Cognitive aspects of work with digital maps

by

Clare Davies

A doctoral thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University

July 1999*

Supervisor: Prof. K. Eason
Director of Research: Dr Susan Harker

Department of Human Sciences
Loughborough University
Loughborough, Leicestershire LE11 3TU, England

---

* Viva conducted January 2000; corrections list received April 2000; revised version delivered June 2000.

Abstract

Digital maps of geographic areas are increasingly common in many types of workplace, in education and in the public domain. Their interactivity and visual features, and the complexity of geographic(al) information systems (GIS) which create, edit and manipulate them, create special cognitive demands on the end-user which are not present in traditional cartographic maps or in most human-computer interaction (HCI). This thesis reviews cross-disciplinary literature regarding cognitive aspects of viewing and interacting with digital maps.

Data from an observational study of GIS use, including real-time recordings of normal workplace activities, was analysed using various approaches to examine the interactive and visual aspects of people's work. The implications for cartographic, psychological and HCI aspects of GIS are discussed, in the context of the actual tasks people perform with them (rather than the computationally advanced analyses assumed by most literature).

The second phase of the research examined the spatial knowledge attained and used during this interaction. The relevance of specific concepts in cognitive psychology, and of factors that create individual differences in cognition, are discussed in depth, alongside work in environmental and educational psychology, cartography and geography.

A controlled experiment examined the degree to which task characteristics induce a different spatial model or reference frame when viewing a digital map. It was shown that even novice users can switch between considering the map as an abstract geometric display or as a geographical representation, without affecting performance. However, tasks forcing subjects to focus entirely on the geometry rather than the geography did affect performance in a surprise post-test photograph identification task. Map users' mental model or reference frame is apparently affected by these task constraints; this has implications for GIS design and practice as well as for understanding spatial cognition. The study also considered the role of expertise and other individual difference factors, although conclusions were limited by sample size. Further research issues are highlighted, particularly regarding the knowledge structures and spatial language used in interpreting digital maps.

Keywords: spatial cognition, cartography, human-computer interaction (HCI), cognitive ergonomics, geographic information systems (GIS), individual differences, maps, task analysis, expertise

Acknowledgements

During this part-time doctoral research I've had four different employers and two periods of unemployment, and lived in six different houses in four counties. This thesis is dedicated to anyone who never lost faith that I'd finish it, despite everything.

Calvin Davies, my ever-patient husband, effectively part-financed this work, and gave endless encouragement and sympathy. It would have been impossible without him, like so much in my life. Family and friends encouraged, nagged or participated, especially Kate Carter, Wei Sun, Val Byrne, John Kirriemuir, Simon Richardson and Alison Scammell. I must also thank my parents Mike and Hazel State, for similar help and for first teaching me the value of perseverance! My late grandfather, Laurence State, was unfailingly proud of his 'erudite granddaughter', and I wish he could have stayed around a few days more for my long-overdue viva, when I finally 'found the answer'.

Thanks to staff at four universities: Ken Eason, my supervisor, for his patience and faith. David Medyckyj-Scott, for his inspiring USIS leadership, lots of early help, long-term book loans, and a helpful degree of challenge after declaring disinterest in 'wishy-washy' cognitive research (hope this is unwishy-washy enough for you, David!). Thom Baguley, for very helpful comments and references. Andy Payne at DMU Bedford, for helping me grapple with a GIS to produce the experiment maps. Mark Lansdale, for early encouragement, advice and a desk, and for proving the potential of controlled cognitive research in realistic contexts. Other current/former staff at Loughborough, Leicester, De Montfort or Surrey: Mary Hewitt, Peter Fisher, Val Byrne, David Walker, Erica Milwain, Peter Robinson, Ian Newman, Charles Oppenheim, Marilyn Deegan, Sophia Goddard, John Eyre, Lisa Dorn, Howell Istance, David Uzzell and Andrew Barnes. Ian Davies and colleagues at the Department of Psychology, University of Surrey, gambled on an unconventional interdisciplinarian, and graciously allowed me time to complete this thesis. Sorry to anyone I've missed out - it's been a long time.

Fellow travellers, en route to Ph.D. or Master's degrees, empathised, discussed, flattered, or pointed me to things I'd otherwise have missed: Gary Burnett, Colin Monckton, Martin Hamilton, Jennifer Sanborn, John Lee, Pauline Smith, Paul Jackson. Also colleagues overseas whose early help enabled me establish my own Ph.D. path: especially Andrew Turk (Murdoch University, Australia), Tim Nyerges (University of Washington, Seattle), Mike Gould (Universidad Complutense de Madrid, Spain), and Reg Golledge (UCSB). I'm especially grateful to Christophe Pallier (EHESS, France) who co-wrote EXPE, mounted it on the Internet for free, offered me a pre-release of a new version that did what I needed, and gave me prompt help by email. David Tremmel in North Carolina wrote the BiblioMan II HyperCard stack (my excellent reference database until the final stages of the thesis) and helped me customise it. They all create a real international research community.

