier to follow, with few class changes (actually 5) and fewer junctions to negotiate than in the more complex part of the network around Woodall. Choosing the easiest route to follow appears to be a major motivation for many people, especially those driving without navigators.

MAP 2

MEAN CHARACTERISTICS OF CHOSEN ROUTES

	Set A	Set B	Differences
Distance (Km)	75.45	75.34	n.s.
Time (mins)	58.35	59.86	**
Class changes	5.64	6.53	*

AGGREGATE ROUTE LENGTHS (Km) BY ROAD CLASS

All links; class ratios in brackets

	Set A		Set B		Differences
Class 1	2498.8	(1.20)	2165.1	(1.25)	n.s.
2	2078.5	(1.67)	1728.4	(1.17)	*
3	1245.2	(1.24)	1475.3	(1.08)	n.s.
4	1002.1	(2.39)	1361.5	(2.59)	**
5	419.5		525.4		n.s.
Gradient	519.8	Km/class	409.9	Km/class	

Category 1 links omitted; % of class total in brackets

	Set A		Set B		Differences
Class 1	1397.6	(55.9%)	778.7	(36.0%)	**
2	1162.9	(55.9%)	676.8	(39.2%)	*
3	204.3	(16.4%)	215.8	(14.6%)	n.s.
4	327.7	(32.7%)	519.4	(38.1%)	n.s.
5	11.5	(2.7%)	69.1	(13.2%)	n.s.

Overall between-map difference in class usage:
$F(4,990)=4.73$ **

With the removal of hierarchical road numbers, map users were all the more dependent on graphical interpretations in order to decode the classifications. A2 stood up to this well, performing similarly to B1 in this respect, with most people assuming the classification to be 'conventional' and thinking little of it. Several, however, said it was more difficult to use than A1. Only a small group was disturbed by the magnitude reversal involving the red class 3. Four people, including at least three users of maps with 'conventional' classifications, interpreted it to be motorway, while another three considered it to be the most prominent. However, in general the continuity aspect was clearly paramount, and only one subject commented that the lack of road class letters had made his task more difficult.

B2 was a different story. 34 of its 100 users spontaneously expressed some form of difficulty or disgust with the appearance of the map. Average travel times were significantly longer than for A2, and the routes involved more class changes. The main cause of this becomes clear in examining the maps for the AB/BA journey where most of the differences lie. A considerable number of people were able to decode the classification satisfactorily. Some sought blue motorways in vain and opted for purple as the closest available colour. However many chose a straight, direct route as they had difficulty in assessing which roads were the best. Consequently 58% went via Venn and Oak Street, the highest proportion taking any single route on any of the maps. This explains the higher number of class changes, and why links of the top two classes involving a detour off the straight line route (i.e. excluding category 1) were used significantly less than on A2. One or two particularly deviant routes, involving up to 45% extra unnecessary travel time, were also selected by people going to great lengths to avoid the apparent complexity of the centre of the map. In contrast to A2,

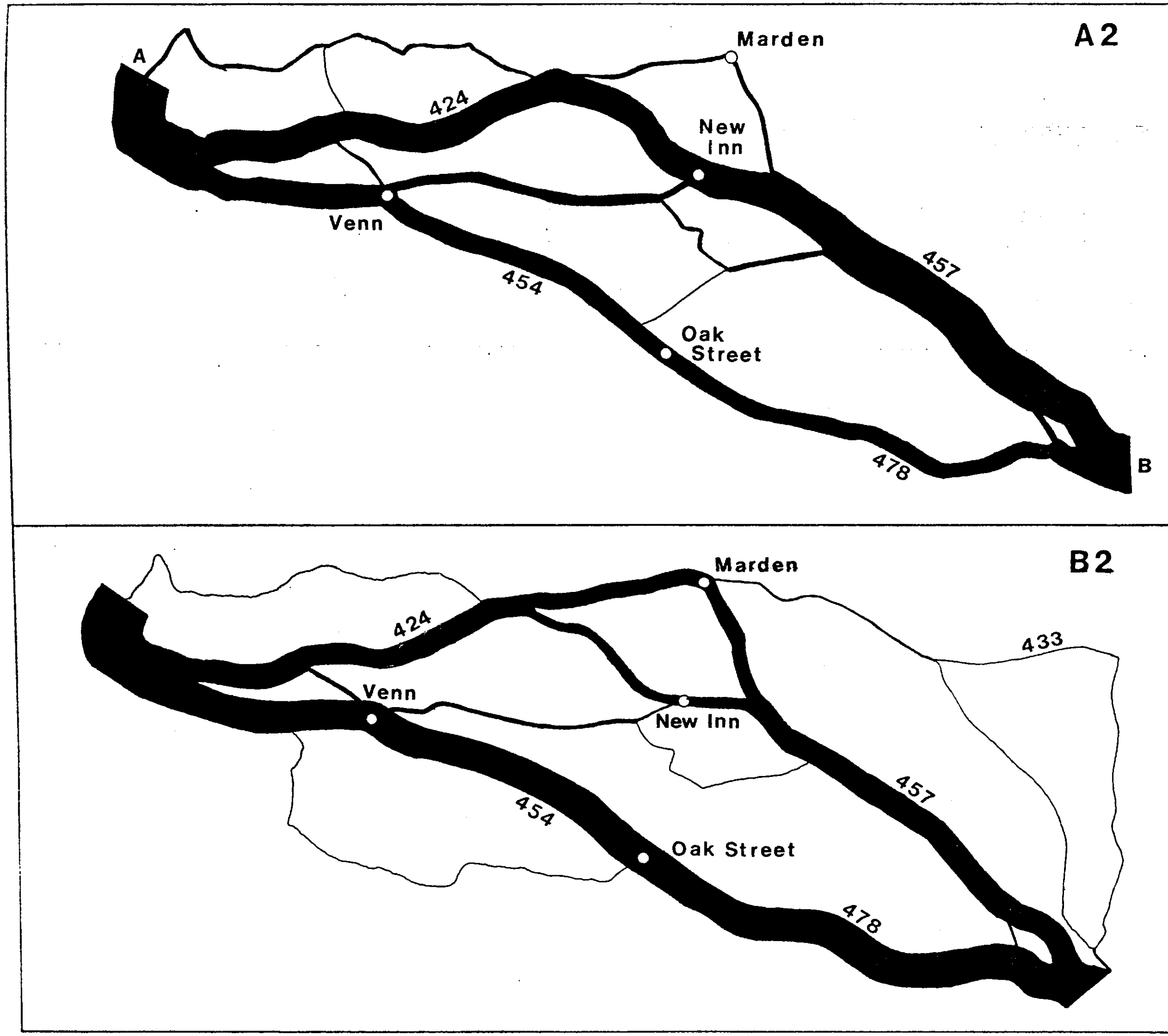


FIGURE 9.3

CHOSEN ROUTES

MAP 2

AB/BA

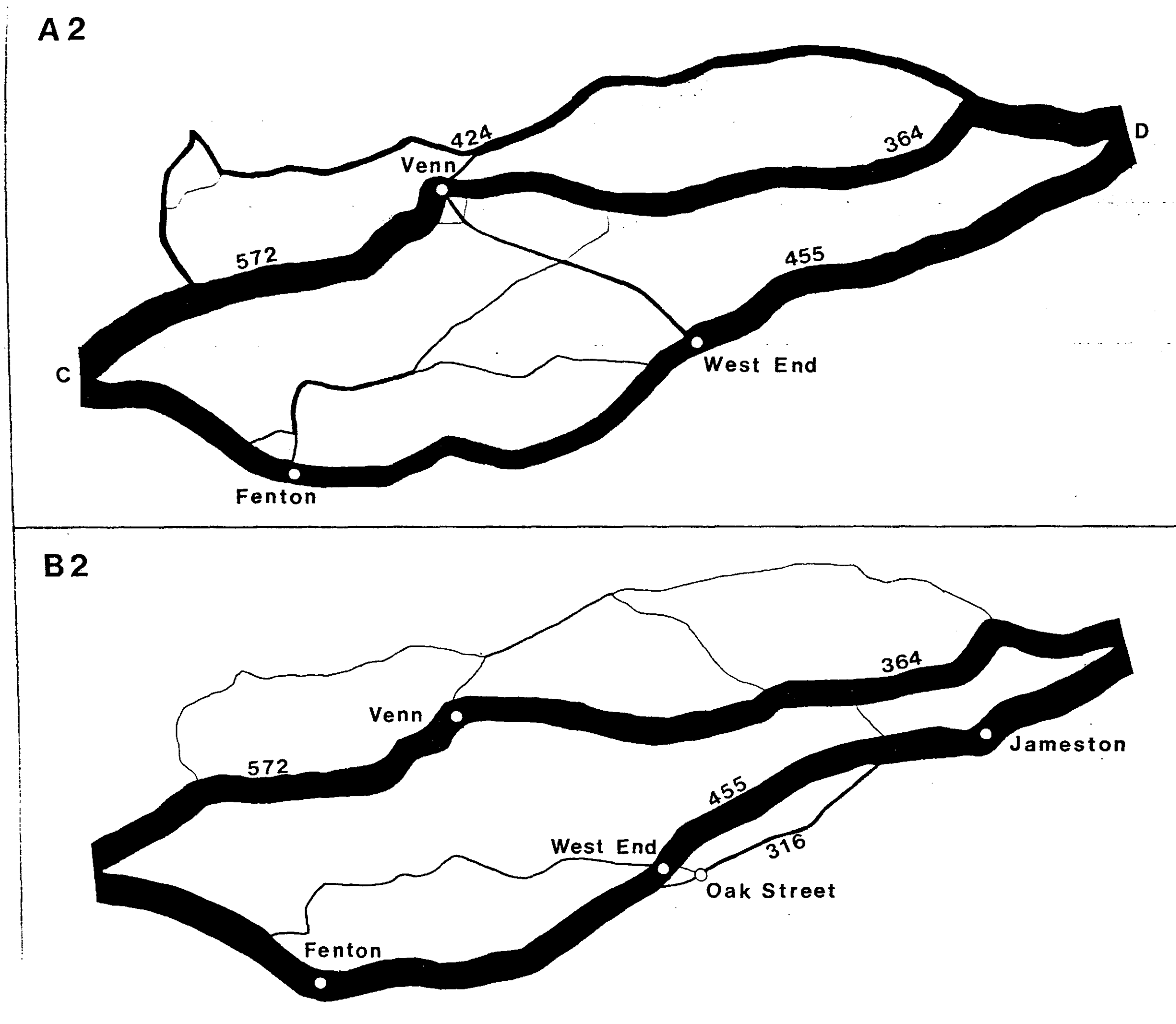
1mm line width
= 6 choices

50 choices per design

FIGURE 9.4
CHOSEN ROUTES

MAP 2

CD/DC



1mm line width
= 6 choices

50 choices per design

several people complained about the uniform road numbers, presumably because these were required as a second string for decoding the classification as the colour scheme was relatively intractable.

The main graphical cause of all this is the lack of distinction between the colours used, the Farmer differences between adjacent classes varying from 7.23 to 8.37, making it difficult to identify individual classes. This was evident right from design and production stages. The design of the colour specification, with the threefold objectives of equal hue steps within a part-spectrum, increasing lightness and a smooth decrease in predicted magnitude, allowed minimal leeway for reprographic distortion, so that on the proofs pink (class 3) and orange (class 4) are especially indistinct. The purple too, a colour still occasionally used by Mairs to denote fast roads, was less brilliant than expected and fairly similar to the maroon class 2. The obvious conclusion is that, at least for uncased lines in a classification with fairly conventional continuity, schemes based on progressive hue differences and lightness ordering within a harmonious part-spectrum are ineffective.

The fluorence of the lighter lines, and particularly the yellow ones, also impaired people's choices. It was necessary for the yellow (class 5) to be of high chroma as no casing could be used and a minimum level of contrast against the background clearly had to be achieved. Although the main consensus about the colour scheme was that yellow formed the lowest class, a few people on CD/DC used the yellow 316 (almost a continuation of the yellow 363) between Jameston and Oak Street in preference to the parallel class 1 455. A similar effect (but with a yellow of higher chroma) was discovered by Cuff (1972a) in a similar red-to-yellow part-spectral scheme on choropleth maps.

Clearly perceptual organisation of similar colours without differences in width is very difficult. Given the additional unfamiliarity of the scheme, B2 was starting with a considerable handicap. This design is clearly unworkable, although route lengths did still decline with each class step. Use of black for the top class could improve hue distinctions slightly, as only four colours would then be needed from the part-spectrum. However, if an equal width scheme is required, use of

1) a wider hue range (losing the ordering aspect of hue)
2) casings, carefully applied, and
3) a redundant chroma progression to back up the lightness series

may together produce an adequate level of comprehension.

MAP A2/B1

MEAN CHARACTERISTICS OF CHOSEN ROUTES

	Map A2	Map B1	Differences
Distance (Km)	75.45	76.19	n.s.
Time (mins)	58.35	60.11	**
-map 1 speeds	61.7	60.11	*
Class changes	5.64	5.08	n.s.

AGGREGATE ROUTE LENGTHS (Km) BY ROAD CLASS

All links; class ratios in brackets

	Map A2		Map B1		Differences
Class 1	2498.8	(1.20)	2741.5	(1.17)	n.s.
2	2078.5	(1.67)	2456.9	(2.36)	*
3	1245.2	(1.24)	1038.0	(1.36)	n.s.
4	1002.1	(2.39)	764.4	(3.36)	*
5	419.5		227.8		*
Gradient	519.8	Km/class	628.4	Km/class	

Map A2/B1 (cont'd)

Category 1 links omitted; % of class total in brackets

	Map A2		Map B1		Differences
Class 1	1397.6	(55.9%)	2216.8	(80.9%)	***
2	1162.9	(55.9%)	1660.3	(67.6%)	*
3	204.3	(16.4%)	135.2	(13.0%)	n.s.
4	327.7	(32.7%)	275.4	(36.0%)	n.s.
5	11.5	(2.7%)	25.2	(11.1%)	n.s.

The table above clearly shows the steeper gradient of class usage for map B1, and the much greater tendency of its users to detour to use class 1 and 2 roads. Is this mainly due to reinforcement of the classification by the hierarchical road numbers, or the effect of built-up areas in deflecting routes onto higher class by-passing roads? Certainly from the comments made, avoidance of large towns is clearly a major element in many road map users' cognitive strategies, but there were also a lot fewer mentions of road classes as reasons for route choice on A2 than on B1. The effect of towns can be gauged by the travel time data. When rural speeds are used everywhere, the routes chosen on A2 were significantly quicker than those of B1. However if the same routes had been chosen with the built-up areas shown (map 1 speed data), they would have been significantly slower. The route maps (figures 9.1-9.4) reveal the influence of the larger towns in particular, with Woodall and Cosby blocking the straight line route on AB/BA for all but one subject. Woodall in particular funnels routes into the M7/M75/A42 corridor, as well as onto the M7 in CD/DC. In doing so it has also had a considerable influence in ironing out the differences in class usage between maps A1 and B1, where Woodall is shown as built-up on both maps.

Furthermore, there was a slightly greater tendency shown on B1 to by-pass the medium-size towns of Eye and Keyworth. However, the lesser use of the 'easy' CD/DC route on B1 (Johnston, Seagrave, Keyworth) is harder to explain confidently, as although it does pass through 3 small towns, it also covers considerable distances on unnumbered and B roads, whose uniform numbering on 2A may lend them more credibility. Evidence of a stronger town effect than otherwise suggested is however provided by some users of map A2, and the other maps with uniform settlements. They still avoided places they considered from the density and shape of the network around them to be large- notably Venn on map 2 (i.e. Woodall on map 1) and Lyng on map 3. As these are in reality Chesterfield and Nottingham respectively, this was quite perceptive map reading.

MAP 3

MEAN CHARACTERISTICS OF CHOSEN ROUTES

	Set A	Set B	Diffces.
Distance (Km)	77.2	77.75	n.s.
Time (mins)	58.78	57.64	**
Class changes:			
With colour change on B3	4.24	3.09	***
Other single/dual joins	2.19	2.89	***
Total	6.43	5.98	n.s.

(continued overleaf)

Map 3 (cont'd)

AGGREGATE ROUTE LENGTHS (Km) BY ROAD CLASS

All links; class ratios in brackets

		Set A		Set B		Differences
Dual	1	2277.6		2843.1		**
	2	423.6		458.1		n.s.
	3	52.2		10.4		**
Total dual		2753.4	(1.13)	3311.6	(1.29)	**
Single	1	2446.5	(2.37)	2577.8	(3.00)	n.s.
	2	1032.5	(1.41)	859.7	(2.64)	n.s.
	3	732.4	(1.75)	326.1	(1.00)	**
	4	419.8		324.6		n.s.
Gradient		583.4	Km/class	746.8	Km/class	

Total length, single and dual; % dual in brackets

Class	1	4724.1	(48.2%)	5420.9	(52.4%)	***
	2	1456.1	(29.1%)	1317.8	(34.8%)	n.s.
	3	784.6	(6.7%)	336.5	(3.1%)	***

Category 1 links omitted; % of class total in brackets

		Set A		Set B		Differences
Dual	1	1778.9	(78.1)	2239.5	(78.8)	**
	2	234.7	(55.4)	305.7	(66.7)	n.s.
	3	48.4	(92.7)	6.6	(63.5)	**
Total dual		2062.0	(74.9)	2551.8	(77.1)	***
Single	1	1926.7	(78.8)	2014.7	(78.2)	n.s.
	2	717.9	(69.5)	637.9	(74.2)	n.s.
	3	591.6	(80.7)	205.5	(63.0)	**
	4	53.2	(12.7)	56.8	(17.5)	n.s.

Overall between-map differences in class usage:

$F(4,990)=6.45$ ***

A3 and B3 were the most successful of the 'improved' maps in terms of the information provided about the effects of graphical variables. Given the relatively unconventional use of colour and the uniform road numbers, subjects were forced to make specific graphical interpretations, which they commented on more freely than with the other maps, without the task appearing to be as difficult as in B2.

One point very clearly demonstrated was the greater significance of continuity of colour than continuity of casing, in two ways:

1) perceptually, in terms of leading the map user along a particular link. This was illustrated by the significantly differing usages of the dual carriageway sub-classes. On A3, where they are symbolised identically, the decline in usage is significantly less steep than on B3, where the filling colours change. For example, the dual class 3 loop at Yarrow (south east from A) is a detouring continuation of a longer stretch of dual class 1. It was selected by 8 viewers of A3, but none of B3, where the casing continues but the filling colour changes from red to yellow. This does nothing to attract people away from the alternative red single carriageway.

2) cognitively, in terms of the user's desire to minimise the number of class changes in the route, where highly significant differences emerged. While mean numbers of class changes were very similar overall, they were much higher (i.e. relatively ignored) for 'other single/dual joins' where the changes are only of casing thickness (map B3) rather than colour (map A3).