Jim Storrar and Patrick Wills at Surrey CC organised trials for my final experiment, and all the participants there and elsewhere gave enthusiastic help. The freely given help of the earlier USIS participants provided many insights. Residents of Saltwood

and Newington showed typical British tolerance, especially the old lady who suggested locations for my photographs.

This doctoral work has been performed with no official funding. However, the Economic and Social Research Council funded the original USIS project, whose data was reanalysed during the early phase of the present work. The HUSAT Research Institute, where my research career began, paid my Ph.D. fees from 1993-7, allowed me continuing access to its building and library, and gave me many friends and ideas.

Finally, there's my goddaughter Alexandra Carter who, by throwing up on me while her mother did my experiment, gave me an ice-breaking anecdote to relax future participants, and showed early promise as either a great psychologist or a well-timing comedian.

A note on collaboration

The analyses and ideas presented in this thesis are the author's independent work. However, the data reanalysed in Chapters 2 and 3 had been previously collected by the collaborative teamwork of the USIS Project, led by Dr David Medyckyj-Scott (although with the present author as the principal researcher and main designer of the observation study), as described in those chapters. The cartographic analysis in the second half of Chapter 2 was also a collaborative effort (led by the author), as will be explained therein. The focus on reference frames within the subsequent literature review was originally suggested by Dr Thom Baguley, and a few cited papers were provided or suggested by colleagues at Loughborough or De Montfort universities, although the direction the thesis took was solely the author's decision. The final experimental study was designed, run and analysed by the author alone.

CONTENTS

Figures and tables in main text

Glossary and abbreviations list

analysis	in the GIS context, general term for commands or macros which can perform sophisticated calculations such as determining where water will flow from a burst pipe, predicting traffic flows, calculating spatial correlations between two phenomena to see how closely related they are, etc.
ARW	Advanced Research Workshop: one of a series of specialist invitation-only conferences funded by NATO.
attributes	the textual or numerical data associated with specific features of digital maps. E.g. a customer's or citizen's house may be associated with a record of the people living there, their recent bill or council tax payments, etc.
cartogram	a cartographic map representation in which the relative sizes (areas) of regions are distorted to reflect values of a variable such as employment, etc. Tend to be used to convey broad messages about inequalities etc., and are only effective where the viewer will already be very familiar with a geographically-proportioned image of the map, e.g. where it is their own country or the world as a whole.
data conversion	a major issue in GIS developments, since a number of different formats became simultaneously common for map data. Still a major problem in many situations, especially when using data not created by major map agencies such as the Ordnance Survey.
deictic	with reference to spatial relations between objects, a deictic relation depends on the observer's understanding of the overall array or image: e.g. something may be 'near the top' of the screen or picture, or 'to the left of' another object.
dialog box	a small window which appears on a computer screen when the user has selected a command from a menu but the system requires more specific information about what the user wishes to do (e.g. which file to open).
digital map	a cartographic map of a real geographical area, displayed on a computer screen and stored as digital data that can be edited and whose on-screen appearance can be altered.
digitising	the process of recording new map features within a digital map: usually with a special input device (often including more buttons than a mouse, to specify aspects of the features being drawn, and may be moved over a touch-sensitive tablet divided into sections for different commands). Sometimes digitising is more automatic, e.g. by scanning a paper map and then correcting the result on screen.

ESRC	Economic and Social Research Council (a UK Government research funding body)
GIS	geographic(al) information system: powerful computer software which can display, edit, manipulate and analyse digital map data, often integrating it with other data such as customer databases or socio-economic census information.
HCI	human-computer interaction: the subdiscipline of computer science/psychology which considers the ways in which people interact with computers, and generally aims to improve computer software and hardware design to increase usability.
help time	one of the usability 'performance metrics' developed by the MUSiC project: time a computer user spends gaining assistance, either from online, human or paper sources, instead of performing the task at hand.
intrinsic	with reference to spatial relations between objects, an intrinsic relation is one which depends on some inherent properties of the objects themselves (e.g. something can't be 'behind' something else unless we assume the latter has a recognisable front and back).
ISO	International Standards Organisation, based in Geneva, and responsible for publishing the standard IS9241 on "Ergonomic requirements for office work with visual display terminals" in the late 1990's, after several years of work by an international 'Technical Committee' (by which time, arguably, some of it was out of date for modern user interfaces).
measurement	as an activity with a digital map, getting the system to calculate e.g. the area or the boundary lengths of a property (or other polygon shape).
MRRL	Midlands Regional Research Laboratory: one of a series of Regional Research Laboratories funded by the ESRC in the early 1990's to facilitate the exploitation of geographic information in the UK. The MRRL was a joint enterprise between the University of Leicester and Loughborough University.
MUSiC	Metrics for Usability Standards in Computing: an applied research project funded by the European Union in the early 1990's, under its 'ESPRIT' programme to improve the competitiveness of the European information technology industry. The project's partners, a mixture of academic and applied research institutions and commercial companies, developed some standard 'usability metrics' to help software developers identify problems with user interfaces.