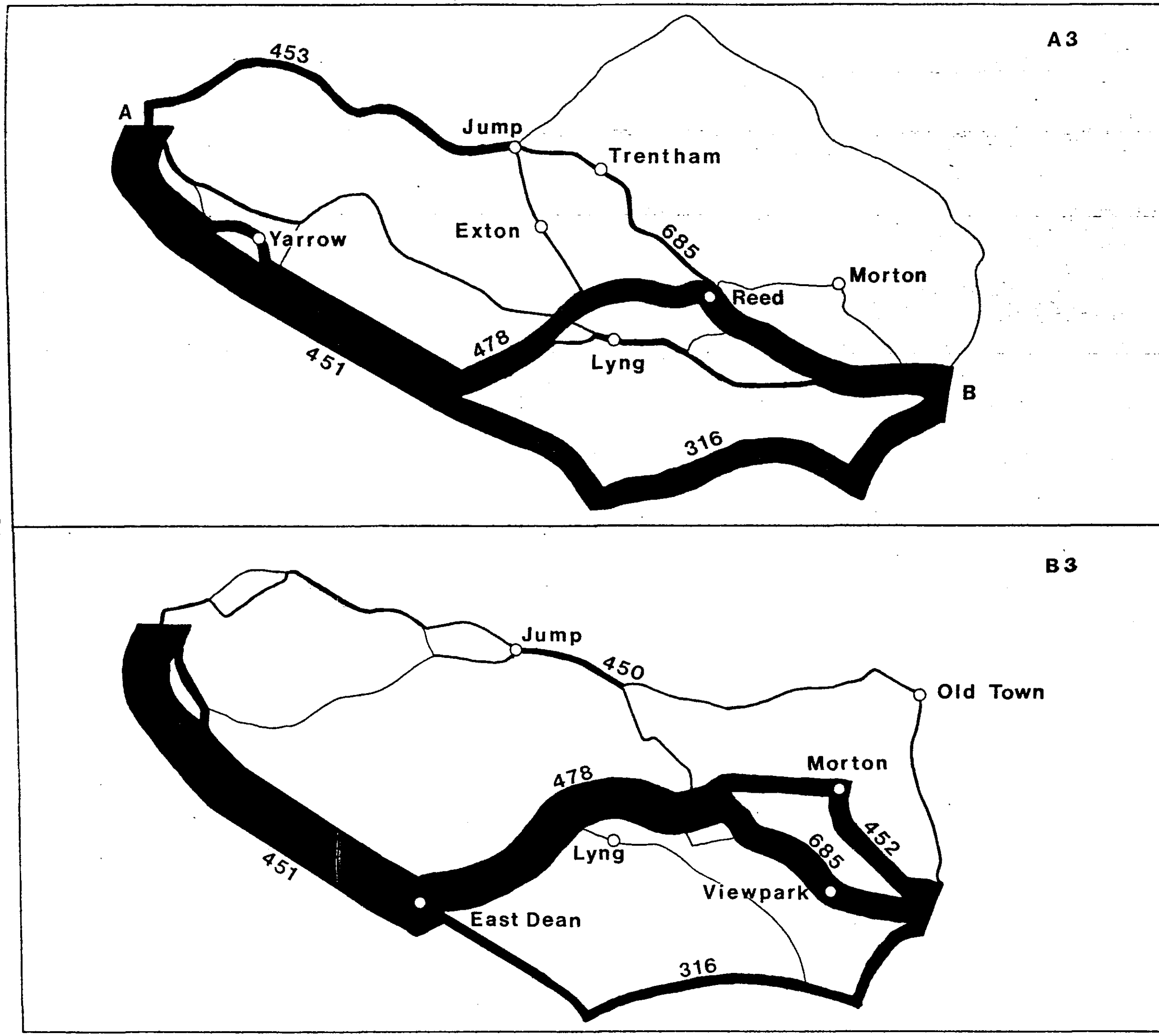
Thus a major discontinuity in road quality such as single to dual carriageway is best indicated by a colour change. On the Shell road atlas this is achieved without loss of perceived continuity of primary routes with dual sections. All dual carriageways are shown in solid red, which contrasts in hue with the amber single 'other A

FIGURE 9.5

CHOSEN ROUTES

MAP 3

AB/BA



1mm line width
= 6 choices
50 choices per design

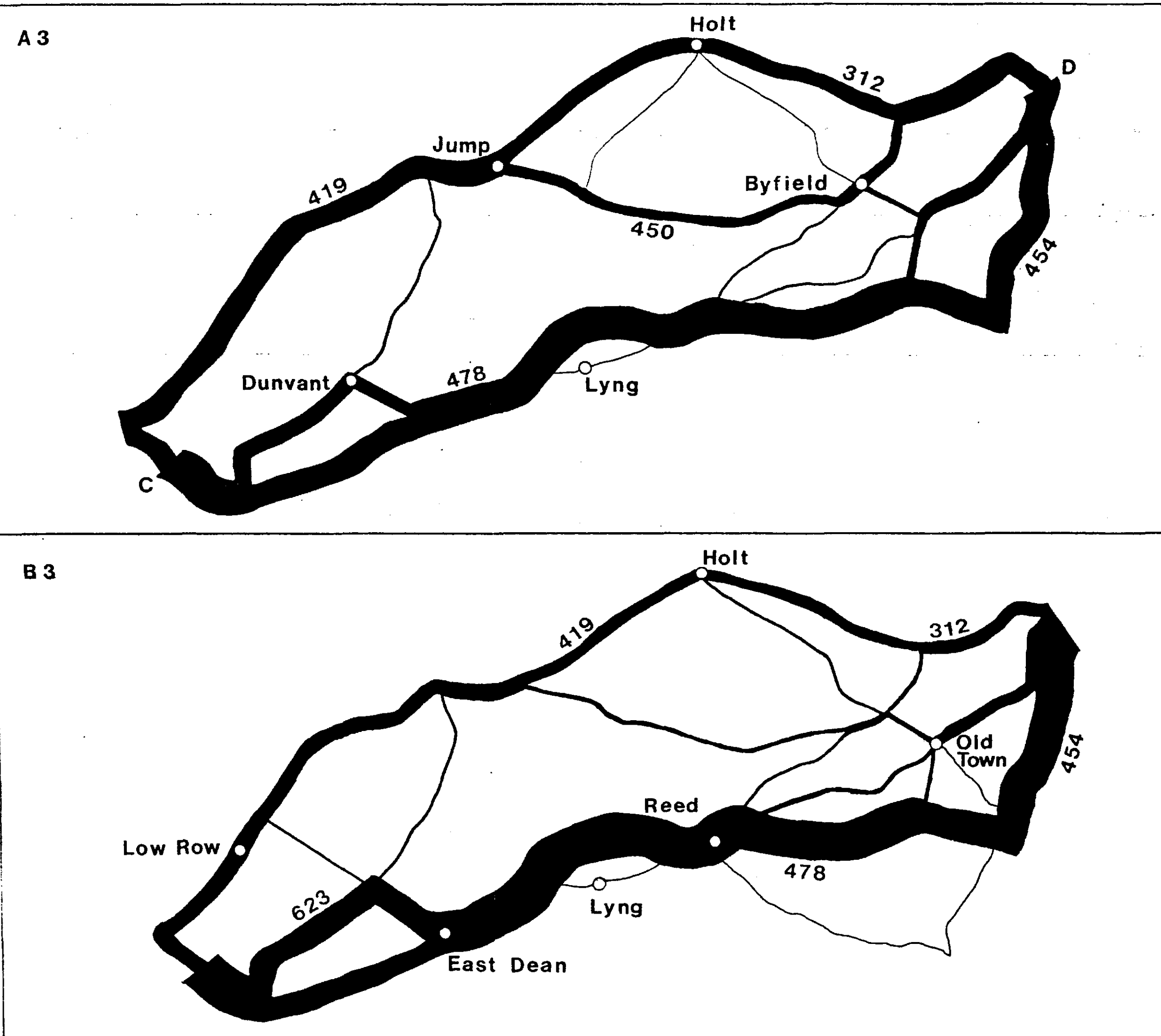


FIGURE 9.6

CHOSEN ROUTES

MAP 3

CD/DC

1mm line width
= 6 choices

50 choices per design

roads' but not with the screened red single primary routes. Thus hue is maintained at single/dual joins, while prominence is altered by a change in lightness and chroma.

After B2, A3 was clearly the least popular of the maps, and attracted 23 complaints for its monotony (or 'redness'), lack of clarity and difficulty to use. One user, after his first look at the map, decided that he would take a train. Again the colours and widths are mostly within a narrow range. Consequently the thin green lines of class 4 single appeared to be so different that many people wondered whether they were still roads. However, the map appeared to perform fairly well. Many subjects (23) commented that the red-cased yellow lines were the most prominent or the best roads, including 13 who took them to represent dual carriageways or motorways. This clearly shows the connotative effect of the thick, coloured casing, which attracted far more comment than the thick black casing for dualled roads in B3 (5 mentions). Although one or two people reported difficulty in assessing its classification relative to the brown (class 1 single), the red/yellow symbol was generally preferred to the darker and slightly wider brown line, despite the brown being qualitatively more distinct from the rest of the classes. As continuity is of little assistance here, only a combination of chroma and connotation can be responsible for this.

The part of the map between Jump and Reed was specifically designed to see classes 2 single and 3 single (the cased and uncased versions of the same line) in direct competition. Of the three virtually equidistant possible routes, only the class 2 ones (via Exton and Trentham) were used, and the class 3 road numbered 418 was ignored. Clearly the casing has no special magic in an unusual situation where the cased line is less continuous than the uncased line, and thus despite its enhanced edge contrast appears to form the weaker figure. It would seem that only thick casings have a significant connotative effect. Character contrast in itself is often buried beneath stronger, well-marked width and colour contrasts as on B3, where the lack of casing on class 2 single (green) was barely noticeable. Only the 'dirtying' effect of black casings remains, with its variety of effects which can be viewed on the proofs, where 5 out of 6 extracts have lines of the same red in various widths and casing states.

On map B3 a very neat hierarchical appearance was obtained by using highly distinct colours of decreasing magnitude and decreasing line widths, and because the discontinuities introduced by stretches of dual carriageway are shown without a colour change. The map demonstrates the reinforcing effect of interpretations of graphical variables working together rather than against each other. This clarity meant, for example, that far fewer people felt the need to select the 'easy' route on AB/BA (following the 451 down to the 316 on the bottom edge of the map) to avoid the more complex area in the centre. It also correlates with the conclusions of a map-based study by Lloyd and Yehl (1979), presenting messages of varying importance in lettering of various sizes, that fewer errors were made in the interpretation of message importance when the connotations of the graphics words were concordant with their visual order.

Effectively it is a map of four visual planes, with red clearly forming the highest plane. A stunning 52 people specifically mentioned that the red roads were the most highly classified- by far the most common comment made about any of the map designs or colours- and there was not a single expressed misinterpretation of the relative classification of the roads. Clearly there was some connotative element to this- the red=major/A roads convention is still strong when not buried beneath green primary routes- but the vast majority commented on their prominence rather than making a specific road type interpretation. One person mentioned that his 'natural instinct' was that red was better than green, although it was the other way round on his atlas.

Overall this map performed the best of all in terms of attracting users to the most highly-classified roads. It has the steepest gradient of class route lengths, with 79.6% of the total distance covered being on the top two classes (i.e. dual and class 1 single). As mentioned above, this was almost entirely due to colour rather than casing, as 73.2% of the total was on red roads alone. An illustration of this is that the route from Reed to point B via Morton rather than Viewpark was used far more often on B3 than A3.

9.4.1 General Conclusions

The most noticeable fact overall was that, despite certain prominence reversals within the schemes and the substantial difficulties with B2, the classification of each map was at least decoded to the extent that the distances travelled declined without exception with each downward class step. Overall the massive effect of the spatial arrangement (continuity and closure) of road classes in organising the map is clear: only a small proportion of users are significantly disturbed by a more prominent but less continuous road class.

The magnitude message was conveyed most strongly by width, as expected from the previous experiment, but the role of chroma was important in certain contexts, despite the illumination at the test sites which tended to reduce the insistence of colours. Chroma was particularly significant where there were no black casings to contain it physically and 'dampen' it down, as on map B2 where the yellow is actually of lower chroma than the one on B3.

A further observation is that most map users are particularly concerned with the extremities of classification schemes, and are relatively indifferent to the middle. Generally the major discontinuities in class usage were below the top two classes (generally thought of as motorways and other major roads), and above the bottom class. This is reflected too in people's comments about the types of roads they went for and avoided. Thus on all the maps the highest ratios were from the 2nd to the 3rd class, and from the 4th to the 5th, except on B2 where the only big jump was from 4:5 (as mentioned above), and on B3 where because of the dominance of the top classes, anything below the 3rd (i.e. green) class appeared to be undesirable or unnecessary.

People clearly prefer and tend to perform better with maps where the classes are in highly distinct and unconfusable colours. This is difficult to achieve solely by a colour scheme of declining magnitude. Thus it is hardly surprising that colour conventions have developed and been perpetuated, in order to use cognitive separation to widen (or override) the relatively small perceptual differences. However, conventions do seem to work best when in harmony with perceptual hierarchies as in map B3. Yet conventional maps certainly held their own in the comparisons with 'improved' maps. Clearly familiarity is to their advantage. However, the overall differences between the maps were less significant than those found in the pilot study, such that the conclusions from the current tests are generally similar but weaker, showing less reliance on the physical prominence of lines. This may relate to the homogeneity of the pilot study sample, the particular routes set, and the more relaxed and detailed interviews possible.

9.4.2 Analysis of Sub-Groups

Finally, it is certainly possible that the map user population is composed of several sub-groups, which might be related to social class and/or map use regularity, with very different levels of dependence upon graphical prominence and connotative aspects such as colour conventions. Clearly people have different cognitive strategies and put very different levels of thought into road map reading.

From the comments made by the respondents, it would appear that the cognitive strategies used in seeking the fastest routes from the maps were composed almost exclusively from one or more of four basic elements. These are tabulated below together with the total number of people from the full sample who made use of them.

Use of major/ highly classified roads	181
Shortest/ most direct route	140
Ease and certainty of route following	88
Avoidance of built-up/ urban areas	75
No usable information supplied	5

A handful of people also mentioned that they looked for 'quiet roads'. Clearly the use of major roads was the most commonly stated element, but how did these subjects decide which were the major roads? Most of them made specific mentions of the cues used. For map 1, where the D.Tp class letters were included, choices were made as follows:

By class letter (/symbol connotation)	111
By symbol prominence	11
By symbol continuity	1

Consequently it is hardly surprising that the prominence reversal on map B1 had little overall effect on route choices, although on B1 the 'prominence' people used on average nearly 56% more of the red class 3 road than the remainder of the representative sample. Clearly the influence of the class letters is dominant here, even on A1 which lacks the advantage of redundancy provided by conventional motorway colouring (blue).

Where the class letters were not included (maps 2 and 3), the cues mentioned were more diverse:

Connotations (of colour and casing)	43
Symbol prominence	29
Continuity	15
Take direct/ easy route instead	16

A further 21 mentioned prominence only where the continuity cues were either concordant (B2,B3) or weak (A3), and an extra 3 relied on prominence where no conventional colours could be found. Also 35 more people (31 on B2; 4 on A3) opted for direct routes when they considered that the road classification was unclear, confirming the need for clear distinction between classes.

The following interpretations were mentioned by the subjects who relied upon connotations:

Blue= motorway	22
Red= A road/ motorway	11
Red/ yellow= dual carriageway/ motorway (A3)	9
Thick casing= dual carriageway (B3)	9
Green= primary route	1

Thus the power of the blue= motorway convention is confirmed; many of these respondents assumed that no motorways were present on maps without blue road symbols. The significance of continuity found in the route choices is not reflected here in the level of comment, implying that it lies at a sub-cognitive level of processing.

Conversely the role of prominence was clearly boosted somewhat by the key, although 40 (70%) of the 57 different users for whom prominence was of some importance had not actually consulted it. Overall, symbol prominence is perceived to be far more important where class letters are not shown and when continuity is an unclear or unreliable cue.

Those people who based their route choices on prominence on A2 (with the same prominence reversal as B1) went on average over 71% further than the rest on the red roads. Consequently it can be seen that a small sub-group particularly dependent on prominence may take significantly different routes from the rest. This group was isolated, by taking the 36 subjects who mentioned prominence as a major cue, and compared to the rest of the sample. No significant differences were found in frequency of map use or social class, except for the tendency of people of lower social classes (III to V) to be less reliant on prominence (chi-squared 4.45,1 df, $p<0.05$). It might perhaps have been expected that less educated map users would be more reliant on low-level processing cues; perhaps this is an indication that they are slightly less flexible in assigning meanings to lines and more likely to stick to the conventions they know.

9.5 RESTRICTED ACCESS JUNCTION SYMBOLS

With the construction of purpose-built high-speed roads, an extra problem has been presented to the road map user by the increasing number of multi-level interchanges where not all the turns between the roads are possible. Misjudgment of these situations can result in a considerable detour being necessary. Some map users now consider information on restricted access junctions to be amongst the most important on the whole map (Gill,1982). Yet in a route-finding experiment conducted by Streeter and Vitello (1986) in the U.S.A., taking an illegal entry or exit onto a limited access road accounted for 94.1% of the map interpretation errors made.

Many road maps indicate which motorway junctions are restricted by changing the colour of the disc symbol, often to red. Fewer attempt to indicate precisely what restrictions apply. Some maps have insets either with enlarged junction plans or lists of restrictions. Others have separate motorway strip maps with written information on access limitations adjacent to the junction symbol. A further group even attempts to show the ramps of all multi-level junctions, or just those on motorways, at main map scale. Even with a certain amount of enlargement and displacement, it becomes very difficult to depict with any clarity complex junctions in cluttered parts of the map (e.g. on urban motorways). The emphasis on morphology is moreover not necessarily helpful. The overriding concern of the map user is not to attempt to navigate a specific path around the 'spaghetti', for which he will generally follow the road signs. However, in route planning he could often benefit from a simple and clear indication of whether a particular access/exit is available. Morphologies are not standard, but vary considerably according to local conditions, such as space limitations and the shape of the local road network, so that several different morphologies may provide the same access opportunities.

However, the vast majority of access limitations fall into a few specific types. For example, the 96 restricted access numbered motorway junctions in Great Britain (as at November 1985) can be categorised as follows:

Fork- forking road ends at junction	21
Fork- forking road continues beyond junction	18
Half interchange	37
Complementary junction	12
One turn missing	3
Other (complex)	5

These general junction types can be described in detail as follows (see figure 9.7):

-Fork (end). The roads fork, and acute-angled turns are not possible.

-Fork (continuing). With two through roads, access is possible from one side of A onto one side of B, and from the other side of B back onto the other side of A.

-Half. With two through roads, both lanes/carriageways of B have access to one side of A, and are accessible from its other side.

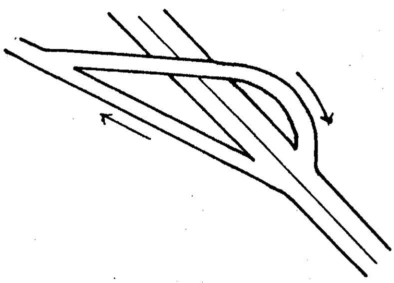
-Complementary. Together with a nearby or adjacent junction, full access is possible.

Some maps have adopted symbols which offer a partial solution to these problems. For example, the AASHO (1962) specification represents interchanges with full access by a rectangle, and half interchanges by a half-size rectangle on the side of the junction where the ramps are situated (figure 9.8). Some map producers have attempted to symbolise entrances to and exits from the major road. However, in a junction with two continuing roads, an exit from the major road may not provide access to the other road <u>in both directions</u>, so that the junction may have a

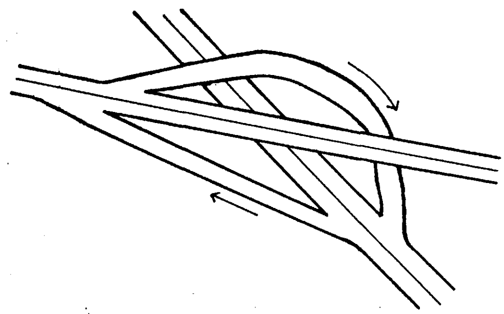
Figure 9.7 TYPICAL LAYOUTS OF RESTRICTED ACCESS JUNCTIONS

All junction diagrams are for left-side driving.

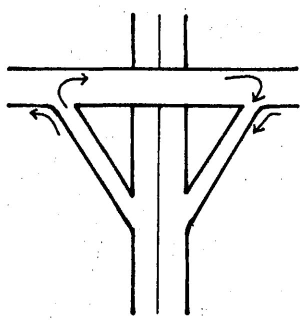
Fork (End)



Fork (continuing)



Half Interchange



Complementary Junctions

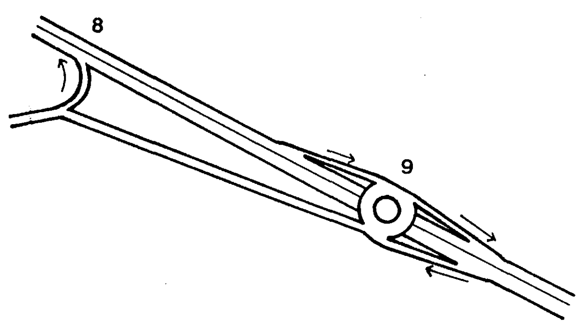


Figure 9.8 AASHO Half Interchange Symbol

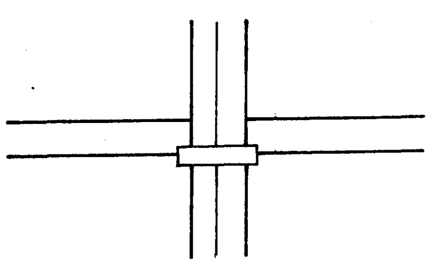
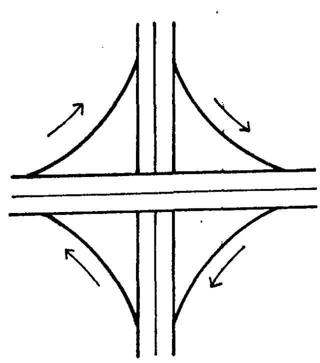
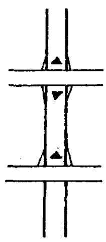
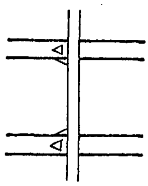


Figure 9.9 Half Interchange, Possible Layout



complete set of entrances and exits, but only half of the possible turns can be made (e.g. figure 9.9- no right turns). They can however be useful in conjunction with thin lines depicting the ramps, or written descriptions of the exit signpost headings (on strip maps). The systems used by Michelin (widely) and Rand McNally (rarely) are as follows:

Michelin	Rand McNally	
	C (red)	Complete access
	N (purple)	Half: northbound entrance southbound exit
	W-E	Westbound entrance only
	E-X	Eastbound exit only

However it is possible to think of ways in which each type of restriction might be symbolised. Complementary junctions could be shown quite simply by running the symbols for the two junctions together. Where one turn is missing, a crossed-out red arrow indicating the outlawed turn could be considered. However, symbolisation for the remaining three types is more problematic, and consequently these have been examined in the current experiments. The overall concept of the designs used was to test a mixture of line-based and area-based symbols, and determine which colour conventions and associations might be of assistance. Details of the individual designs used are as follows:

1) Fork (end), in junction 2. The symbol on set A (design I), analagous to a 'Y'-point on a railway track, was designed by Alastair Morrison and used on his Road Speed Map of South East England (1983). The alternative

(set B, design II) includes the red outline disc to see whether a standard symbol element could be utilised without loss of information.

2) Fork (continuing), in junctions 1 and 4. Five different symbols (designs I to V) were tested. For junction 1, these were outline discs containing coloured sectors shaped to link together the road segments between which access is possible. Given the small area of the discs, it was considered that for clarity and technical ease of reproduction, only one colour should be printed within the disc. Consequently each colour used is effectively tested against white with respect to connotations of linking (or barring). Here the linking sectors are green (set A, design I), with associations with 'go', and red (set B, design II), conventionally used for 'stop' or 'danger'. Thus in II, colour and shape may be emitting opposing messages. With junction 4, a green line is used to link the relevant road segments, again within a disc for emphasis. The line has arrowheads on set A (III), but as these might be difficult to produce clearly at small scales, they have been dropped on set B (IV) to see whether the symbol is as effective without them. During the experiments, when it emerged that a popular interpretation of design II was 'red=barrier, white=link', the reversal of this symbol (i.e. red sector larger than white, design V) was substituted for III in set A junction 4.

3) Half, in junctions 3 and 5. Again five designs were tested. For junction 3, set B (II) shows the ramps by thin unfilled black double lines, whereas set A (I) uses a triangular block, an expanded version of the Michelin symbol, but still more compact than II. For junction 5, coloured sector symbols have again been used, the colours tested being black (III), with its obvious high physical and connotative contrast with white, and blue (IV), intended to imply access through continuity of colour from

the blue motorway. Red (V) was again substituted in later in the trials, this time for the blue. As the disc is divided into equal halves, interpretation of colour is necessary in order to decode the symbol, as shape provides no clues as to which of the two areas denotes access. Unlike junctions 1-4, no interpretations of 5 were deemed to be 'errors'.

The method of testing was to indicate one turn on each of the junction insets by red chinagraph arrows. In each case the subject was asked whether, if he were to encounter this symbol on a map, he would consider that this turn was possible, impossible or that the symbol was not clear. Up to 4 different turns were investigated for each junction, and these were rotated in cycles to balance the numbers of subjects shown each turn. Comments made about the symbols were noted.

9.5.1 Results

Given the obvious unfamiliarity of these symbols, it was anticipated that many might find this the hardest part of the tests. Thus the overall level of comprehension was surprisingly high. Clearly, however, in focussing specifically upon these symbols, the experiment was inevitably not entirely realistic of a real route planning situation. The analysis is concerned more with the level of error than numbers of 'not clears', as the latter could be mostly eliminated by a learning effect if the symbols were ever to be widely adopted. Obviously there was no key present in this case.

Of the full sample, 170 (68.5%) made no errors, or graphically justifiable 'errors' only (e.g. red sector on 2.II is barring access). 25 subjects (10.1%) made no real attempt to discriminate between the symbols. Five said they were all 'not clear', but the remaining twenty gave answers from a priori reasons- e.g. all the junctions were roundabouts so access was always possible, or right turns could not be made on or onto motorways. Consequently these answers had to be excluded from further analysis, as they could unfairly distort comparisons between symbols. However it should be noted that unintended (roundabout) connotations of disc symbols can sometimes present a problem.

It might be expected that regular and more educated map users would be better at making these more complex interpretations. However, there is surprisingly only a very weak relationship between the comprehension level of the symbols and frequency of map use or social class. Only one relationship is statistically significant: social classes I and II combined are slightly better at interpreting the symbols than 'the rest' combined (chi-squared 7.327,2 df, $p<0.05$).

The tables for the overall comprehension levels of individual symbols are presented below. Figures are percentages. The C/I ratio is the number of correct interpretations divided by the number of incorrect ones.

1) Fork (end):

Junction 2.

	1.I	1.II
Correct	90	73
Incorrect	6	12
Not Clear	4	15
C/I ratio	14.1	5.9

Clearly the 'Y-point' symbol (I) performed very well, although a few people were reluctant to take the legitimate fork because they considered the black line across their path to be a barrier. II was significantly less clear overall (chi-squared 11.94,2 df, $p<0.01$) but still well understood. However, far more people considered that the acute turns were possible. For some the symbol appeared to denote a roundabout, while for others the small 'Y' within the disc was insufficiently sharp-pointed to discourage the acute turns.

2) Fork (continuing):

Junction 1/4:

	Sectorial Symbols			Line Symbols	
	2.I (green)	2.II (red,+ve)	2.V (red,-ve)	2.III (arrow)	2.IV (line)
Correct	42	26	64	74	59
Incorrect	14	38	8	14	23
Not Clear	45	36	28	12	18
C/I ratio	3.1	0.7	8	5.3	2.6

This is the most difficult type of junction to symbolise adequately. Conventional wisdom and pretesting suggested that lines carried stronger connotations of linking than sectors, and this was shown to be true. For example, with the arrowed line (III) 100% comprehension was achieved for the acceptable turn (right off M46 northbound). The performance of the non-arrowed line (IV) was not significantly worse. However, 37% of those asked on III and IV thought going straight across the motorway on the A46 westbound was not possible, as the line appeared to indicate a compulsory turn left onto the motorway. Others were tempted to think that way until they realised that it would leave the western segment of the A46 completely unconnected. There is perhaps a problem here with the design of the insets, as if the road segments were continued all the way to the edges of the box, the continuity of the A46 would probably be more obvious. However, nobody thought going straight across was impossible on V, presumably because the strong figure of the dominantly red disc 'raised' it so that the A46 could be considered to pass through underneath the junction. Possible solutions to this problem with a line symbol would be to strengthen the outline of the disc, or add in another linking line straight across, although that might detract from the clarity of the other link by cluttering the symbol. (Logically a 'straight across' link for the motorway would also have to be considered.)

Alternatively, a 'softer' sector-type symbol could be used. More people found them not clear, but the overall error level was low. The exception was II, where the colour connotation of red=barrier, white=empty (no barrier) was stronger overall than the sector shape message. Again this discouraged some subjects from going straight across because they would have to cross the red sector, in contrast to V, which is the reversal of the same symbol. V works the best of the sector-type symbols, better than I (green/white), because of the blocking power of red, but is less clear than the line symbols at showing the acceptable turns.

3) Half interchange:

Junction 3.

	3.I (block)	3.II (ramps)
Correct	63	81
Incorrect	14	9
Not Clear	24	11
C/I ratio	4.6	9.1

Junction 5.

	3.III (black)	3.IV (blue)	3.V (red)
Colour=block	40	22	72
Colour=access	14	34	0
Not Clear	46	44	28
Bl/Ac ratio	2.9	0.6	inf.

The ramps symbol (II) is remarkably effective. Most of the errors are caused by the perception that the northern ramp could not be used for right turns onto the A461 southbound. Design I is also well understood but significantly less clear (chi-squared 9.14,2 df, $p<0.05$), with more subjects viewing it as a flyover without access. The sector symbols are unsurprisingly less successful because of the lack of redundant coding mentioned above and the consequent reliance on a colour interpretaion alone. Thus a high proportion of interviewees found these unclear. Of the colour connotations, it appears that blue (continuity)= access (IV) is the weakest. III has a similar performance level to symbol 2.I, without the shape advantage, suggesting that black= barrier is stronger than green= access. However, the blocking effect of red is best of all, significantly better than black (chi-squared 9.55,2 df, $p<0.01$), and with no errors at all. Although some people might have problems in going straight across a red semi-circle, this symbol could clearly be used as an alternative to II.

Recommendations for Symbols

1) Fork (end):	I
2) Fork (continuing):	III or IV (modified), V
3) Half interchange:	II,V

10. CONCLUSIONS

The major findings of this study can be summarised as follows. Firstly, it was discovered that differences in the symbolisation of roads on maps with the same road classification can have a significant effect on routes chosen from them. Within the range of solid line symbols generally used on road maps, line width appears to be the most significant influence upon their prominence or visual weight, which is proportional to the logarithm of line width. This result was replicated in experiments with three sets of subjects on three differently-coloured backgrounds. The prominence of uncased lines on a white background, for observers with normal colour vision, is also affected by the lightness contrast with the background, saturation (colour strength) and affective value of the line colour, confirming the implications of the literature, and these variables can all be related together by the equation in section 8.2.3. Their relative significance was confirmed in a road map context (experiment 3).

The introduction of black casings to a coloured line serves to strengthen its edge contrast but reduce its apparent chroma. Consequently casings do not greatly boost the perceived prominence of individual lines other than those of light colours such as yellow, in which case the amount of the boost increases with both casing and filling widths. Clearly on some maps where cased lines represent a higher class than uncased lines of the same colour (e.g.red), the correct perception of the classification is reliant upon their superior continuity. Thick (black or coloured) casings also carry connotations of road quality. Background colour complicates the prominence issue, causing increased interpersonal variation in the assessment of the role of the chromatic variables, but it appears that chromaticity contrast may be generally more relevant than saturation contrast, while the relative influence of width is somewhat reduced. In terms of conspicuity, it would seem that colour is a more powerful cue than line width in the perceptual segregation of different line types and the attraction of attention. Line colours with a Farmer contrast of less than about 14 may be confused in extrafoveal vision. Other contrasts being equal, the colours that are more prominent in central vision appear to be the more attention-getting.

For a representative sample of British motorists, the most important cue in the discernment of the road class hierarchy is the Department of Transport class prefix, followed in declining order of significance by the relative continuity of lines of each class, the use of familiar graphical conventions and the prominence of the symbols. The significance of continuity is such that even the interruption of a line by a differently-coloured road number box may hinder the perception of the symbol hierarchy. However, reversals of colour-based prominence in an administrative classification involving continuity cues and conventions do still cause a disturbance to a minority of road map users, and prominence is clearly much more important in schemes where the other cues are weak or absent. Conspicuity may also be more important in in-vehicle map use; although this could not unfortunately be tested. The schemes which perform best in terms of the correct perception of the hierarchy and the maximum use of high class roads are those where the cues are concordant, as in conventions where connotation and physical prominence work together (such as red for major roads), and especially where this is reinforced by relative continuity (e.g. experiment 3, map B3).

Users tend to prefer maps on which the line colours are highly distinctive, and opt for straighter routes where this is not the case. Conventions considered to be meaningful (e.g. blue= motorway) appear to have been learned fairly quickly, whereas few people seem to be aware of green= primary route or the significance of primary routes in signposting. The tendency in a 5-class series to maximise travel on the top two classes and minimise use of the bottom class was also noted.

The level of comprehension of the restricted access junction symbols was surprisingly high, such that for all the common junction types symbols could be used that, despite their unfamiliarity, would only be misinterpreted by a maximum of 20% of road map users if full attention is paid to them. The strongest relevant colour connotations are red=barred versus white=access.

10.1 Road Map Design Recommendations

The major recommendation that can be made from this study is that the hierarchy of the symbols, whilst not being the major influence on route choice in most map designs, should reflect in its steepness and smoothness the relative quality of the roads. For example, the quantitative distinction between motorways and rural dual carriageways should be less than that between the latter and rural single carriageways. Major discontinuities in road quality are best represented by a change in colour. Highly-saturated reds are appropriate for the top class in a series, but in violating existing conventions or trying to establish new ones, map uses other than fastest-route planning must also be considered. This is especially the case on larger scale road maps which are particularly used for navigation, the purpose for which the blue=motorway and green=primary route conventions are theoretically appropriate. However, given people's lack of use of the primary route convention, presumably based on the lack of perceived or actual advantage in quality of many primary routes over 'other A roads', it cannot be considered to be sacrosanct.

Qualitative differences between classes should be as even as possible through the series, otherwise individual very different classes may not be interpreted as roads at all. In administratively-based schemes with redundant continuity, the use of widespread hues and a minimum colour difference of about 14 Farmer units between lines differentiated solely by colour are desirable. This could still allow (except with very narrow lines) for the use together of a fully-saturated red, yellow, green, blue, and purple with black, even without differences in width and character.

Where continuity is disrupted by a classification partially or wholly based on road quality, use of less contrasting and more harmonious hues is desirable to avoid violent colour changes in the middle of a road section. In such circumstances, particular care is required to ensure that the distinctions are not lost in the printing process. 'Two-ended' colour schemes sometimes used in these circumstances do not appear to function well, and the avoidance of slow, urban roads is best encouraged by a prominent depiction of the extent of the built-up area. Care should also be taken to ensure that classes in which many adjacent roads are grouped (e.g. medium speeds on a speed map) are not shown too prominently, in order that the monotony might not impede the perception of hierarchy. Where a change in line width is precluded, use of an increasing lightness/ decreasing chroma colour series is necessary, with casings to eliminate potential dazzle or fluorence. Grey casings (preferably not screened) might be considered as a compromise to maintain the distinctiveness and order of the filling colours as much as possible. Such a series could be extended if necessary by the use at the top end of thick, saturated coloured casings, or a fluorescent colour, although this would restrict the variety of conditions in which the map could be usefully employed. Recommendations for symbols for restricted access junctions are provided in section 9.5.1.

These recommendations are hardly revolutionary, and some existing maps incorporate them to some extent, such as the aforementioned practice on the Shell Road Atlas (George Philip,1983) of showing stretches of dual carriageway (red) along a primary (pink) or non-primary A road (amber) in a contrasting but harmonious colour. New road map display technologies such as videodisc and CRT also open up new problems for the representation and perception of colour in particular which this study has barely touched on. The findings presented above could make a minor contribution towards the recovery of the unwanted resource wastage. However, one particularly intriguing problem remains: how to convince road map producers that it might be in their interests to incorporate the suggested design changes.

APPENDIX A

PILOT STUDY: MAP SPECIFICATIONS

Listed below are those aspects of the specifications of the respective maps which are relevant to the specific journeys planned in the pilot study.

MAP 1A: Ordnance Survey Motoring Atlas
(Process printing)

Casing colour: Black

Class	Filling Colour		Line width	(casings)
Motorway	Blue	10B 6/8	1.4mm	(0.3mm)
A roads	Magenta	2.5RP 6/8	1.0	(0.1)
B roads	Brown	10R 4/8	1.0	(0.1)
Unclassified	Unfilled		0.5	(0.1)

Background	Layers, mostly 10YR 9/1
Towns	Yellow, 7.5Y 9/10
Road Numbers	magenta, unboxed, with prefixes. Trunk roads differentiated by (T) suffix.