NCGIA	National Center for Geographic Information and Analysis: a research consortium of American universities, funded by the US National Science Foundation to perform research into effective use of geographic information, between 1988 and 1997.
plotting	producing colour output of a digital map on paper, usually using a pen-plotter to draw it out to a high degree of accuracy.
projection	the choice of algorithm or strategy for representing the Earth's curved surface on a flat 2D plane (such as a sheet of paper or a computer screen). The relative locations of objects, or the relative sizes of areas, tend to be represented slightly inaccurately as a result.
query	generally, an action whereby a user points or clicks the mouse or digitiser at a map feature, to retrieve a table or window detailing textual and/or numeric attributes of that feature. Sometimes involves additional actions such as setting the software into 'query mode', or specifying the location by some other means than physically pointing (e.g. specifying co-ordinates or names).
raster data	data usually stored as individual pixels, each with a categorical value which may be displayed using different colours or shades. Examples include satellite-derived data showing land use (urban, arable etc.). Raster and vector data may be stored and displayed simultaneously within some more recent versions of GIS, but most systems still tend to reflect one data model more than the other in their functionality and jargon.
search time	one of the usability 'performance metrics' developed by the MUSiC project: time spent with a computer user searching for the means to perform a required command or action, e.g. if unfamiliar with the command menus.
search	in GIS, generally refers to a command instructing the system to find and display (or report about) one or more map features matching certain criteria, e.g. type, size, attribute values, labels.
snag time	one of the usability 'performance metrics' developed by the MUSiC project: defines all time spent doing actions which a computer user then cancels or backtracks, so that nothing is achieved towards the task. Does not include time spent browsing through data, e.g. to check it or to learn from it.
SUS	System Usability Scale: a 10-question 'quick and dirty' usability evaluation questionnaire for computer software user interfaces, developed by John Brooke of the Digital Equipment Corporation in the late 1980's.

USIS	Usable Spatial Information Systems Project: an applied research project run by the Departments of Computer Science and Geography at Loughborough University from 1991-1993, financed by the ESRC
vector data	data stored as individual spatially-separated objects, such as lines, points and polygonal areas. Typical applications include the storage of water pipelines: the pipes can be stored as line objects, and street furniture such as water hydrants as point objects, each linked to an alphanumeric *attributes* record detailing material, size and maintenance records.
VPA	verbal protocol analysis: a method of extracting quantitative measures from people's spoken discourse, e.g. the frequencies of mentioning a particular topic or type of description.
VR	virtual reality: computer-generated imagery, generally made to appear as a life-size environment through which the user can 'move'.
zooming in	altering the displayed scale of a digital map so that details are increased in size, and a smaller geographical area is displayed on the screen at one time.
zooming out	altering the displayed scale of a digital map so that a larger geographical area fits on the screen, but map features become smaller in size.

Chapter 1: Introduction

This thesis reports a series of research activities concerned with understanding what happens when people try to use or interpret a digital map. It begins with the explanatory literature review in this chapter, which will summarise what a digital map is, how it differs from a paper map, why it is an interesting cognitive artefact, and what work has been done and hypotheses suggested regarding our understanding of people's interaction with it.

Chapters 2 and 3 then describe some exploratory analyses of some real-world observational data on people's use of the systems generally employed to store, display, edit and manipulate digital maps, i.e. *geographic(al) information systems (GIS)*. The data was explored with a view to identifying interesting aspects of people's real-life handling of digital maps, rather than depending upon the conjectured hypotheses of the non-empirical literature. Analysis methods were borrowed from the human-computer interaction (HCI) literature, as well as from cartography. The impetus for this whole programme of doctoral research arose out of the author's observations of real-world use of digital maps, and her strong desire to maintain ecological validity and relevance throughout the work. For this reason, rather than produce a series of laboratory experiments imposing controlled but potentially unrealistic tasks, the reanalyses described in Chapters 2 and 3 allowed a more considered understanding of the context and constraints affecting people's use of a digital map.