MAP 1B: Map Productions RAC South Wales
(5 printing colours)

Class	Filling Colour		Line width	(casings)
Motorway	Blue	10B 7/8	2.1mm	(0.4mm, blue)
Primary	Yellow	7.5Y 9/10	1.8	(0.5, green)
Other A	Orange	5YR 7/10	1.0	(0.3, grey)
B roads	Cream	7.5Y 9/6	1.0	(0.3, grey)
Unclassified	Unfilled		0.75	(0.2, grey)

Green casing 10GY 5/8

Background	White
Towns	Purple-grey
Road Numbers:	
Motorway	Boxed, white on blue, with prefixes
Primary	Boxed, yellow on green, without prefixes
Other A/B	Unboxed, red, without prefixes

MAP 2A: Geographers' A-Z Road Atlas
(5 printing colours)

Class	Filling Colour		Line width (casings)	
Primary	Green	5GY 7/8	1.3mm	(0.15, black)
Other A	Red	7.5R 6/12	1.3	(0.15, black)
B roads	Yellow	10YR 8/10	1.3	(0.15, black)
Unclassified	Unfilled		0.6	(0.1, grey)

Background	Layers, mostly 7.5Y 9/4
Towns	Peach, 5YR 9/4
Road Numbers:	
Primary	Boxed, yellow on green, with prefixes
Other A/B	Unboxed, red, with prefixes

MAP 2B: Shell Road Atlas
(5 printing colours)

Casing colour: Black

Class	Filling Colour		Line width	(casings)
Primary	Pink	5R 7/8	1.1mm	(0.1mm)
Other A	Orange	5YR 7/10	0.8	(0.1)
B roads	Yellow	7.5Y 9/8	0.8	(0.1)
Unclassified	Unfilled		0.8	(0.1)

Background	White
Towns	Orange, 10YR 9/4
Road Numbers:	
Primary	Boxed, yellow on green, no prefixes
Other A/B	Black, outline boxes, no prefixes

MAP 3A: Ordnance Survey Routeplanner
(Process printing)

Class	Filling Colour	Line width (casings)
Motorway	Unfilled	1.1mm (0.45, cyan)
Primary dual	Yellow 7.5Y 9/10	1.1 (0.45, magenta)
Primary single	Magenta 7.5RP 5/12	0.7 Uncased
Other A	Magenta 7.5RP 5/12	0.5 Uncased
B/ unclass.	Brown 5YR 6/10	0.2 Uncased

Cyan casing	7.5B 7/8
Background	Buff, 2.5Y 9/2
Towns	Yellow, 7.5Y 9/10
Road Numbers:	
Motorway	Cyan, outline boxes, with prefixes
Primary	Green, outline boxes, with prefixes
Other A	Magenta, outline boxes, with prefixes
B roads	Brown, unboxed, with prefixes

MAP 3B: Road Speed Map of South East England
(6 printing colours)

Speed Class	Filling Colour	
Fast	Red	7.5R 6/12
Fairly fast	Brown	2.5YR 6/6
Medium	Yellow	5Y 8.5/10
Fairly slow	Yellow-grey	5Y 8/6
Slow	Grey	N 7.0

Casing colour: Black

Administrative Class	Line Width	Casings
Motorway	1.2 mm	0.3 mm
Primary Route	0.9	0.3/0.1
Other	0.7	0.1

Background	White
Towns	Point symbols only
Road Numbers	Black, unboxed, with prefixes

MAP 4A: AA Touring Map
(5 printing colours)

Casing colour: Black

Class	Filling Colour		Line width	(casings)
Motorway	Blue	7.5B 7/6	1.3mm	(0.25mm)
Primary dual	Green	7.5GY 7/8	1.0	(0.15)
Primary single	Green	7.5GY 7/8	0.6	(0.1)
Other A dual	Pink	10RP 7/8	1.0	(0.15)
Other A single	Pink	10RP 7/8	0.6	(0.1)

Motorways and dual carriageways have an additional black 0.1mm centre line.

Background	Cream, 5Y 9/2
Towns	Peach, 2.5Y 9/4
Road Numbers:	
Motorway	Black, outline boxes, with prefixes
A roads	Red, unboxed, with prefixes
B roads	Black, unboxed, with prefixes

MAP 4B: Map Productions England and Wales
(5 printing colours)

Casing colour: Red, 5R 5/14

Class	Filling Colour		Line width	(casings)
Motorway	Blue	10B 5/10	1.0mm	Uncased
Primary dual	Yellow	2.5Y 8.5/8	1.1	(0.3mm)
Primary single	Yellow	2.5Y 8.5/8	0.7	(0.1)
Other A dual	Unfilled		1.1	(0.3)
Other A single	Red	5R 5/14	0.5	Uncased

Background	White (Peak District: Cream, 10Y 9/2)
Towns	Grey, N 8.0
Road Numbers:	
Motorway	Boxed, white on blue, with prefixes
Primary	Boxed, yellow on red, without prefixes
Other A	Red, unboxed, without prefixes

APPENDIX B. EXPERIMENT 1: SPECIFICATIONS AND SCORES

TABLE B.1 STIMULUS ARRAY SPECIFICATIONS

PRINTING COLOURS

Colour	Selected Colour Munsell	Selected Colour Pantone	Actual Colour Munsell	x	y	Y
Red	7.5R 6/12	032U	7.5R 5/14	0.5889	0.3071	19.04
Green	10GY 5/8	362U	10GY 5/10	0.2755	0.5417	20.96
Yellow	7.5Y 9/10	102U	7.5Y 8.5/12	0.4348	0.4930	73.20
Blue	7.5B 7/6	306U	7.5B 6/8	0.1724	0.2321	29.42
Brown	10R 4/8	491U	5R 3/6	0.4678	0.3026	5.77
Black		Process	N 2.0	0.3274	0.3315	2.44
Background Colours						
Yellow	10% tint,	102U	7.5Y 9/2	0.3314	0.3470	84.85
Brown	10% tint,	491U	5R 8/1	0.3148	0.3133	71.87
White			N 9.5	0.3113	0.3175	86.22

CIE (x,y,Y) specifications are based on Illuminant D_{65} (artificial daylight).

LINE SPECIFICATIONS (WHITE BACKGROUND)

Line Number	Line Width (mm) Designed	Line Width (mm) Actual	(Filling) Colour	Casing Width (mm)
1	0.2	0.2	Red	-
2	0.2	0.15	Green	-
3	0.2	0.15	Yellow	-
4	0.2	0.15	Blue	-
5	0.2	0.15	Brown	-
6	0.2	0.15	Black	-
7	0.7	0.7	Red	-
8	0.7	0.7	Green	-
9	0.7	0.7	Yellow	-
10	0.7	0.7	Blue	-
11	0.7	0.7	Brown	-
12	0.7	0.65	Black	-
13	0.7	0.7	Red	0.1
14	0.7	0.7	Green	0.1
15	0.7	0.7	Yellow	0.1
16	0.7	0.7	Blue	0.1
17	0.7	0.7	Brown	0.1
18	0.7	0.7	White	0.1
19	1.0	1.05	Black	-
20	1.0	1.05	Red	-
21	1.0	1.05	Green	-
22	1.0	1.05	Yellow	-
23	1.0	1.1	Blue	

/cont'd

Table B.1: Line Specifications (cont'd)

Line Number	Line Width (mm) Designed	Actual	(Filling) Colour	Casing Width (mm)
24	1.0	1.05	Brown	-
25	1.0	1.0	White	0.1
26	1.0	1.0	Red	0.1
27	1.0	1.0	Green	0.1
28	1.0	0.95	Yellow	0.1
29	1.0	1.0	Blue	0.1
30	1.0	1.0	Brown	0.1
31	1.0	1.0	Brown	0.3
32	1.0	1.0	White	0.3
33	1.0	1.0	Red	0.3
34	1.0	1.0	Green	0.3
35	1.0	1.0	Yellow	0.3
36	1.0	1.0	Blue	0.3
37	1.4	1.4	Brown	-
38	1.4	1.4	Black	-
39	1.4	1.4	Blue	-
40	1.4	1.4	Red	-
41	1.4	1.4	Green	-
42	1.4	1.4	Yellow	-
43	1.4	1.4	Blue	0.1
44	1.4	1.4	Brown	0.1
45	1.4	1.4	Red	0.1
46	1.4	1.4	Green	0.1
47	1.4	1.4	Yellow	0.1
48	1.4	1.4	White	0.1
49	1.4	1.4	Red	0.3
50	1.4	1.4	Green	0.3
51	1.4	1.4	Yellow	0.3
52	1.4	1.4	Blue	0.3
53	1.4	1.4	Brown	0.3
54	1.4	1.4	White	0.3

Line Number	Line Width (mm) Designed	Actual	(Filling) Colour	Casing Colour
55	1.0	1.0	White	Blue
56	1.0	1.05	White	Red
57	1.0	1.0	Yellow	Red
58	1.0	1.0	Yellow	Green
59	1.4	1.5	White	Blue
60	1.4	1.4	White	Red
61	1.4	1.4	Yellow	Red
62	1.4	1.4	Yellow	Green

All coloured casings are 0.3mm wide.

TABLE B.2 MEAN PERCEIVED MAGNITUDE SCORES (WHITE BACKGROUND)

Line No.	Specification	Mean Score	Rank
1	Red,0.2,uncased	-1.572	57
2	Green,0.15,uncased	-2.078	58
3	Yellow,0.15,uncased	-2.577	62
4	Blue,0.15,uncased	-2.12	59
5	Brown,0.15,uncased	-2.158	60
6	Black,0.15,uncased	-2.228	61
7	Red,0.7,uncased	-0.016	39
8	Green,0.7,uncased	-0.33	49
9	Yellow,0.7,uncased	-0.842	55
10	Blue,0.7,uncased	-0.316	48
11	Brown,0.7,uncased	-0.306	47
12	Black,0.65,uncased	-0.355	50
13	Red,0.7,thin	-0.206	43=
14	Green,0.7,thin	-0.231	46
15	Yellow,0.7,thin	-0.703	53
16	Blue,0.7,thin	-0.206	43=
17	Brown,0.7,thin	-0.352	51
18	White,0.7,thin	-1.153	56
19	Black,1.05,uncased	+0.565	17
20	Red,1.05,uncased	+0.538	18
21	Green,1.05,uncased	+0.2	32
22	Yellow,1.05,uncased	-0.176	42
23	Blue,1.1,uncased	+0.333	23
24	Brown,1.05,uncased	+0.264	28
25	White,1.0,thin	-0.754	54
26	Red,1.0,thin	+0.494	19
27	Green,1.0,thin	+0.29	24
28	Yellow,0.95,thin	-0.087	40
29	Blue,1.0,thin	+0.241	29=
30	Brown,1.0,thin	+0.279	25
31	Brown,1.0,thick	+0.421	21
32	White,1.0,thick	-0.108	41
33	Red,1.0,thick	+0.425	20
34	Green,1.0,thick	+0.241	29=
35	Yellow,1.0,thick	+0.107	36
36	Blue,1.0,thick	+0.269	26
37	Brown,1.4,uncased	+0.783	11
38	Black,1.4,uncased	+0.901	7
39	Blue,1.4,uncased	+0.651	15
40	Red,1.4,uncased	+1.058	2
41	Green,1.4,uncased	+0.614	16
42	Yellow,1.4,uncased	+0.14	35
43	Blue,1.4,thin	+0.793	10
44	Brown,1.4,thin	+0.919	6
45	Red,1.4,thin	+0.988	3
46	Green,1.4,thin	+0.885	8
47	Yellow,1.4,thin	+0.356	22
48	White,1.4,thin	-0.549	52
49	Red,1.4,thick	+1.067	1
50	Green,1.4,thick	+0.677	14
51	Yellow,1.4,thick	+0.707	12

/cont'd

Table B.2: Mean Perceived Magnitude Scores (cont'd)

Line No.	Specification	Mean Score	Rank
52	Blue,1.4,thick	+0.84	9
53	Brown,1.4,thick	+0.972	4
54	White,1.4,thick	+0.096	37
55	Blue/white,1.0	-0.221	45
56	Red/white,1.05	+0.196	33
57	Red/yellow,1.0	+0.174	34
58	Green/yellow,1.0	+0.028	38
59	Blue/white,1.5	+0.213	31
60	Red/white,1.4	+0.267	27
61	Red/yellow,1.4	+0.935	5
62	Green/yellow,1.4	+0.695	13

TABLE B.3 LINE SPECIFICATIONS (COLOURED BACKGROUNDS EXPERIMENT)

Line Number	Specification (colour,	casing,	width)	Equivalent Line, White background
1	Red,	uncased,	0.7 mm	7
2	Green,	uncased,	1.0	21
3	Yellow,	uncased,	1.4	42
4	Blue,	uncased,	0.7	10
5	Brown,	uncased,	1.0	24
6	Black,	uncased,	1.4	38
7	Red,	thin,	1.0	26
8	Green,	thin,	1.4*	46
9	Yellow,	thin,	0.7*	15
10	Blue,	thin,	1.4*	43
11	Brown,	thin,	0.7*	17
12	White,	thin,	1.4	48
13	Red,	thick,	1.0	33
14	Green,	thick,	1.4	50
15	Yellow,	thick,	1.4	51
16	Blue,	thick,	1.0	36
17	Brown,	thick,	1.0	31
18	White,	thick,	1.4	54
19	White,	blue,	1.0	55
20	White,	red,	1.0	56
21	Yellow,	red,	1.0	57
22	Yellow,	green,	1.4	62
23	White,	uncased,	1.0	(brown background only)

* Lines which on either or both of the coloured backgrounds slightly exceed their designed width.

TABLE B.4 MEAN PERCEIVED MAGNITUDE SCORES (COLOURED BACKGROUNDS)

Line Number	Specification	Mean Scores Yellow Bckgd.	Brown Bckgd.
1	Red,uncased,0.7	-0.28	+0.005
2	Green,uncased,1.0	-0.157	+0.102
3	Yellow,uncased,1.4	+0.305	-0.22
4	Blue,uncased,0.7	-0.415	-0.329
5	Brown,uncased,1.0	-0.146	+0.133
6	Black,uncased,1.4	+1.057	+0.315
7	Red,thin,1.0	+0.403	+0.512
8	Green,thin,1.4*	+0.648	+0.712
9	Yellow,thin,0.7*	-0.586	-0.091
10	Blue,thin,1.4*	+0.473	+0.678
11	Brown,thin,0.7*	-0.471	-0.092
12	White,thin,1.4	-1.12	-1.356
13	Red,thick,1.0	-0.151	+0.325
14	Green,thick,1.4	+0.432	+0.541
15	Yellow,thick,1.4	+0.487	+0.162
16	Blue,thick,1.0	+0.013	-0.005
17	Brown,thick,1.0	+0.568	+0.348
18	White,thick,1.4	-0.266	-0.093
19	White,blue,1.0	-0.545	+0.071
20	White,red,1.0	-0.685	+0.023
21	Yellow,red,1.0	+0.018	+0.126
22	Yellow,green,1.4	+0.418	+0.089
23	White,uncased,1.0	-	-1.956

* Lines which on either or both of the coloured backgrounds slightly exceed their designed width.

TABLE B.5 RANKINGS OF MEAN SCORES ON EACH BACKGROUND FOR EQUIVALENT LINES

Line Number	Specification	Background Colour White	Yellow	Brown
1	Red,uncased,0.7	17	16	15
2	Green,uncased,1.0	12	14	11
3	Yellow,uncased,1.4	15	9	20
4	Blue,uncased,0.7	19	17	21
5	Brown,uncased,1.0	11	12	9
6	Black,uncased,1.4	1	1	7
7	Red,thin,1.0	7	8	4
8	Green,thin,1.4*	2	2	1
9	Yellow,thin,0.7*	22	20	17
10	Blue,thin,1.4*	3	5	2
11	Brown,thin,0.7*	20	18	18
12	White,thin,1.4	21	22	22
13	Red,thick,1.0	8	13	6
14	Green,thick,1.4	6	6	3
15	Yellow,thick,1.4	4	4	8
16	Blue,thick,1.0	10	11	16
17	Brown,thick,1.0	9	3	5
18	White,thick,1.4	16	15	19
19	White,blue,1.0	18	19	13
20	White,red,1.0	13	21	14
21	Yellow,red,1.0	14	10	10
22	Yellow,green,1.4	5	7	12
23	White,uncased,1.0	-	-	23

* Lines which on either or both of the coloured backgrounds slightly exceed their designed width.

Overall rank correlations between:

White and yellow backgrounds	0.883
White and brown backgrounds	0.833
Yellow and brown backgrounds	0.743

APPENDIX C. EXPERIMENT 2: SPECIFICATIONS AND SCORES

TABLE C.1 ARRAY SPECIFICATIONS

Array no.	Target Line	Background line types 1	2	3
1	Red,1.4,u	Brown,0.7,u	Green,0.7,u	White,0.7,n
2	Yellow,0.7,u	Brown,1.0,u	Green,1.0,k	Blue,1.0,k
3	Yellow,1.4,u	Brown,1.0,u	Green,1.0,k	Blue,1.0,k
4	Red,1.4,u	Brown,1.0,k	Green,1.0,k	Black,1.0,u
5	Red,1.4,u	Yellow,0.7,u	Yellow,1.0,u	Yellow,1.4,u
6	White,0.7,n	White,1.0,n	Green,1.4,k	Red,1.4,k
7	Brown,0.7,u	Black,0.7,u	Red,1.0,u	Green,1.4,u
8	Blue,1.0,u	Blue,0.7,u	Blue,1.4,u	Blue,1.4,k
9	Blue,0.7,u	Green,1.0,u	Red,1.0,k	Yellow,1.4,k
10	Blue,0.7,u	Red,1.0,u	Brown,1.4,k	White,1.4,k
11	Yellow,1.0,k	White,1.0,k	Yellow,1.4,k	Blue,0.7,u
12	Green,0.7,u	Brown,1.0,k	Yellow,1.0,k	Blue,1.0,u
13	Green,1.4,u	Brown,1.0,k	Yellow,1.0,k	Blue,1.0,u
14	Black,1.4,u	White,1.4,k	Green,1.0,u	Blue,1.4,u
15	Black,1.4,u	Red,1.4,k	Green,1.0,u	Blue,1.4,u
16	White,1.4,n	Red,1.4,u	Brown,1.4,u	Black,1.4,u
17	Black,1.4,u	Green,1.4,k	Brown,1.4,k	White,1.4,n
18	Blue,1.4,k	Blue,1.0,k	Red,0.7,u	Green,1.4,u
19	Red,0.7,u	Brown,0.7,u	Yellow,1.4,u	White,1.0,k
20	Red,0.7,u	Blue,0.7,u	Yellow,1.4,u	White,1.0,k

Key: u uncased
n thin (0.1mm) casing
k thick (0.3mm) casing

TABLE C.2 MEDIAN SEARCH TIMES

Rank	Array Number	Median Time (secs.)
1	5	0.81
2	4	0.92
3	3	1.05
4	2	1.17
5	1	1.17
6	20	1.18
7	10	1.20
8	15	1.29
9	9	1.34
10	16	1.35
11	19	1.47
12	14	1.57
13	13	1.82
14	12	2.07
15	17	2.72
16	18	4.73
17	11	6.05
18	6	7.05
19	7	7.40
20	8	11.55

APPENDIX D. EXPERIMENT 3: MAP SPECIFICATIONS AND SUBJECTS' CHARACTERISTICS

1. MAP SPECIFICATIONS

MAP 1A

Road Class	Colour/ Munsell Specification	Line Width (mm)	Magnitude Estimate
1	Red, 7.5R 5/14	1.05	0.605 c
2	Green, 10GY 5/10	1.05	0.215 c
3	Blue, 7.5B 6/8	0.7	-0.350 c
4	Black N 2.0	0.4	-0.955 c
5	White (0.1mm casing)	0.4	-1.567 c

C=computed, O=observed (in experiment 1). Magnitude estimates could not be calculated for cased coloured lines as no totally consistent formula was discovered. Thus the choice of these lines is limited to those used in the previous experiment, with their observed values. The affective value element of the computed estimates is based on a two-thirds male, one-third female sample.