Based on the indications from these analyses, Chapters 4 and 5 then examine the literature regarding the cognitive processes undergone when people try to understand geographic space, especially when using a map to do so. The studies and theories put forward by psychologists and cartographers are thus discussed in the context of the actual activities of a digital map user, rather than in general terms. Chapter 4 looks at theories and studies concerning spatial cognition, and spatial memory, where deemed relevant to map use. Chapter 5 examines issues concerning individual differences in people's use and understanding of spatial tasks.

The methodological and theoretical concerns raised in those two chapters then direct the final piece of empirical work. Chapters 6 and 7 describe an experimental study designed primarily to investigate one specific aspect of people's cognitive processes with digital maps: namely the cognitive 'reference frame' or mental model people develop when performing different types of task with them. Chapter 4's conclusions about the meaning of reference frames and other mental spatial representations are used to test the relationship between language, task and spatial understanding, and between interpretation of a digital map and photographs of the 'real' environment it represents. Individual difference issues drawn from Chapter 5 are also examined, although practical restrictions on the study limited the conclusiveness of this. Finally, Chapter 8 points to an agenda for further research, and summarises the limitations and findings of the present research approach.

1.2 Introducing digital maps

When a cartographer designs a paper-based map of a city or region, many decisions have to be taken including:

- the map's scale, and its projection (i.e. how the Earth's curvature is to be resolved into a flat plane)

- which features of the area should be shown, how they should be symbolised or encoded, whether they should be accompanied by text labels giving names or descriptions, and where to place those text labels so as not to obscure other features

- how to represent different levels of a variable, such as height above sea level or population density

- what colour and/or texture schemes to use: e.g. for any single-point features such as monuments and telephone boxes, for any linear features such as paths and boundaries, and for any polygon areas such as fields or factories. (Note that at different scales, some features such as buildings may be either a point or a polygon; a river may be either a single uniform line or a pair of lines a variable width apart.)

- whether and how to include a reference grid to enable location of specific features
- how to 'generalise' features such as roads or rivers, whose every bend and angle cannot be shown at a small scale, so that their general 'bendiness' is shown and important bends are specifically noted
- how much to overlap one map sheet with the next, and the exact region to be covered by each one
- how to explain the above choices, via scale bars, legends, titles, etc.

The resulting map is a single, fixed product: although its users can study it and perhaps draw extra details or notes onto it, they cannot obtain any more information from it than has been explicitly drawn, nor easily remove any information to make it simpler to read. The single sheet of paper also limits the amount of information that can ever be included before the map becomes too crowded and complex to understand. Transparent sheets, e.g. made of tracing paper or polythene, could be placed over the top as 'overlays' and used to add extra information or features, but this would be cumbersome and is generally avoided.

The information conveyed by a paper map is under the control of the cartographer: she or he may choose to emphasise some things, simplify others, and omit yet others altogether, and the user is unable to alter the result. In other words, the map may be seen as a potentially powerful act of one-way communication from the cartographer to the user, and every map therefore contains a degree of subjectivity and misrepresentation (Monmonier, 1991). An apparently empty space may contain a secret military installation; the cut-off points chosen in illustrating unemployment levels may be carefully doctored to mislead; the projection used for a world map can make northern countries seem more significant than southern ones.

As with the difference between a printed page and a word-processor file, placing maps on a computer allows the user to edit and change the appearance of the information they contain. The user gains more control over the information, and can choose what to display; Monmonier (1996) argues that this should undermine traditional ways in which maps were designed to mislead their users, but can still prevent them from seeing more data than has been made available. However, in digital form, far more information can be made available to the user than could be fitted onto a paper map, as follows:

- Users can choose to hide or display different combinations of 'layers', showing different feature types or variables, and can thus observe various different views or relationships. 'Layers' can include a reference grid, text labels, and other explanatory features, as well as actual geographical entities (see Figure 1-1 below).

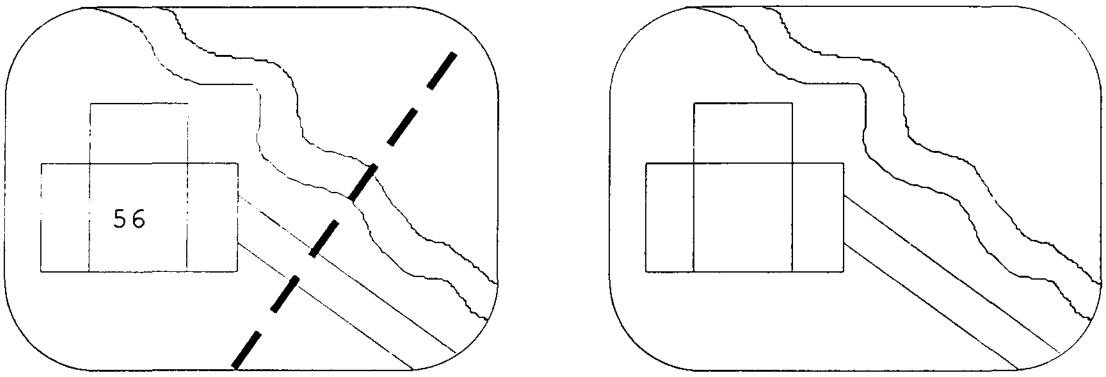


Figure 1-1. Hiding and showing data layers: on the left, house numbers and a cable are displayed along with waterways, paths and buildings, but on the right the first two are 'hidden' or 'turned off'.

- The user may be given the choice over some or all aspects of the map's appearance: symbolisation, categorisation, colour, texture, scale, projection, label placement, generalisation and description.

- Spatial correlations and other statistical relationships between features or variables can be calculated and displayed, to test whether apparent effects are really significant.

- A particular phenomenon (such as floods, emigration or erosion) can be modelled, and the model animated, to show its changes of extent or distribution over time.

- Rather than having definite edges to a sheet, which then have to be matched up with another sheet, a digital map can be continuous and can be much larger than the screen at a given scale: the user can choose to 'zoom out' or 'zoom in' and to 'pan' across the map, to change the area displayed at any given moment.

- A database can be linked to the map so that displayed objects (e.g. a building) can be selected with a mouse click, to display further information (e.g. about the building's history or owners) in a pop-up window. The data linked to the map may include more than simple text records: aerial or other photographs, numeric tables or spreadsheets, and hypermedia entities such as video clips or hypertext could also be included.

Furthermore, the *function* fulfilled by the map tends to differ between a traditional hand-drawn paper product and a digital screen image, as suggested by Unwin (1997, p. 2/108), in the table reproduced below (Table 1-1).

HAND DRAWN GRAPHICS	COMPUTER DISPLAYS
Use symbolism	Often aim for realism (VR)
Are selective	Try to use as much data as possible
Are end products	Are aids to understanding
Demonstrate the known	Detect the unknown
Intended for many viewers	Used by one person
Permanent	Temporary
Used many times	Used once
Restricted dimensions (x,y)	Multidimensional

Table 1-1. Changes in the properties of maps (and other pictures) from traditional hand drafting to computer display, reproduced from Unwin (1997).

One could take issue with some of Unwin's assertions, when considering digital maps: where they are incorporated into a learning package or delivered as part of a general-purpose online service, there are likely to be many users and more than one use, and digital maps are as likely to be used to illustrate known phenomena or decisions as to uncover new patterns in information. As for realism, except for the case of virtual reality (VR), digital maps are often even less similar to the real environment they represent than some paper maps, especially where only selective layers are shown or unusual colour choices selected. Nevertheless, Unwin's points are true to the extent that digital map displays are often partly *intended* to be used in the different ways he suggests, and certainly do fulfil these purposes in academic research settings.

Non-academic uses of digital maps vary enormously, however. One GIS vendor, in a product information sheet (MapInfo®, 1997), lists 50 potential uses for its product. These included, for example, those shown in Table 1-2 below.

Police analyse crimes
Local authorities manage planning applications
Retail companies find locations for new stores
Distribution companies route delivery trucks
Telecommunications companies analyse cellular networks
Farmers optimise fertiliser placement
The military tracks troop movements
Estate agents recommend appropriate listings
Oil companies manage pipelines, wells, and off-shore rigs
Emergency management agencies prepare disaster plans
Natural resource agencies analyse trees, soil type, erosion
Archaeologists map their dig sites
Security companies track stolen vehicles
Political parties understand voter characteristics

Table 1-2. Some uses of GIS suggested by MapInfo (1997)

Some of these uses were illustrated by data from the USIS project, described in Chapters 2 and 3.

Despite this increasing popularity of GIS as a means of representing space and spatially-distributed data, it is worth remembering that the virtual space of a digital map is not the same as the space we see around us. In real space objects do not exist in 'layers' which can be removed independently. In real space, although what we can see is limited by the extents of our eyes and windows and by the horizon, just as our view of a digital map is limited by the screen, we cannot normally 'zoom' in and out to see a greater or lesser amount (except, e.g., by climbing a hill or a building, or by using binoculars). Nor can we, in real life, effectively 'teleport' like a science fiction character directly from one point to another, simply by entering appropriate co-ordinates; we always have to travel through the intervening distance. We are familiar with certain colour schemes and find it unnatural at first to accept others: e.g. roads and buildings are not usually the bright red, green or purple with which they may be coloured in a (paper or digital) map.