MAP 1B/2A

Road Class	Colour/ Munsell Specification	Width (mm) Total	Width (mm) Casing	Magnitude Estimate
1	Blue, 7.5B 6/8	1.1	0.2	0.43 o
2	Green, 10GY 5/10	1.0	0.1	0.29 o
3	Red, 7.5R 5/14	1.0	0.1	0.494 o
4	Yellow, 7.5Y 8.5/12	0.7	0.1	-0.703 o
5	White	0.4	0.1	-1.567 c

MAP 2B

Road Class	Colour/ Munsell Specification	Line Width (mm)	Magnitude Estimate
1	Purple, 5P 4/10	1.0	0.541 c
2	Maroon, 10RP 4.5/9	1.0	0.268 c
3	Pink, 5R 6/10	1.0	0.034 c
4	Orange, 5YR 7/10	1.0	-0.179 c
5	Yellow, 5Y 8.5/10	1.0	-0.353 c

MAP 3A

Road Class	Colour/ Munsell Specification		Width (mm) Total	Casing	Magnitude Estimate
Dual	Red/	7.5R 5/14		0.3	
	Yellow,	7.5Y 8.5/12	1.0		0.174 o
S 1	Brown,	5R 3/6	1.05		0.264 o
S 2	Red,	7.5R 5/14	0.7		-0.016 o
S 3	Red,	7.5R 5/14	0.7	0.1	-0.206 o
S 4	Green,	10GY 5/10	0.4		-1.24 c

S=single

On maps 3A and 3B, dual road was allocated to classes 1 to 4 in the ratio 12:7:2:0

MAP 3B

Road Class	Colour/ Munsell Specification		Width (mm) Total	Casing	Magnitude Estimate
D 1	Red,	7.5R 5/14	1.4	0.3	1.067 o
2	Green,	10GY 5/10	1.4	0.3	0.677 o
3	Yellow,	7.5Y 8.5/12	1.4	0.3	0.707 o
S 1	Red,	7.5R 5/14	1.4	0.1	0.988 o
2	Green,	10GY 5/10	1.05	-	0.200 o
3	Yellow,	7.5Y 8.5/12	1.0	0.1	-0.087 o
4	White.		0.7	0.1	-1.153 o

D=dual, S=single

2. SUBJECTS' PERSONAL CHARACTERISTICS

Full Sample (248 subjects)

Totals by Social Class

Social Class	Numbers Set A	Set B	Percentages In sample	U.K. drivers
I	17	13	12.1	4.4
II	26	37	25.4	22.7
IIIN	14	15	11.7	13.6
IIIM	22	25	19.0	22.7
IV	8	8	6.5	9.2
V	3	2	2.0	1.8
Retired	13	9	8.9	7.6
Other	21	15	14.5	18.2
Total	124	124	(100.1)	(100.2)

Figures for 'U.K. drivers' are from a combination of DTp (1985) and OPCS (1983).

Representative Sample (200 subjects)

Totals by Social Class

Social Class	Numbers Set A	Set B	% of U.K. drivers
I	4	4	4.4
II	23	23	22.7
IIIN/M	36	39	36.3
IV/V	11	10	11.0
Ret/other	26	24	25.8
Total	100	100	(100.2)

Classes with similar educational levels (IIIN and IIIM, IV and V) have been grouped together. Where the quotas could not quite be met in set B some substitution has been necessary.

Representative Sample (cont'd)

Totals by Age and Sex

Age range	No. of males		No. of females		Total
Under 25	16		6		22
25-44	61		33		94
45-64	45		22		67
Over 64	11		6		17
Total	133	(66.5%)	67	(33.5%)	
U.K. drivers		67.0%		33.0%	

Set A: 68 males, 32 females
Set B: 65 males, 35 females

BIBLIOGRAPHY

(* indicates reference consulted but not referred to in the text)

AASHO (1962), Report on Uniform Map Symbols. American Association of State Highway Officials, Washington D.C.

Adams K.R. (1967), Cartographic Design Standards- A Preliminary Investigation. Final technical report for the European Research Office, U.S. Army, contract DAJA 37-671C-0413.

Agoston G.A. (1979), Color Theory and its Application in Art and Design. Springer-Verlag, Berlin.

Allen E.C. and Guilford J.P. (1936), Factors determining the Affective Values of Color Combinations. American Journal of Psychology, 48, p.643-648

Anderson J.M. (1975), A Road Map Study based on the 1974 Official Alberta Road Map. Unpublished MSc thesis, University of Alberta, Edmonton.

Anderson J.M. (1977), Map User Research and Modern Map Design. Canadian Surveyor, 31 (1), p.52-64.

Armstrong B.D. (1977), The Need for Route Guidance. Supplementary Report 330, Transport and Road Research Laboratory, Crowthorne, Berks.

Arnberger E. (1966), Handbuch der Thematischen Kartographie. Franz Deuticke, Vienna.

Arnheim R. (1954), Art and Visual Perception. Faber and Faber, London.

Astley R.W. (1969), A Note on the Requirements of Road Map Users. Cartographic Journal, 6 (2), p.130.

Audley R.J., Bickmore D.P. and Phillips R.J. (1974), Legibility Criteria for the Efficient Use of Maps. Final report to the Social Science Research Council, Project HR2167.

Barbur J.L. (1979), Visual Periphery. In: Clare J.N. and Sinclair M.A. (eds.), Search and the Human Observer. Taylor and Francis, London.

Bartholomew J.C. and Kinniburgh I.A.G. (1973), The Factor of Awareness. Cartographic Journal, 10 (1), p.59-62.

Bartlett Sir F.C. (1958), Thinking: An Experimental and Social Study. George Allen and Unwin, London.

Bartram D.J. (1980), Comprehending Spatial Information: The Relative Efficiency of Different Methods of Presenting Information about Bus Routes. Journal of Applied Psychology, 65 (1), p.103-110

Bartz B.S. (1970), Experimental Use of the Search Task in an Analysis of Type Legibility in Cartography. Cartographic Journal, 7 (2), p.103-112.

BBC (British Broadcasting Corporation) (1985), Colourful Notions. BBC2 'Horizon' programme, first broadcast 7th January 1985.

Beck J. (1975), The Perception of Surface Colour. Scientific American, 233, p.62-75.

Benshoof J.A. (1970), Characteristics of Drivers' Route Selection Behaviour. Traffic Engineering and Control, 11, p.604-606,609

Berlin B. and Kay P. (1969), Basic Colour Terms: their Universality and Evolution. University of California Press, Berkeley, CA.

Bertin J. (1981), Graphics and Graphic Information Processing. Translated by W.J.Berg and P.Scott. Walter de Gruyter, Berlin.

Bertin J. (1983a), Semiology of Graphics. Translation by W.J.Berg. University of Wisconsin Press, Madison,WI.

Bertin J. (1983b), A New Look at Cartography. In: D.R.F.Taylor (ed.), Graphic Communication and Design in Contemporary Cartography, p.69-86. John Wiley, Chichester.

Bevan W. and Dukes W.F. (1953), Color as a Variable in the Judgment of Size. American Journal of Psychology, 66, p.283-288.

Birren F. (1930), A Chart to Illustrate the Fundamental Discrepancies of Color Study in Physics, Art and Psychology. Psychological Review, 37, p.271-273.

Birren F. (1941), The Story of Color. Crimson Press, Westport, Conn.

Birren F. (1969), Light, Color and Environment. Van Nostrand Reinhold Co., New York.

Bishop H.P. and Crook M.N. (1961), Absolute Identification of Color for Targets Presented against White and Coloured Backgrounds. Wright Air Development Division Technical Report 60-611.

Board C. (1975), Map Reading Tasks Appropriate in Experimental Studies in Cartographic Communication. Paper presented to the 1st International Symposium on Cartographic Communication, London.

Board C. and Buchanan A. (1974), Evaluating Thematic Maps- Current Trends in the United Kingdom. Cartographic Journal, 11 (2), p.124-129.

Bouma H. (1980), Visual Reading Processes and the Quality of Text Displays. In: Grandjean E. and Vigliani E. (eds.), Ergonomic Aspects of Visual Display Terminals. Taylor and Francis, London.

Bourdon-Michelin M. (1952), L'histoire d'une Carte Routière. Société des Ingénieurs Civils de France, Mémoires, p.309-338.

Boynton R.M. (1978), Color in Contour and Object Perception. In: Carterette E.C. and Friedman M.P. (eds.), Handbook of Perception, Volume VIII Perceptual Coding, p.173-199. Academic Press Inc., London.

Boynton R.M. (1979), Human Color Vision. Holt, Rinehart and Winston, New York.

Braidwood,S. (1981), Birth of the Intelligible Bus Map. Design, 2393, p.53-54.

Bruce M. and Foster J.J. (1982), The Visibility of Coloured Characters on Coloured Backgrounds in Viewdata Display. Visible Language, 16 (4), p.382-390.

BSI (British Standards Institution) (1970), Specification for Colours for High Visibility Clothing BS4610. BSI, London.

Bundesen C. and Pedersen L.F. (1983), Color Segregation and Visual Search. Perception and Psychophysics, 33 (5), p.487-493.

Burnham R.W. (1951), The Dependence of Color upon Area. American Journal of Psychology, 64, p.521-533.

Cahill M.C. and Carter R.C. (1976), Color Code Size for Searching Displays of Different Density. Human Factors, 18 (3), p.273-280.

Campbell D.T. and Stanley (1966), Experimental and Quasi-Experimental Designs for Research. Rand McNally, Chicago.

Carpenter S.M. (1979), Drivers' Route Choice Project: Pilot Study. Report prepared by Oxford University Transport Studies Unit for the Transport and Road Research Laboratory.

Carter E.C. and Carter R.C. (1981), Color and Conspicuousness. Journal of the Optical Society of America, 71 (6), p.723-729.

Castiglioni M. (1961), Les Récentes Publications Cartographiques du Touring Club Italiano. International Yearbook of Cartography, I, p.182-189.

Castner H.W. (1978), Perceptual Processes during the Initiation of Map Viewing. Paper presented to the 9th International Cartographic Conference.

Castner H.W. (1983), Research Questions and Cartographic Design. In: D.R.F.Taylor (ed.), Graphic Communication and Design in Contemporary Cartography, p.87-113. John Wiley, Chichester.

Castner H.W. and Robinson A.H. (1969), Dot Area Symbols in Cartography: The Influence of Pattern on their Perception. American Congress on Surveying and Mapping (Cartographic Division) Technical Monograph CA-4.

Chaloogin E.I. (1967), Application of Photoluminescence in Cartography. International Yearbook of Cartography, VII, p.93-96.

Christ R.E. (1975), Review and Analysis of Color Coding Research for Visual Displays. Human Factors, 17 (6), p.542-570.

Christner C.A. and Ray H.W. (1961), An Evaluation of Selected Combinations of Target and Background Coding on Map Reading Performance. Human Factors, 3 (2), p.131-146.

Clark B.A.J. (1985), Field Measurement of Visibility. Paper presented to 14th Commonwealth Defence Conference on Operational Clothing and Combat Equipment, Melbourne.

Clarke J.I. (1959), Statistical Map Reading. Geography, 44, p.96-104.

Comrey A.L. (1950), A Proposed Method for Absolute Ratio Scaling. Psychometrika, 15, p.317-325.

Connal K.A. (1983), The Usefulness of Environmental Information for Activities in Road Map Use. Unpublished MA thesis, Queen's University, Kingston, Ontario.

Consumers' Association (1963), Motoring Maps. Which?, July, p.196-203.

Consumers' Association (1971), Road Maps. Motoring Which?, April, p.75-79.

Consumers' Association (1979), Road Maps. Which?, July, p.388-390.

Consumers' Association (1983a), Good Map Guide, Holiday Which?, May, p.152-159.

Consumers' Association (1983b), Road Atlases. Which?, December, p.572-575.

*Cooke D.F. (1985), Vehicle Navigation Appliances. Auto-Carto VII, Proceedings, p.108-115.

Coren S. and Girgus J.S. (1980), Principles of Perceptual Organisation and Spatial Distortion: The Gestalt Illusions. Journal of Experimental Psychology, Human Perception and Performance, 6 (3), p.404-412.

Cox C.W. (1973), Adaptation-level Theory as an Aid to the Understanding of Map Perception. Proceedings of the American Congress on Surveying and Mapping, 33rd Annual Meeting, p.334-369.

Crandall C.J. (1973), Fluorescent Map and Chart Development. Cartographic Journal, 10 (2), p.95-98.

Crawford P.V. (1971), The Perception of Grey-tone Symbols. Annals of the Association of American Geographers, 61, p.721-735.

Cuff D.J. (1972a), The Magnitude Message: A Study of Effectiveness of Colour Sequences on Quantitative Maps. Unpublished PhD thesis, Geography Department, Pennsylvania State University.

Cuff D.J. (1972b), Value versus Chroma in Colour Schemes on Quantitative Maps. Canadian Cartographer, 9 (2), p.134-140.

Cuff D.J. (1973), Colour on Temperature Maps. Cartographic Journal, 10 (1), p.17-20.

Cuff D.J. (1975), Conflicting Goals in Choosing Colours for Quantitative Maps. Proceedings of the International Symposium on Computer-Assisted Cartography, Auto-Carto II, p.286-288.

Davidoff J. (1974), The Psychological Relationship between Lightness and Saturation. Perception and Psychophysics, 16, p.79-83.

DeLucia A.A. (1976), How People Read Maps: Some Objective Evidence. Proceedings of the American Congress on Surveying and Mapping, 36th Annual Meeting, p.135-144.

Dent B.D. (1972), Visual Organisation and Thematic Map Communication. Annals of the Association of American Geographers, 62, p.79-93.

Department of Transport (1985), Transport Statistics Great Britain 1974-1984. H.M.S.O., London.

DeValois R.L. and DeValois K.K (1975), Neural Coding of Colour. In: Carterette E.C. and Friedman M.P. (eds.), Handbook of Perception, Volume V: Seeing, p.117-166.

Dewar R.E. and Ells J.G. (1984), Methods of Evaluation of Traffic Signs. In: Easterby R. and Zwaga H. (eds.), Information Design, p.77-90. John Wiley, Chichester.

Dixon O. (1986), The Motorist's Route Planner 1985 of Great Britain. Cartographic Journal, 23 (1), p.74.

Dobson M.W. (1979a), Visual Information Processing during Cartographic Communication. Cartographic Journal, 16 (1), p.14-20.

Dobson M.W. (1979b), The Influence of Map Information on Fixation Localisation. American Cartographer, 6 (1), p.51-65.

Dobson M.W. (1980), Alternative Research Strategies and Map Reading. Auto-Carto IV, Proceedings, p.474-480.

Dobson M.W. (1981), Effectiveness of the Visual over the Verbal: Additional Factors. Journal of Mental Imagery, 5, p.20-24.

Dobson M.W. (1983a), Human Factors in the Cartographic Design of Real-Time, Color Displays. Auto-Carto VI, Proceedings, p.421-426.

Dobson M.W. (1983b), A High Resolution Microcomputer-based Color System Examining the Human Factors Aspect of Cartography in a Real-Time User Environment. Auto-Carto VI, Proceedings, p.352-360.

Dobson M.W. (1983c), Visual Information Processing and Cartographic Communication: The Utility of Redundant Stimulus Dimensions. In: D.R.F.Taylor (ed.), Graphic Communication and Design in Contemporary Cartography, p.149-175. John Wiley, Chichester.

Dobson M.W. (1984), Effective Color Display for Map Task Performance in a Computer Environment. Proceedings of the International Symposium on Spatial Data Handling, Zurich, Volume II, p.332-348.

Dooley R.P. and Harkins L.E. (1970), Functional and Attention-Getting Effects of Color on Graphic Communications. Perceptual and Motor Skills, 31, p.851-854.

Drewitt B. (1973), The Changing Profile of the Map User in Great Britain. Cartographic Journal, 10 (1), p.42-48.

*Durbin J. and Watson G.S. (1951), Testing for Serial Correlation in Least Squares Regression II. Biometrika, 38, p.159-178.

Easterby R.S. (1980), A Commentary and Forecast for Colour Applications in Man/Machine Systems. Institute of Electrical Engineers, Digest 1980/76, p.8/1-8/4.

Easterby R.S. (1984), Tasks, Processes and Display Design. In: Easterby R.S. and Zwaga H. (eds.), Information Design, p.19-36. John Wiley, Chichester.

Eastman J.R. (1985), Cognitive Models and Cartographic Design Research. Cartographic Journal, 22 (2), p.95-101.

Eastman J.R. and Castner H.W (1983), The Meaning of Experience in Task-Specific Map Reading. In: D.R.F.Taylor (ed.), Graphic Communication and Design in Contemporary Cartography, p.115-147. John Wiley, Chichester.

Eckert M. (1925), Die Kartenwissenschaft, Band 2. De Gruyter, Berlin.

Edwards A.S. (1955), Effect of Color on Visual Depth Perception. Journal of General Psychology, 52, p.331-333.

Ekman G. and Junge K. (1961), Psychophysical Relations in Visual Perception of Length, Area and Volume. Scandinavian Journal of Psychology, 2, p.1-10.

Engel F.L. (1969), In Search of Conspicuity- Attempts to Measure and Comprehend Visual Conspicuity. I.P.O. Annual Progress Report, 4, p.89-95. Institute of Perception Research, Eindhoven, Netherlands.

Engel F.L. (1971), Visual Conspicuity, Directed Attention and Retinal Locus. Vision Research, 11, p.563-576.

Engel F.L. (1974), Visual Conspicuity and Selective Background Interference in Eccentric Vision. Vision Research, 14, p.459-471.

Engel F.L. (1977), Visual Conspicuity, Visual Search and Fixation Tendencies of the Eye. Vision Research, 17, p.95-98.

Engel F.L. and Bos T.M. (1971), Size and Luminance Contrast in Visual Conspicuity. I.P.O. Annual Progress Report, 6, p.45-49. Institute for Perception Research, Eindhoven, Netherlands.

Engel F.L., Ligtenbarg H.G. and Bos T.M. (1972), Conspicuity and Colour. I.P.O. Annual Progress Report, 7, p.55-62. Institute for Perception Research, Eindhoven, Netherlands.

Evans R.M. (1948), An Introduction to Color. Wiley, New York.

Evans R.M. (1959), Fluorescence and Gray Content of Surface Colours. Journal of the Optical Society of America, 49, p.1049-1059.

Evans R.M. (1974), The Perception of Color. Wiley, New York.

Farmer E.W. and Taylor R.M. (1980), Visual Search through Color Displays: Effects of Target-Background Similarity and Background Uniformity. Perception and Psychophysics, 27 (3), p.267-272.

Farmer E.W., Taylor R.M. and Belyavin A.J. (1980), Large Color Differences and the Geometry of Munsell Color Space. Journal of the Optical Society of America, 70 (2), p.243-245.

*Fisher R.A. and Yates F. (1963), Statistical Tables for Biological, Agricultural and Medical Research, 6th edition. Oliver and Boyd, London.

Flannery J.J. (1956), The Graduated Circle: A Description, Analysis and Evaluation of a Quantitative Map Symbol. Unpublished PhD thesis, University of Wisconsin, Madison,WI.

Forrest A.R. (1985), Antialiasing in Practice. In: R.A.Earnshaw (ed.), Fundamental Algorithms for Computer Graphics, p.113-134 . Springer-Verlag, Berlin.

Forrest D. and Castner H.W. (1985), The Design and Perception of Point Symbols for Tourist Maps. Cartographic Journal, 22 (1), p.11-19.

Frenzel K. (1965), Zur Frage des Optischen Gewicht von Signaturen für thematische Karten. Erdkunde, 19 (1), p.66-70.

Fullard H. (1965), Problems of Communication between Editors and Users of Atlases. International Yearbook of Cartography, V, p.187-192.

Fulton E.J., Kirkby C. and Stroud P.G. (1980), Daytime Motorcycle Conspicuity. Supplementary Report 625, Transport and Road Research Laboratory, Crowthorne, Berks.