In addition, the boundaries of an oblong screen are very artificial, a fact which is frequently overlooked in our media-driven society. This point was nicely discussed by Arnheim (1969):

> The reactions of African natives... make it clear that the human mind does not spontaneously accept the rectangular limits of a picture. Visual reality is boundless; therefore when a film showed persons going off the edge of the screen, the audience wanted to know how and why they had disappeared... Many of our own children learn to accept such breaks of spatial or temporal continuity at an early age, although even they will run into the problem when they face unfamiliar conditions. In a useful study of how well pupils in elementary and secondary schools handle geographic maps, Barbara S. Bartz observed that children sometimes assume a country to end where the map ends. She noted that border lines are often so neat as to give a misleading impression of completeness. [p.310]

Although adult students in contemporary Western society would be unlikely to have these problems, nevertheless one study (Cocks, 1991) showed that people who had viewed a map as a series of zoomed-in screenfuls, rather than as a single paper sheet, recalled features less accurately afterwards. This was largely because they tended to recall the information in 'chunks' which were based on the screen boundaries rather than on meaningful groupings of map features. In other words, the artificial boundary of the screen can reduce viewers' ability to gain a holistic overview of the space which is represented in the digital map. The importance of this depends on whether such a view is important for the task at hand, and whether it can still be gained from zooming far enough out of the map so that all of it appears on the screen simultaneously (at a very small scale).

Simply to interact with a digital map larger than a single screenful thus requires an understanding of the artifice of the screen boundary, and skill in overcoming this via zooming and panning and by retaining off-screen information in memory. Other cognitive processes and technical procedures relating to digital map use will be discussed in Chapters 2-5.

1.3 Previous research

Digital maps are intended to represent geographical space, partly in cartographic form (and partly in the form of numeric values, textual records and often 3D modelling), displayed on a computer, and viewed and manipulated by a human. As such, their use impinges upon at least four academic disciplines: geography, cartography, computer science and psychology. In particular, the subdisciplines where these overlap are crucial to understanding and optimising digital map use: namely human-computer interaction (HCI), cognitive cartography studies, environmental psychology, and spatial aspects of cognitive science. In the past few years the synergy of these areas has been labelled 'geographic information science', and the former *International Journal of Geographical Information Systems* changed its last word to 'Science' to reflect this developing sense of a new body of knowledge. It is worth stepping back a little through the history of this 'science', focusing on the aspects relevant to understanding people's interaction with digital maps.

In the 1980s, as the GIS industry and its penetration into the workplace expanded at an almost explosive rate, the academic community became understandably interested in researchable aspects of these highly complex and powerful systems. The main root from which such research sprang was the geographic/cartographic research tradition, since initially GIS were intended as tools for researchers and decision-makers who wished to visualise, analyse or model the geographic environment. Computer scientists, often 'converts' from a geography background, devoted increasing amounts of time to developing algorithms and decision-support software for modelling geographic phenomena and spatial relations. Much of the research reported at specialist GIS conferences (e.g. Harts, Ottens, & Scholten, 1993) has focused on the development of hardware and software solutions for data analyses or models, or on descriptions of adaptations of the technology to new application areas with unique requirements.

By contrast, the interaction between a GIS and its user was often ignored, except when novel user interface designs were demonstrated (e.g. Raper & Bundock, 1991). This was so despite the increasing and widely recognised tendency for GIS to be used by personnel who had not received specialist cartographic education, in environments where profit or public service replaced research as the organisational goals (Eason, 1993).

However, at the end of the 1980's a small research community began to consider psychological issues regarding GIS use (e.g. Hearnshaw & Medyckyj Scott, 1990; Mark & Frank, 1990; Nyerges, 1991; Medyckyj Scott & Hearnshaw, 1993; Turk & Mackaness, 1993). This grouping was largely based in the US, and partly owed its existence to funding from the US NCGIA, and its successor initiative called 'Varenius'. These researchers, often again originating from the geographical rather than the psychological tradition, have assumed that GIS places unique demands upon its users, particularly with regard to its manipulation of spatial data. This (it is assumed) obliges GIS users to harness their spatial cognitive abilities, and also to gain some knowledge of cartographic terms and concepts, on top of handling a complex information system. Thus it has become normal within this research community to assume that psychological research onto spatial cognition will be crucial to understanding users' ability to learn, understand and use a GIS (e.g. Williams, 1989; Medyckyj Scott & Blades, 1991; Freundschuh & Gould, 1991; Mark, 1993).

1.3.1 HCI aspects

The tendency to focus on GIS as a geographic tool, even among researchers taking a more psychological viewpoint, meant that for some time it was still rarely considered in human-computer interaction (HCI) terms (Medyckyj Scott, 1991). No research had examined how existing GIS shaped up when examined in the context of usability guidelines or standards, let alone how well cognitive user models arising from very different application contexts could be applied to GIS. While several survey studies had examined factors affecting the *organisational* impact of GIS introduction (Firns, 1990; Cornelius & Medyckyj Scott, 1991; Campbell & Masser, 1992; Pinto & Onsrud, 1993), and this is itself an important human factors issue (Eason, 1993), none had penetrated to the level of day-to-day interaction with the system.

Early in 1992, this knowledge gap prompted the initiation of the USIS (Usable Spatial Information Systems) Project at the Midlands Regional Research Laboratory (MRRL[1]), a joint enterprise of the departments of Geography and Computer Studies at the then Loughborough University of Technology and the University of Leicester. The USIS Project aimed to identify problems (if any) in the *usability* of current GIS, in the context of 'real-world' use rather than academic research and teaching, with a view to broadening the agenda for research investigating cognitive aspects of GIS use. Its methods and findings will be described further in Chapter 2.

---

[1] The Regional Research Laboratories initiative was funded by the UK Economic and Social Research Council between 1988 and 1991, primarily to encourage research and development facilitating the use of local/regional data within the UK. Most of the RRLs interpreted this to include research into GIS and related issues. The RRL identity was preserved by several of the research groups beyond the end of the initial ESRC funding.

After USIS had finished, a largely unrelated event took place which aimed to clarify the research agenda for cognitive aspects of digital maps and of GIS functionality. With funding from NATO, the US National Science Foundation and the US Advanced Research Projects Agency, an Advanced Research Workshop (ARW) was convened in March 1994 in Mallorca, Spain, to which the present author was invited. This was the first occasion to be specifically devoted to cognitive aspects of GIS use, and so its contributions and findings were expected by this author to form a focus for her planned Ph.D. work.

The experience of the workshop itself, as with many academic conferences, proved at first to be a little disappointing. The author was the only one of 26 participants to produce empirical data about people's use of GIS (presenting two papers); some of the other participants had not produced a paper for the workshop before it happened; others simply presented user interface designs with only scant reference to their relevance to supporting user cognition; still others wrote from a theoretical perspective but at that time had not yet clarified their speculations into predictive models or hypotheses.

The small-group discussion sessions within the workshop, as summarised in the proceedings volume (Nyerges, Mark, Laurini, & Egenhofer, 1995), did produce a few insights which proved relevant to the work in this thesis, as shown in the table below Table 1-3). In general, however, the disparity of approaches and interests did not lead to any single coherent direction. The focus on fairly shallow and generic HCI aspects of GIS meant that half of the topics listed are in fact discussed before the end of Chapter 3 of this thesis. Beyond that, the author proceeded independently with reviewing data and literature, to establish a more specific research programme.

Discussion topic at ARW	**Insights bearing on present work**	**How addressed in present work**
1. What is special about spatial knowledge?	The distinction between 'what' identity and 'where' location) of objects; object identity not as obvious in geographic maps as in 'real' spaces.	Discussed in Ch 4, and contributing to the basis of the study developed in Ch 6.
2.Primitives of spatial knowledge	None	Not addressed
3. User behaviour	How does a GIS's representation of space influence people's performance of their tasks?	Discussed in Ch. 3; partly addressed by later study in Chs 6 & 7.
4. User interfaces	Software should become more adaptable to users, rather than the reverse	Task suitability issues examined in Chs 2 and 3.
5. Task taxonomy	How to develop one, levels of abstraction within it, use of Rasmussen's hierarchy	Led directly to work described in Ch. 3.
6. Cross-cultural issues	Cultural differences in spatial cognition, esp. relating to digital maps/GIS	Considered in experimental study described in Ch. 6; otherwise not addressed.
7. Computer-supported co-operative work (CSCW)	None	Not addressed.
8. Design for improving GIS	Lack of integration between models of cognition, between HCI and spatial cognition models	Discussed in Ch. 3.

Table 1-3. Issues discussed by the 'break-out' groups at the 1994 NATO ARW

1.3.2 Cartographic aspects

Cartography is a very different discipline from those of geography, psychology or computer science. Cartographers have long viewed themselves as craftsmen and designers, even in the academic context, and their publications are usually either actual maps, or descriptions/studies of them. A more scientific approach, attempting to improve clarity and communication of map features to their users, became gradually popular in the 1970s, when cartographers discovered psychophysics and suddenly put great efforts into studies of 'just noticeable differences' between circle diameters, shades of light and dark, colours, lettering sizes and line thicknesses (see e.g. Gilmartin, 1981). However, after psychologists had moved on from behaviourism to rediscover cognition as a focus for research, cartographers disillusioned with the lack of integration in their psychophysical work also turned their attention more to cognitive aspects of map use.