Garland H.C., Haynes J.J. and Grubb G.C. (1979), Transit Map Color Coding and Street Detail: Effects on Trip Planning Performance. Environment and Behavior, 11 (2), p.162-184.

Garner W.R. and Felfoldy G.L. (1970), Integrality of Stimulus Dimensions in Various Types of Information Processing. Cognitive Psychology, 1, p.225-241.

Gatrell A.W. (1966), The Design and Production of Road Maps. Cartographic Journal, 3 (1), Symposium '65 p.25-28.

Gaussen H. (1964), La Couleur dans les Cartes Thématiques. International Yearbook of Cartography, IV, p.53-61.

Gill G.A. (1982), British Road Map Design and the Representation of Settlement. Unpublished diploma dissertation, University College, Swansea.

Gilmartin P.P. (1981), The Interface of Cognitive and Psychophysical Research in Cartography. Cartographica, 18 (3), p.9-20.

Ginsburg A.P. (1981), Proposed new Vision Standards for the 1980s and Beyond: Contrast Sensitivity. U.S.A.F. Medical Research Laboratory, Wright-Patterson Air Force Base, Report AFAMRL-TR-80-121.

Godlove I.H. (1951), Improved Color-Difference Formula, with Applications to the Perceptibility of Fadings. Journal of the Optical Society of America, 41, p.760-772.

Goldring P. (1978), The Development of Esso Road Maps. Society of University Cartographers Bulletin, 12 (1), p.70-72.

Gordon J. and Abramov I. (1977), Color Vision in the Peripheral Retina II: Hue and Saturation. Journal of the Optical Society of America, 67 (2), p.202-207.

Graves M. (1952), Color Fundamentals. McGraw Hill, New York.

Gray P.G. and Russell R. (1962), Drivers' Understanding of Road Traffic Signs. Social Survey Report SS347, Central Office of Information, London.

Green B.F. and Anderson L.K. (1956), Color Coding in a Visual Search Task. Journal of Experimental Psychology, 51, p.19-24.

Greenberg G.L. (1971), Irradiation Effect on the Perception of Map Symbology. International Yearbook of Cartography, XI, p.120-126.

*Gregory R.L. (1977), Eye and Brain- The Psychology of Seeing, 3rd edition. Weidenfeld and Nicolson, London.

Grether W.F. and Baker C.A. (1972), Visual Presentation of Information. In: Van Cott H.P. and Kinkade R.G. (eds.), Human Engineering Guide to Equipment Design. U.S.Government Printing Office, Washington D.C.

Grice G.R., Canham L. and Boroughs J.M. (1983), Forest before Trees? It depends where you look. Perception and Psychophysics, 33 (2), p.121-128.

Griffin T.L.C. (1983), Problem Solving on Maps- The Importance of User Strategies. Cartographic Journal, 20 (2), p.101-109.

Griffith D.L. and Kelly J.J. (1965), The Quarter Inch to One Mile Map of Great Britain. International Yearbook of Cartography, V, p.94-102.

Guilford J.P. (1954), Psychometric Methods. Tata McGraw-Hill Publishing Co., Bombay.

Guilford J.P. and Smith P.C. (1959), A System of Color-Preferences. American Journal of Psychology, 72 (4), p.487-502.

Gundlach C. and Macoubrey C. (1931), The Effect of Color on Apparent Size. American Journal of Psychology, 43, p.109-111.

Haber R.N. (1981), The Power of Visual Perceiving. Journal of Mental Imagery, 5, p.1-40.

Haber R.N. and Hershenson M. (1973), The Psychology of Visual Perception. Holt, Rinehart and Winston, New York.

*Hammond R. and McCullagh P.S. (1978), Quantitative Techniques in Geography- An Introduction, 2nd edition. Clarendon Press, Oxford.

Harms L. and Bundesen C. (1983), Color Segregation and Selective Attention in a Non-Search Task. Perception and Psychophysics, 33 (1), p.11-19.

Head C.G. (1984), The Map as Natural Language: A Paradigm for Understanding. Cartographica, 21 (1), p.1-32.

Hecht S. and Mintz E.U. (1939), The Visibility of Single Lines at Various Illuminations and the Retinal Basis of Visual Resolution. Journal of General Physiology, 22, p.593-612.

Heinemann E.G. (1972), Simultaneous Brightness Induction. In: Jameson D. and Hurvich L. (eds.), Handbook of Sensory Physiology, Volume VII/4. Springer-Verlag, New York.

Helson H. and Rohles F.H.Jr. (1959), A Quantitative Study of Reversal of Classical Lightness Contrast. American Journal of Psychology, 72, p.530-538.

Hiatt C.K. (1982), The Function of Color in Legibility of Linear Symbology on Maps for the Partially Sighted. PhD thesis, University of Washington. Dissertation Abstracts International, 43 (6), p.2098-A.

Hill A.R. (1975), Eye Movement Studies in Cartographic Communication: Problems of Experimental Design. Paper presented to the 1st International Symposium on Cartographic Communication, London.

Hilliard J.A. (1977), The Application of Computer-Assisted Cartographic Methods to Road Mapping. Proceedings of the American Congress on Surveying and Mapping, Fall Meeting, p.271-281.

Hoffman J.E. (1980), Interaction between Global and Local Levels of a Form. Journal of Experimental Psychology, Human Perception and Performance, 6 (2), p.222-234.

Hopkin V.D. and Taylor R.M. (1979), Human Factors in the Design and Evaluation of Aviation Maps. AGARDograph no.225. NATO Advisory Group for Aerospace Research and Development, Neuilly-sur-Seine, France.

Huddleston J.H.F. (1982), Colour Coding for Computer-Driven Displays. Ferranti Computer Systems Report 4733.

Hunt R.W.G. (1979), Colour Vision. In: Clare J.N. and Sinclair M.A. (eds.), Search and the Human Observer, p.132-140. Taylor and Francis, London.

Igbozurike M.U. (1974), The Africa Road Map: Issues in Conceptualisation, Design, Standardisation and Rationalisation. Cartographic Journal, 11 (2), p.99-101.

Imhof E. (1972), Thematische Kartographie. Walter de Gruyter, Berlin.

Imhof E. (1982), Cartographic Relief Representation. Walter de Gruyter, Berlin.

Ishihara (1976), Tests for Colour-Blindness, '38 plates edition'. H.K.Lewis and Co., London.

Ittleson W.H. and Kilpatrick F.P. (1951), Experiments in Perception. Scientific American, 185 (2), p.50-55.

Janssen W.H. (1975), De Psychologie van de Automobilist-Kaartgebruiker. Kartografisch Tijdschrift, 1 (3), p.15-19.

Jeffery D.J. (1981a), The Potential Benefits of Route Guidance. Laboratory Report 997, Transport and Road Research Laboratory, Crowthorne, Berks.

Jeffery D.J. (1981b), Ways and Means for Improving Driver Route Guidance. Laboratory Report 1016, Transport and Road Research Laboratory, Crowthorne, Berks.

Jeffery D.J. (1985), Options for the Provision of Improved Driver Information Systems: The Role of Microelectronics and Information Technology. Paper presented to the Institute of Electrical Engineers Colloquium on Vehicle Route Guidance, London.

Johns E.H. and Sumner F.C. (1948), Relation of the Brightness Differences of Colours to their Apparent Distances. Journal of Psychology, 26, p.25-29.

Judd D.B. (1943), Facts of Color-Blindness. Journal of the Optical Society of America, 33 (6), p.294-307.

Judd D.B. and Yonemura G.T. (1969), Target Conspicuity and its Dependence on Color and Angular Subtense for Grey and Foliage Green Surrounds. Preliminary Report 10-120, National Bureau of Standards. U.S.Government Printing Office, Washington D.C.

Judd D.B. and Eastman A.A. (1971), Prediction of Target Visibility from the Colors of Target and Surround. Illumination Engineering, 66, p.256-266.

Kahneman D. and Henik A. (1981), Perceptual Organisation and Attention. In: Kubovy M. and Pomerantz J.R. (eds.), Perceptual Organisation, p.181-211. Lawrence Erlbaum Associates, Hillsdale, NJ.

Kaiser P.K. (1968), Color Names of Very Small Fields varying in Duration and Luminance. Journal of the Optical Society of America, 58, p.849-852.

Katz D. (1935), The World of Colour. Kegan, Paul, Trench and Trubner, London.

Kaufman L. (1975), The Uses of Color. Proceedings of the International Symposium on Computer-Assisted Cartography, Auto-Carto II, p.270-276.

Keates J.S. (1962), The Perception of Colour in Cartography. Proceedings of the Cartographic Symposium held in the Department of Geography, University of Glasgow, p.19-28.

Keates J.S. (1973), Cartographic Design and Production. Longman, London.

Keates J.S. (1982), Understanding Maps. Longman, London.

Keller H. and Cremer M. (1982), Performance and Effectiveness of Route Guidance Systems. Proceedings of the International Conference on Road Traffic Signalling, p.160-164. Institute of Electrical Engineers, London.

Kelly K.L. (1976), Twenty-two Colors of Maximum Contrast. Color Engineering, 3 (6), p.26-27.

Kendall M.G. (1955), Further Contributions to the Theory of Paired Comparisons. Biometrics, 11, p.43-62.

Kimerling A.J. (1980), Color Specification in Cartography. American Cartographer, 7 (2), p.139-153.

Kinchla R.A., Solis-Macias V. and Hoffman J. (1983), Attending to Different Levels of Structure in a Visual Image. Perception and Psychophysics, 33 (1), p.1-10.

Kinnear P.R. (1979), The Effects of Coloured Surrounds on Colour Naming and Luminosity. Vision Research, 19, p.1381-1387.

Kinneir J. (1984), The Practical and Graphic Problems of Road Sign Design. In: Easterby R. and Zwaga H. (eds.), Information Design, p.341-. John Wiley, Chichester.

Kinney J.A.S. (1962), Factors affecting Induced Color. Vision Research, 2, p.503-525.

Kinney J.A.S. (1979), The Use of Color in Wide Angle Displays. Proceedings of the Society for Information Display, 20 (1), p.33-40.

Kirby R.P. (1970), A Survey of Map User Practices and Requirements. Cartographic Journal, 7 (1), p.31-39.

Kirk R.E. (1968), Experimental Design: Procedures for the Behavioural Sciences. Brooks and Cole, Belmont,CA.

Kosslyn S.M. (1981), Species of Visual Displays. Journal of Mental Imagery, 5, p.24-26.

Künnapas T. (1958), Measurements of Subjective Length in the Vertical-Horizontal Illusion. Acta Psychologica, 14, p.371-374.

Land E.H. (1977), The Retinex Theory of Colour Vision. Scientific American, 237, p.108-128.

Langran G. (1984), Map Design for Computer Processing: Literature Review and DMA Product Critique. Naval Ocean Research and Development Activity Report 72, NSTL, Mississippi.

Lansdown J. (1985), Visual Perception and Computer Graphics. In: R.A.Earnshaw (ed.), Fundamental Algorithms for Computer Graphics, p.1005-1026. Springer-Verlag, Berlin.

Lloyd R. and Yehl S. (1979), Orderly Maps and Map Communication. American Cartographer, 6 (2), p.149-156.

LPTRG (London Passenger Transport Research Group) (1985), Getting There: An Assessment of LT's Endeavour to Improve Bus Passenger Information Literature for Central London, 1979-1985. Lodon Passenger Transport, 13, May, p.639-670.

Luckiesh M. (1944), Light, Vision and Seeing. D.Van Nostrand Co. Inc., New York.

Luder C.B. (1982), Redundant Colour Coding of Airborne CRT Displays. Ergonomics, 25, p.333 (abstract).

Lundqvist G. (1963), Generalisation of Communication Networks. Nachrichten aus dem Karten und Vermessungswesen, V (5), p.35-42.

Lunn S.E. (1978), Route Choice by Drivers. Supplementary Report 374, Transport and Road Research Laboratory, Crowthorne, Berks.

McGranaghan M. (1985), Symbol and Background Value Effects in Choropleth Maps for Color CRT Display. Technical Papers, American Congress on Surveying and Mapping, 45th Annual Meeting, p.356-363.

McGrath G. (1971), The Mapping of National Parks: A Methodological Approach. Cartographica Monograph, 2, p.71-76.

McGrath G. and Kirby R.P. (1970), Survey Methods on the Use of Maps and Atlases. Surveys and Mapping Branch reprint no.30 (from Canadian Cartographer, 6 (2), 1969, p.132-148).

McLean M.V. (1965), Brightness Contrast, Color Contrast and Legibility. Human Factors, 7 (6), p.521-526.

McManus I.C., Jones A.L. and Cotrell J. (1981), The Aesthetics of Colour. Perception, 10 (6), p.651-666.

MacAdam D.L. (1935), Maximum Visual Efficiency of Colored Materials. Journal of the Optical Society of America, 25, p.361-367.

Macioch A. (1983), The System of Cartographic Denotations: A Scientific Language for Cartography. In: D.R.F.Taylor (ed.), Graphic Communication and Design in Contemporary Cartography, p.219-229. John Wiley, Chichester.

Mair V. (1963), Strassenkarten aus Mairs Geographischen Verlag. International Yearbook of Cartography, III, p.163-170.

Makowski A. (1967), Aesthetic and Utilitarian Aspects of Colour in Cartography. International Yearbook of Cartography, VII, p.62-85.

Menu J-P. and Santucci G.F. (1982), Rôle de la Couleur dans la Symbologie de Pilotage des Aeronefs. NATO AGARD-CP-329, Neuilly-sur-Seine, France.

Metfessel M. (1947), A Proposal for Quantitative Reporting of Comparative Judgments. Journal of Psychology, 24, p.229-235.

Michelin (Les Services de Tourisme du Pneu Michelin) (1963), Histoire et Evolution des Publications Cartographiques Michelin. International Yearbook of Cartography, III, p.171-180.

Miller G.A. (1956), The Magic Number Seven Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review, 63, p.81-97.

Mollon J. (1985), Studies in Scarlet. The Listener, 10th January, p.6-7.

Morrison A. (1966), Principles of Road Classification for Road Maps. Cartographic Journal, 3 (1), p.17-30.

Morrison A. (1971), Experimental Maps of Road Travel Speed. Cartographic Journal, 8 (2), p.115-132.

Morrison A. (1974), Testing the Effectiveness of Road Speed Maps and Conventional Road Maps. Cartographic Journal, 11 (2), p.102-116.

Morrison A. (1975), Methods of Evaluating Motoring Maps. International Yearbook of Cartography, XV, p.110-122.

Morrison A. (1979a), The Frequency of Map Use for Planning Journeys by Road. Society of University Cartographers Bulletin, 13 (1), p.35-41.

Morrison A. (1979b), Finding your own Way. Geographical Magazine, 51, 9th June, p.661

Morrison A. (1980a), Existing and Improved Road Classifications for British Road Maps, Related to the Speed of Traffic. Cartographic Journal, 17 (2), p.111-122.

Morrison A. (1980b), Maps for Business and Pleasure. Geographical Magazine, 53, 1st October, p.76.

Morrison J.L. (1976), The Science of Cartography and its Essential Processes. International Yearbook of Cartography, XVI, p.84-97.

Morrison J.L. (1984), Applied Cartographic Communication: Map Symbolisation for Atlases. Cartographica, 21 (1), p.44-84.

Muller J-C. (1978), The Mapping of Travel Time in Edmonton, Alberta. Canadian Geographer, 22 (3), p.195-210.

Munsell Color (1966 and 1970) Munsell Book of Color: Glossy Finish Collection. Kollmorgen Corporation, Newburgh,NY.

Myers W.S. (1967), Accommodation Effects in Multi-Color Displays. Report AFFDL-TR-67-161, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio.

Neisser U. (1967), Cognitive Psychology. Appleton-Century-Crofts, New York.

Neisser U. (1976), Cognition and Reality: Principles and Applications of Cognitive Psychology. W.H.Freeman, San Francisco.

*Neter J., Wasserman W. and Kutner M.H. (1983), Applied Linear Regression Models. Richard D.Irwin Inc., Homewood,IL.

Newhall S.M. (1939), The Ratio Method in the Review of the Munsell Colours. American Journal of Psychology, 52, p.394-405.

Newhall S.M., Nickerson D. and Judd D.B. (1943), Final Report of the O.S.A. Subcommittee on the Spacing of the Munsell Colors. Journal of the Optical Society of America, 33 (7), p.385-418.

Nickerson D. and Newhall S.M. (1943), Theoretical Chroma Limits for Munsell Hues. Journal of the Optical Society of America, 33 (7), p.419-421.

Nicholson T.R. (1983), Wheels on the Road (Maps of Britain for the Cyclist and Motorist, 1870-1940). Geo Books, Norwich.

Niemelä (1979), The Road Map of Finland GT as a Multipurpose Map. International Yearbook of Cartography, XIX, p.90-95.

*Norcliffe G.B. (1977), Inferential Statistics for Geographers. Hutchinson and Co., London.

Noyes L. (1979), Are Some Maps Better than Others? Geography, 64, p.303-306.

Olson J.M. (1981), Spectrally Encoded Two-Variable Maps. Annals of the Association of American Geographers, 71 (2), p.259-276.

*OPCS (Office of Population Censuses and Surveys) (1980), Classification of Occupations, 1980. H.M.S.O., London.

OPCS (Office of Population Censuses and Surveys) (1982), General Household Survey. H.M.S.O., London.

*OPCS (Office of Population Censuses and Surveys) (1983), Census 1981 Great Britain- National Report, Parts 1 and 2. H.M.S.O., London.

Outram V.E. (1976), Route Choice. Department of the Environment Mathematical Advisory Unit note 256, March 1976.

Outram V.E. and Thompson E. (1978), Drivers Perceived Cost in Route Choice. Proceedings of the PTRC Summer Annual Meeting, University of Warwick.

Padgham C.A. (1971), The Direct Estimation of the Luminosity of Coloured Light Sources. Vision Research, 11, p.577-589.

Padgham C.A. and Saunders J.E. (1975), The Perception of Light and Colour. G.Bell and Sons, London.

Patton J.C. and Crawford P.V. (1977), The Perception of Hypsometric Colours. Cartographic Journal, 14 (2), p.115-127.

Payne M.C. (1958), Apparent Weight as a Function of Color. American Journal of Psychology, 71, p.725-730.

Payne M.C. (1961), Apparent Weight as a Function of Hue. American Journal of Psychology, 74, p.104-105.

Pearson E.S. (1930), A Further Development of Tests of Normality. Biometrika, 22, p.239-248.

Pearson K. (1971), The Relative Visual Importance of Selected Line Symbols. Unpublished Master's thesis, Department of Geography, University of Wisconsin, Madison,WI.

Petchenik B.Bartz (1983), A Map Maker's Perspective on Map Design Research 1950-1980. In: D.R.F.Taylor (ed.), Graphic Communication and Design in Contemporary Cartography, p.37-68. John Wiley, Chichester.

Peterson M.P. (1979), An Evaluation of Unclassed Crossed-Line Choropleth Mapping. American Cartographer, 6, p.21-37.

Peucker K. (1898), Schattenplastik und Farbenplastik. Beiträge zur Geschichte und Theorie der Geländedarstellung. Artaria, Vienna.

Phillips R.J. (1981), Estimating the Area of Peripheral Vision employed for Map Search. Cartography, 12 (2), p.104-107.

Phillips R.J. (1982), An Experimental Investigation of Layer Tints for Relief Maps in School Atlases. Ergonomics, 25 (12), p.1143-1154.

Phillips R.J. (1984), Experimental Method in Cartographic Communication: Research on Relief Maps. Cartographica, 21 (1), p.120-128.

Phillips R.J. and Noyes L. (1980), A Comparison of Colour and Visual Texture as Codes for Use as Area Symbols on Thematic Maps. Ergonomics, 23 (12), p.1117-1128.

Pinkerton E. and Humphrey N.K. (1974), The Apparent Heaviness of Colours. Nature, 250, p.164-165.

Potash L.M. (1977), Design of Maps and Map-Related Research. Human Factors, 19 (2), p.139-150.

Poulton E.C. (1975), A Recommended Ergonomic Colour Code. Applied Ergonomics, December, p.231-235.

Preston P.R. (1980), Six Road Maps of Tasmania: An Assessment. Cartography, 11 (3), p.153-.

Rado S. and Dudar T. (1971), Some Problems of Standardisation of Transportation Map Symbols in Thematical Mapping. International Yearbook of Cartography, XI, p.160-164.

Raisz E. (1962), Principles of Cartography. McGraw-Hill, New York.

Reay A. and Hobbs G. (1979), The Judo Manual. Stanley Paul, London.

Rhind D.W., Shaw M.A. and Howarth R.J. (1973), Experimental Geochemical Maps- A Case Study in Cartographic Techniques for Scientific Research. Cartographic Journal, 10 (2), p.112-118.

Richardson L.F. (1929), Quantitative Mental Estimates of Light and Colour. British Journal of Psychology, 20, p.27-37.

Richardson L.F. and Ross J.S. (1930), Loudness and Telephone Current. Journal of General Psychology, 3, p.288-306.

Ristow W.W. (1964), A Half-Century of Oil Company Road Maps. Surveying and Mapping, 24 (4), p.617-637.

Robb M.C. (1985), An Automatic Route-Planning and Mapping System. Unpublished PhD thesis, University of Glasgow.

Robertson P.K. and O'Callaghan J.F. (1986), The Generation of Color Sequences for Univariate and Bivariate Mapping. IEEE Computer Graphics and Applications, February, p.24-32.

Robinson A.H. (1952), The Look of Maps- An Examination of Cartographic Design. University of Wisconsin Press, Madison,WI.

Robinson A.H. (1967), Psychological Aspects of Colour in Cartography. International Yearbook of Cartography, VII, p.50-59.

Robinson A.H. (1977), Research in Cartographic Design. American Cartographer, 4 (2), p.163-169.

Robinson A.H. (1982), A Program of Research to Aid Cartographic Design. American Cartographer, 9 (1), p.25-29.

Robinson A.H. and Sale R.D. (1969), Elements of Cartography, 3rd edition. John Wiley, New York.

Robinson A.H., Sale R.D. and Morrison J.L. (1978), Elements of Cartography, 4th edition. John Wiley, New York.

Robinson A.H., Sale R.D., Morrison J.L. and Muehrcke P.C. (1984), Elements of Cartography, 5th edition. John Wiley, New York.

Royce I. (1986), CARIN: An Application of CD-ROM. Optical Information Systems, 6 (1), p.60-63.

Russon K. (1986), Interview at the Transport and Road Research Laboratory on BBC Radio 4 'Science Now: In Passing.' First broadcast 8th September 1986.

*Ryan T.A.Jr., Joiner B.L. and Ryan B.F. (1981), Minitab Reference Manual. Pennsylvania State University, University Park, PA.

Salichtchev K.A. (1983), Cartographic Communication: A Theoretical Survey. In: D.R.F.Taylor (ed.), Graphic Communication and Design in Contemporary Cartography, p.11-35. John Wiley, Chichester.

Saunders B.G.R. (1958), Design and Emphasis in Special Purpose Maps. Cartography, 2, p.108-113.

Saunders B.G.R. (1961), Map Design and Colour in Special Purpose Maps. Cartography, 4, p.5-9.

Schiede H. (1968), Kritische Gedanken zur Kartographie der Strassenkarte. In: K-H. Meine (ed.), Kartengeschichte und Kartenbearbeitung, p.163-169. Kirschbaum Verlag, Bad Godesberg.

Schultz G.M. (1963), New Developments in American Road Maps. Professional Geographer, XV (3), p.13-15.

Severud K. (1968), A Study of the Relative Visual Importance of Selected Line Symbols. Paper presented at the University of Wisconsin Cartography Seminar 962.

Shellswell M. (1976), Towards Objectivity in the Use of Colour. Cartographic Journal, 13 (1), p.72-84.

*Sheppard D. (1971), Characteristics of Drivers obtained from Large Scale Enquiries. Laboratory Report 389, Transport and Road Research Laboratory.

Sheppard D. and Adams J.M. (1971), A Survey of Drivers' Opinions on Maps for Route Finding. Cartographic Journal, 8 (2), p.105-114.

Sheppard J.J.Jr. (1968), Human Color Perception: A Critical Study of the Experimental Foundation. American Elsevier Publishing Co., New York.

Shortridge B.G. (1979), Map Reader Discrimination of Lettering Size. American Cartographer, 6 (1), p.13-20.

Shortridge B.G. and Welch R.B. (1980), Are we Asking the Right Questions? Comments on Instructions on Cartographic Psychophysical Studies. American Cartographer, 7 (1), p.19-23.

Silverstein L.D. (1982), Human Factors for Color CRT Displays. Society for Information Display, Seminar notes, Volume II, p.7-1 to 7-41.

Skelton R.A. (1964), The Early Map Printer and his Problems. The Penrose Annual, 27, p.171-186.

Smith S.L. (1962), Color Coding and Visual Search. Journal of Experimental Psychology, 64 (5), p.434-440.

Smith S.L. (1963), Color Coding and Visual Separability in Information Displays. Journal of Applied Psychology, 47 (6), p.358-364.

Stams W. (1964), Die Strassenatlanten der Sozialistischen Länder. Vermessungstechnik, 12 (9), p.337-342.

Stevens S.S. (1956), The Direct Estimation of Sensory Magnitudes- Loudness. American Journal of Psychology, 69 (1), p.1-25.

Stevens S.S. (1957), On the Psychophysical Law. Psychological Review, 64 (3), p.153-180.

Stevens S.S. and Galanter E.H. (1957), Ratio Scales and Category Scales for a Dozen Perceptual Continua. Journal of Experimental Psychology, 54 (6), p.377-411.

Stevens S.S. and Guirao M. (1963), Subjective Scaling of Length and Area and the Matching of Length to Loudness and Brightness. Journal of Experimental Psychology, 66, p.177-186.

Streeter L.A. and Vitello D. (1986), Profile of Drivers' Map Reading Abilities. Human Factors, 28 (2), p.223-239.

Stringer P. (1973), Colour and Base in Urban Planning Maps. Cartographic Journal, 10 (2), p.89-94.

Sumner F.C. (1932), Influence of Color on Legibility of Copy. Journal of Applied Psychology, 16, p.201-204.

Swezey R.W. (1984), Optimising Legibility for Recall and Retention. In: Easterby R. and Zwaga H. (eds.), Information Design, p.145-155.

Tansley B.W. and Boynton R.M. (1976), A Line, Not a Space, Represents Visual Distinctiveness of Borders Formed by Different Colors. Science, 191 (4230) March, p.954-957.

Taylor I.L. and Sumner F.C. (1945), Actual Brightness and Distance of Individual Colours when their Apparent Distance is held Constant. Journal of Psychology, 19, p.79-85.

Taylor N.B. and Jeffery D.J. (1981), Determining Optimum Routes from Maps. Laboratory Report 1001, Transport and Road Research Laboratory, Crowthorne, Berks.

Taylor R.M. (1974), Subjective Methods in Map Evaluation. Paper presented to the British Cartographic Society Annual Technical Symposium.

Taylor R.M. (1977), Empirical Derivation of Map Colour Specifications. Paper presented to U.K. Working Group on Cartographic Communication, Royal Society, London.

Taylor R.M. (1981), UV Light Activated Fluorescent Maps for Helicopter PNG Operations. Unpublished RAF Institute of Aviation Medicine memorandum.

Taylor R.M. (1984), Colour Coding in Information Displays: Heuristics, Experience and Evidence from Cartography. Paper presented to NATO WG5 Workshop on Colour v Monochrome Electronic Displays, Farnborough.

Taylor R.M. (1985), Colour Design in Aviation Cartography. Displays, October, p.187-201.

Taylor R.M. (1986) Perceptual Organisation and Information Management. Proceedings of NATO AGARD Aerospace Medical Panel Symposium, Toronto. NATO, Neuilly-sur-Seine, France.

Taylor R.M. and Belyavin A.J. (1980), Colour Differences and Map Design for Moving Map Displays. RAF Institute of Aviation Medicine Report 587.

Taylor R.M. and Farmer E.W. (1980), Visual Search through Colour Displays: The Effects of Colour Differences. Annual Scientific Meeting Preprint, Aerospace Medical Association, Washington D.C., p.153-154.

Teghtsoonian M. (1965), The Judgment of Size. American Journal of Psychology, 78, p.392-402.

Thake J.E. (1979), Maps for the Road User. Society of University Cartographers, Proceedings of the Oxford Summer School, p.43-.

Thorndyke P.W. and Stasz C. (1980), Individual Differences in Procedures for Knowledge Acquisition from Maps. Cognitive Psychology, 12, p.137-175.

Tobler W.R. (1957), An Empirical Investigation of some Aspects of Hypsometric Colours. Unpublished MA thesis, Geography Department, University of Washington.

Treisman, A. (1982), Perceptual Grouping and Attention in Visual Search for Features and for Objects. Journal of Experimental Psychology, Human Perception and Performance, 8 (2), p.194-214.

Trumbo B.E.A. (1981), Theory for Coloring Bivariate Statistical Maps. American Statistician, 35 (4), p.220-226.

Van der Weiden F.L.T. and Ormeling F.J. (1972), An Inquiry into the Associative Properties of Colour as Applied to Cartography. Geografisch Tijdschrift, VI (3), p.282-289.

Van Meeteren A. (1974), Beoordeling van Enkele Wegenkaarten van Belgie en Frankrijk. Report 12FC 1974-3, Institute for Perception TNO, Soesterberg, Netherlands.

Vickerstaff T. and Woolvin C.S. (1944), Investigation of the Legibility and Aesthetic Value of Coloured Printing Inks on Coloured Papers. Communication (Journal, Oil and Colour Chemists' Association), XXVII, 293 (November), p.209-219.

Voisin R.L. (1976), Maps for the American People. American Cartographer, 3 (2), p.101-106.

Wallis W.A. (1935), The Influence of Color upon Apparent Size. Journal of General Psychology, 13, p.193-199.

Walraven J. (1984), Perceptual Artifacts that may Interfere with Colour Coding on Visual Displays. In: Colour Coded v Monochrome Electronic Displays, Proceedings of a NATO Workshop, Farnborough.

Ward A.W. (1977), The State of State Road Maps. American Cartographer, 4 (1), p.5-9.

Warden C.J. and Flynn E.L. (1926), The Effect of Color on Apparent Size and Weight. American Journal of Psychology, 37, p.398-401.

Wasserman, G.S. (1979), The Physiology of Color Vision. Color Research and Application, 4 (2), p.57-65.

Watts G.R. (1984), Evaluation of Conspicuity Aids for Pedal Cyclists. Laboratory Report 1103, Transport and Road Research Laboratory, Crowthorne, Berks.

Weale R.A. (1951), The Foveal and Para-Central Spectral Sensitivities in Man. Journal of Physiology, 114, p.435-446.

Weale R.A. (1953), Spectral Sensitivity and Wavelength Discrimination of the Peripheral Retina. Journal of Physiology, 119, p.170-190.

Wetherell A. (1984), Some Factors affecting Spatial Memory for Route Information. In: Easterby R. and Zwaga H. (eds.), Information Design, p.321-338. John Wiley, Chichester.

Williams J.E. (1953), The Relief Map. Professional Geographer, September, p.7-10.

Williams L.G. (1966), Target Conspicuity and Visual Search. Human Factors, 8 (2), p.80-92.

Williams L.G. (1967a), The Effects of Target Specification on Objects Fixated during Visual Search. Acta Psychologica, 27, p.355-360.

Williams L.G. (1967b), A Study of Visual Search using Eye Movement Recordings. Honeywell document 12009-IR. Honeywell Systems and Research Center, Minneapolis.

Williams L.G. (1971), Obtaining Information from Displays with Discrete Elements. Cartographica Monograph 2, p.29-34.

Williams R.L. (1956), Statistical Symbols for Maps: Their Design and Relative Values. Report of Contract NR 088-006, Nonr 609(03) of the Office of Naval Research.

Witzel R.F., Burnham R.W. and Onley J.W. (1973), Threshold and Superthreshold Perceptual Colour Differences. Journal of the Optical Society of America, 63, p.615-625.

Wong K.W. and Yacoumelos N.G. (1973), Identification of Cartographic Symbols from TV Displays. Human Factors, 15 (1), p.21-31.

Wood C.H. (1976), Brightness Gradients Operant in the Cartographic Context of Figure-Ground Relationship. Proceedings of the American Congress on Surveying and Mapping, 36th Annual Meeting, p.5-34.

Wood M. (1968), Visual Perception and Map Design. Cartographic Journal, 5 (1), p.54-64.

Wood M. (1972), Human Factors in Cartographic Communication. Cartographic Journal, 9 (2), p.123-132.

Woodworth R.S. and Schlosberg H. (1955), Experimental Psychology, 3rd edition. Methuen, London.

Wootton H.J., Ness M.P. and Burton R.S. (1981), Improved Direction Signs and the Benefits for Road Users. Traffic Engineering and Control, May, p.264-268.

Wright B. and Rainwater L. (1962), The Meanings of Color. Journal of General Psychology, 67, p.89-99.

Wright J.K. (1942), Map Makers are Human: Comments on the Subjective in Maps. Geographical Review 32 (4), p.527-544.

Wright R.D. (1967), Selection of Line Weights for Solid Qualitative Line Symbols in Series on Maps. Unpublished PhD thesis, University of Kansas. Dissertation Abstracts International, 28 (8), p.3333-B.

Wrigley N. (1985), Categorical Data Analysis for Geographers and Environmental Scientists. Longman, London.

Wysecki G.W. and Stiles W.S. (1967), Color Science, Concepts and Methods, Quantitative Data and Formulae. John Wiley, New York.

Yarbus A.L. (1967), Eye Movements and Vision. Plenum Press, New York.

Zeki S.M. (1977), Colour Coding in the Superior Temporal Sulcus of Rhesus Monkey Visual Cortex. Proceedings of the Royal Society of London B, 197, p.195-223.

MAP SOURCES

Automobile Association (1982), Touring Map of Great Britain, 1:633,600.

Automobile Association (1984), Members' Handbook 1984/85, 1:633,600.

Collins (1985), Road Atlas Britain, 1:200,000.

Geographers' A-Z Map Co. (c.1983), Road Atlas Great Britain, 1:200,000.

Geographia (c.1980), Dunlop Width of Road Map, no.5, Southern Scotland, 1:316,800.

Geographia (1984/5), Travellers Britain: The Data File for People on the Move, 1:250,000.

George Philip (1980), Esso Map no.2, South East England, 1:316,800.

George Philip (1983), The Shell Road Atlas of Great Britain, 1:200,000.

Mair (c.1981), Baedeker's AA Road Map of Switzerland, 1:505,000.

Map Productions (c.1982), England and Wales, 1:633,600.

Map Productions (c.1982), RAC South Wales, 1:190,080.

Map Productions (1983), RAC Navigator 1, 2nd edition, 1:100,000.

Michelin (1984), Autoroutes, 10th edition (map 400), various scales.

Michelin (1985), France Route Planning, 1st edition (map 911), 1:1,000,000.

Morrison, A. (1983), Road Speed Map of South East England, 1982-3, 1:625,000. (Transport and Road Research Laboratory contract TRR 842/385)

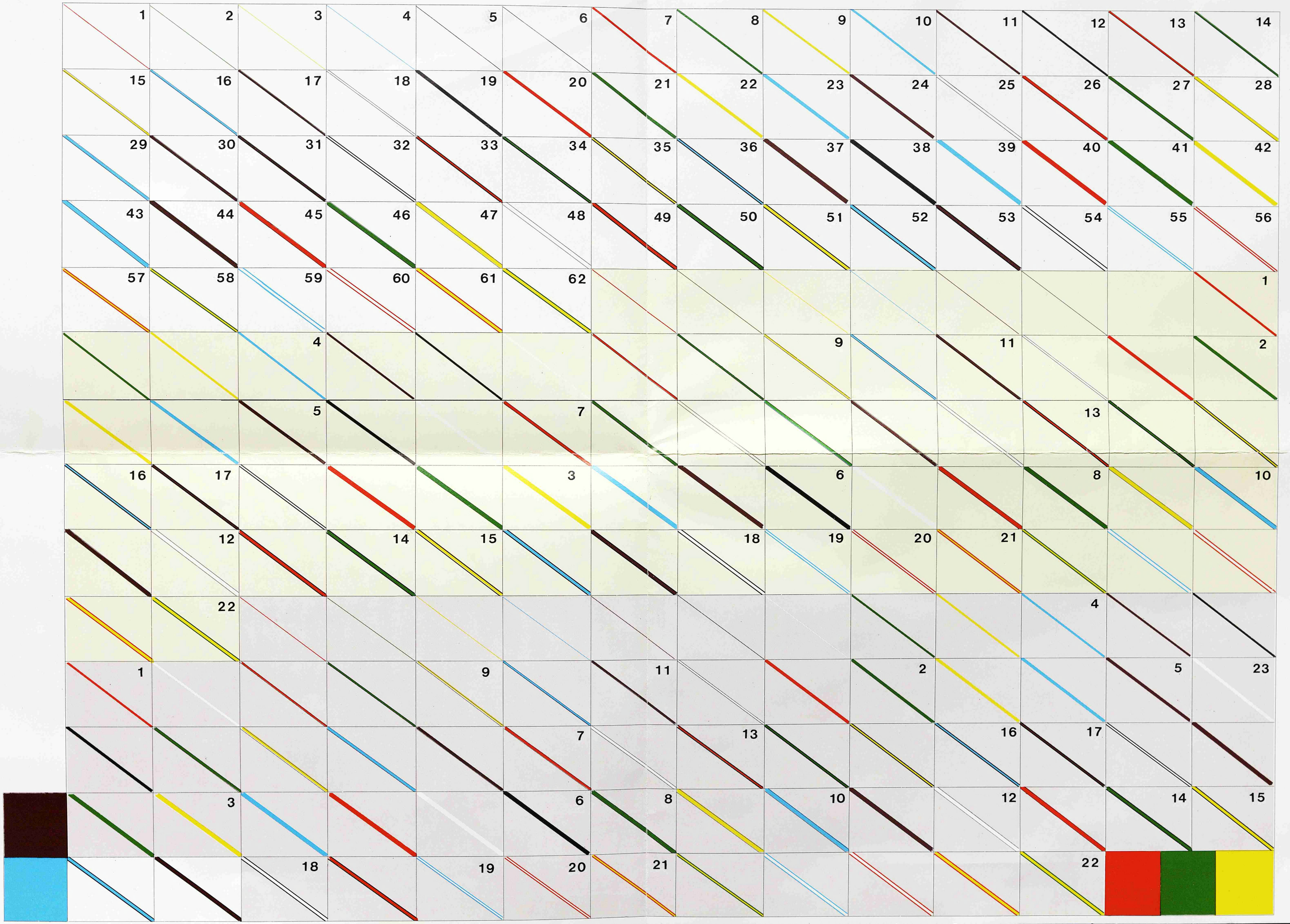
Ordnance Survey (1983), The Motorist's Routeplanner 1984 of Great Britain, 23rd edition, 1:625,000.

Ordnance Survey (1983), Motoring Atlas of Great Britain, 1st edition, 1:190,080. (Ordnance Survey/ Temple Press)

Ordnance Survey. Landranger Series of Great Britain, 1:50,000, sheets 119, 120, 128, 129, 135, 145, 146, 159, 187, 188, 197, 198.

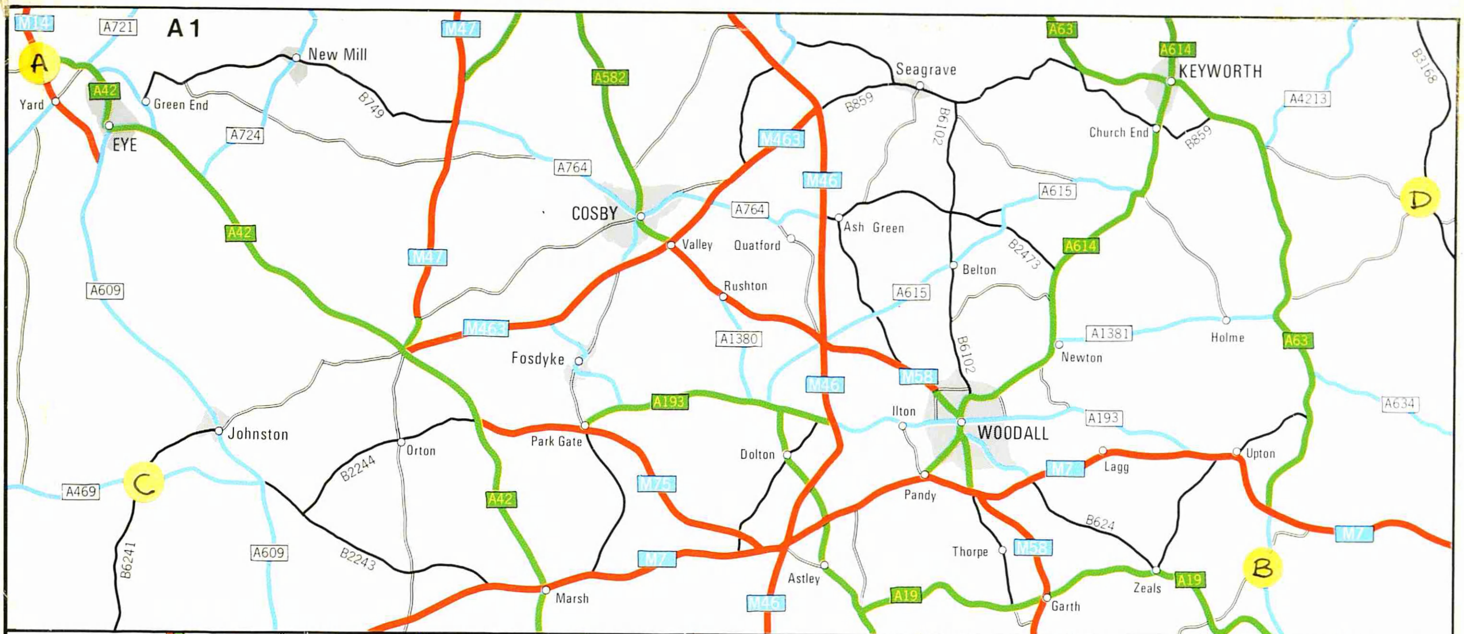
Rand McNally (c.1963), Standard Oil Map of Toll Roads (Illinois to New England).

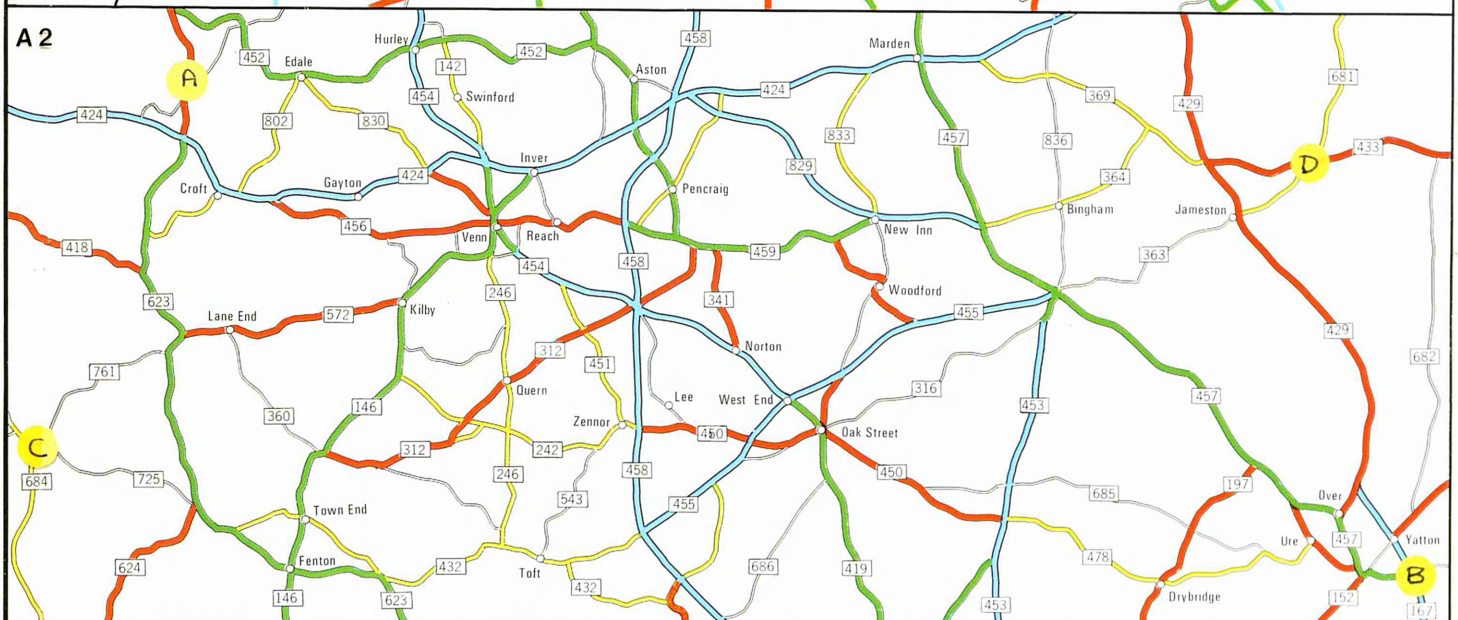
Reader's Digest/AA (1986), New Book of the Road 1986/7, 1:250,000

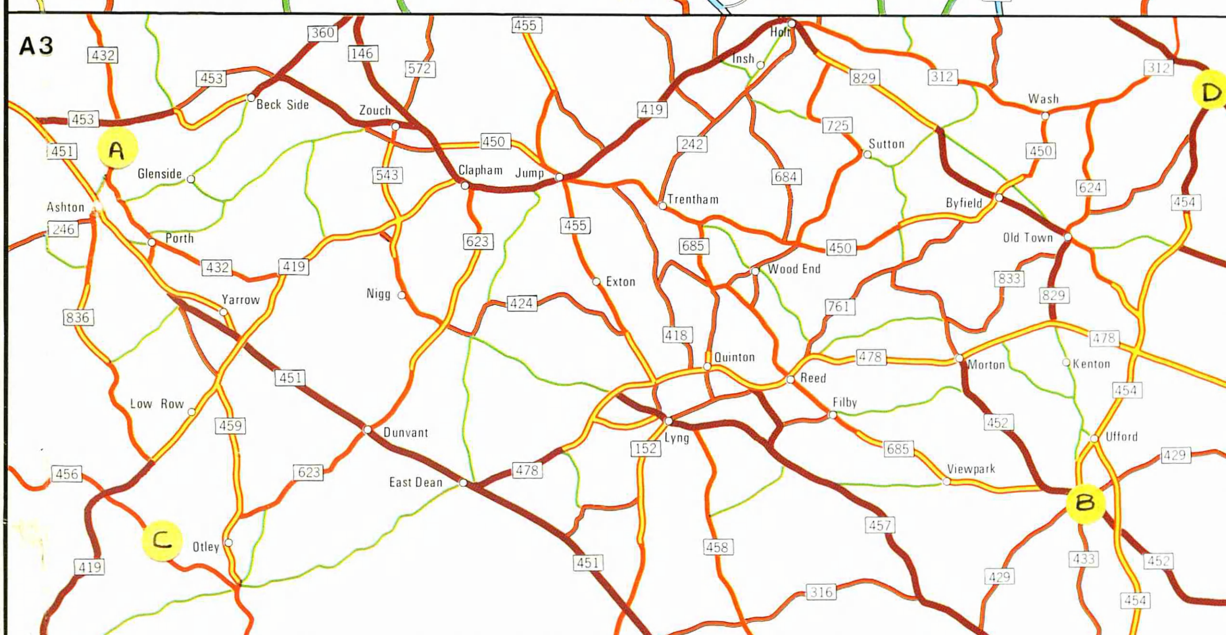

1 2 3 4 5 6 7 8 9 10 11 12 13 14
15 16 17 18 19 20 21 22 23 24 25 26 27 28
29 30 31 32 33 34 35 36 37 38 39 40 41 42
43 44 45 46 47 48 49 50 51 52 53 54 55 56
57 58 59 60 61 62 1
4 9 11 2
5 7 13
16 17 3 6 8 10
12 14 15 18 19 20 21
22 4
1 9 11 2 5 23
7 13 16 17
3 6 8 10 12 14 15
18 19 20 21 22

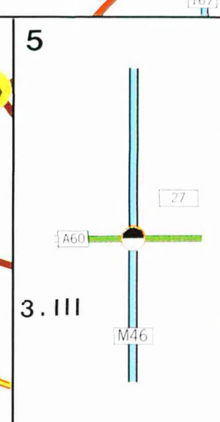
STIMULI, EXPMTS. 1 and 2

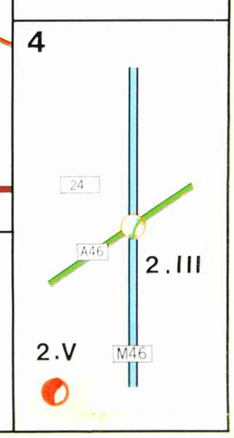
SET A

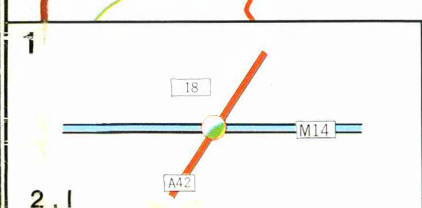

A1
A
B
C
D
M14
A721
A42
Yard
Green End
EYE
New Mill
B749
A724
M47
A582
A764
COSBY
Valley
Quatford
Rushton
A609
Fosdyke
A1380
Johnston
Orton
B2244
Park Gate
A469
B6241
A609
B2243
Marsh
M463
Seagrave
B859
B6102
M46
Ash Green
A615
Belton
B2473
B6102
A193
M58
Ilton
WOODALL
Dolton
M75
Pandy
M7
Astley
Thorpe
A19
Garth
Zeals
B624
Lagg
Upton
A63
A614
KEYWORTH
Church End
B859
A4213
B3168
A1381
Newton
Holme
A634

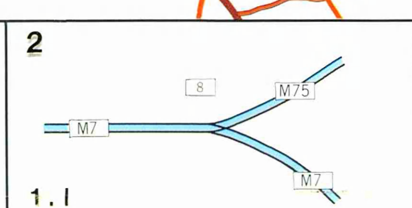

A2
A
B
C
D
Hurley
Edale
Swinford
Aston
Marden
Inver
Gayton
Croft
Pencraig
Venn
Reach
New Inn
Bingham
Jameston
Woodford
Lane End
Kilby
Norton
Quern
Lee
West End
Zennor
Oak Street
Town End
Fenton
Toft
Drybridge
Over
Ure
Yatton

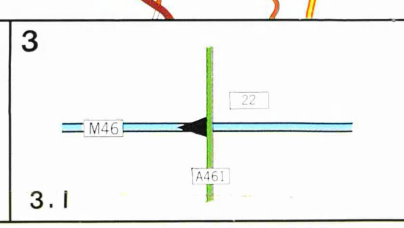

A3
A
B
C
D
Beck Side
Zouch
Holt
Insh
Glenside
Ashton
Porth
Clapham
Jump
Sutton
Wash
Trentham
Byfield
Old Town
Wood End
Exton
Nigg
Yarrow
Quinton
Reed
Morton
Kenton
Low Row
Filby
Lyng
Dunvant
Ufford
East Dean
Viewpark
Otley


5
27
A60
3.III
M46

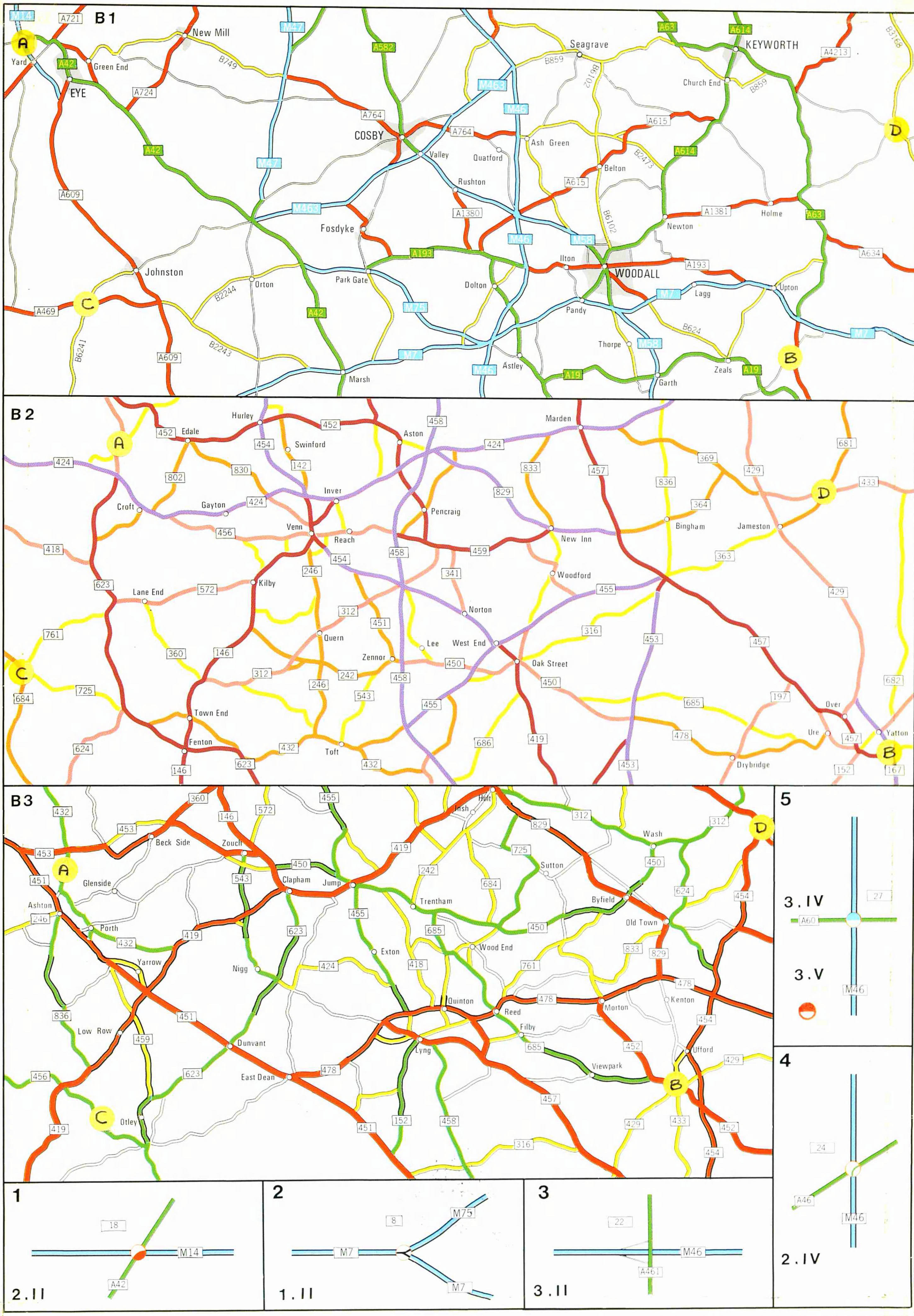

4
24
A46
2.III
2.V
M46


1
18
M14
A42
2.I


2
8
M75
M7
M7
1.I


3
22
M46
A461
3.I

SET B



B 1
M14
A721
A
Yard
A42
Green End
EYE
New Mill
B749
A724
M47
A582
A764
COSBY
A764
Valley
Quatford
M463
M46
Seagrave
B859
B6102
A63
A614
KEYWORTH
A4213
B3168
Church End
B859
A615
Ash Green
B2473
Belton
A614
D
A42
A609
M47
M463
Fosdyke
Rushton
A1380
A615
B6102
A1381
Holme
A63
Newton
M46
M58
A193
Ilton
WOODALL
A193
A634
Johnston
Orton
B2244
Park Gate
Dolton
M7
Lagg
Upton
A469
C
A42
M75
Pandy
B624
B6241
A609
B2243
M7
Thorpe
M58
M7
B
Marsh
M46
Astley
A19
Garth
Zeals
A19
B 2
Hurley
452
452
Edale
A
454
Swinford
142
Aston
458
424
Marden
681
424
830
802
833
457
369
429
433
836
Inver
829
Pencraig
364
D
Croft
Gayton
424
Venn
Reach
456
Bingham
Jameston
New Inn
418
454
458
459
363
246
Woodford
623
Lane End
572
Kilby
341
455
429
Norton
312
451
761
Quern
316
Lee
West End
453
457
360
146
Zennor
450
Oak Street
C
312
242
725
246
458
450
682
684
543
455
685
197
Over
Town End
Ure
457
Yatton
686
419
478
Fenton
432
Toft
Drybridge
B
624
146
623
432
453
152
167
B 3
360
455
432
146
572
Holt
453
Beck Side
Zouch
Insh
829
312
312
D
453
A
450
419
725
Wash
451
Glenside
543
Clapham
Jump
242
684
Sutton
450
Ashton
Trentham
Byfield
624
454
246
455
Old Town
Porth
623
685
450
432
419
Exton
Wood End
833
829
Yarrow
Nigg
424
418
761
478
Kenton
836
Quinton
Reed
478
Morton
454
Low Row
451
Filby
459
Lyng
Dunvant
685
452
Ufford
Viewpark
429
456
623
East Dean
478
B
457
C
Otley
152
458
429
433
452
419
451
316
454
5
3.IV
27
A60
3.V
M46
4
24
A46
M46
2.IV
1
18
M14
A42
2.II
2
8
M75
M7
M7
1.II
3
22
M46
A461
3.II