Cognitive cartography studies became quite popular in the 1980s and 1990s, culminating with MacEachren's 1995 book *How Maps Work*, but many such studies have aimed at geographic visualisation of quite a high order (e.g. mentally visualising overall topography), or else at cartographers' attempts to improve the visual design of maps by focusing on prescriptive findings. We still have limited understanding of how map reading skills are learned, how different spatial reference frames and potential interpretations are applied at different scales even in traditional (non-zoomable) maps, or how the many cognitive studies of expert map use can be applied to novice users (e.g. in public access systems, or in education). In addition, the role of individual differences in spatial and other cognitive aptitudes has been only patchily investigated. When we introduce the interactive element by using digital maps, allowing the previously passive user to alter and manipulate the displayed representation, this adds further to our uncertainty about people's behaviour.

Another problem with much cognitive cartography has been *who* it studies (i.e. which types of map reader, performing which types of task). We already know quite a lot about ordinary map-reading and map use by partly or extensively *trained* adults (such as geography undergraduates and military personnel): it is these groups which tend to feature most heavily as subjects in cartographic experiments. A number of such studies in recent years have made their subjects study a map and then recall it later, either by drawing it, describing it, or answering questions about it (e.g. Rossano & Hodgson, 1994; Kulhavy & Stock, 1996; Curiel & Radvansky, 1998). The point of this type of task is not always entirely obvious - outside the military, memorising maps is hardly a skill that demands daily application.

However, these experiments have at least demonstrated a consistent tendency for our spatial memory for maps, like our memory for larger-scale spaces and for other small-scale scenes, to be organised hierarchically into 'clusters' based on significant landmarks and familiar features, and to be biased according to expectations of the rules and conventions of the environments we live in. Although one must always be wary of confusing memory recall processes with initial perception, Chapter 4 will show that the same biases seem to influence our perception and understanding of a map placed in front of us, and therefore we have some idea of how a user may be influenced by the design of a digital map. However, we will see that little work has been performed in that context.

More specifically, in the educational context, certain recent studies have suggested aspects of map design that can facilitate or inhibit successful inference and learning from the map display. For example, one study (Rittschof & Kulhavy, 1998), while finding that regions and information about them were remembered better when learned from a map than from a table, choropleth maps were also more successful than proportional-symbol maps in terms of student recall.

Work on children's abilities with maps has suggested a surprisingly early ability to learn and understand them (e.g. Blades & Spencer, 1987), shrugging off the older Piagetian assertion that such skills have to be learned relatively late in childhood owing to their symbolic nature, and that they depend heavily on dedicated instruction. Yet at the same time, the literature throws up evidence of real problems in spatial understanding among unskilled adults (Giraudo & Péruch, 1992; Golledge, 1992). Clearly the individual cognitive skills involved in map reading deserve careful teasing out. Broadly speaking, certain themes have tended to run through the (relatively few) papers which appear to directly imply direct hypotheses about these skills:

1. Differences and possible interactions between users' understanding of the perceived 'small-scale' space (the map, and the screen that displays it) and 'real' or 'large-scale' environments (the space we walk around in).
   (Nunes, 1991; Mark & Freundschuh, 1995)

2. The contrast and possible conflict between the system- and map-oriented cognitive models: in other words, between the demands made by the user interface of the GIS, and those of the graphics and semiotics of the displayed maps.
   (Davies & Medyckyj Scott, 1995; Mark & Freundschuh, 1995)

3. Understanding, describing and interpreting relations between spatial objects, using various 'reference frames'.
   (Hirtle & Heidorn, 1993; Lloyd, 1993; Logan, 1995)

4. The role of prior expectations/expertise (both generic expertise, and expectations peculiar to that map) in interpreting the visible display.
   (MacEachren, 1995)

5. The hierarchical categorisation of areas or 'chunks' of the display as representing (different types of) specific features or objects.
   (Nunes, 1991; Tversky, 1992; MacEachren, 1995)

The diagram below (Figure 1-2) attempts to broadly sketch out the relationships between these and other ideas concerning people's cognition when using digital maps, and to show where they will be discussed in this thesis.

Topics covered in this thesis are shown in rounded-corner boxes; the major contributing disciplines are shown in bold in square-cornered boxes. The diagram shows (with solid lines) which discipline has made inroads towards understanding a topic; where two or more topics tend to be closely linked within that discipline their lines meet at a point. Where topics have generally *not* been well linked in the literature, but where attempts are made to draw closer links between them in this thesis, this is shown with dotted lines.

Thus, the tendency for generic and specific knowledge/memory of a map to be considered by the same branch of psychology (e.g. Denis, 1996), but for long-term expertise/training and short-term specific interpretation to be treated largely as separate issues in cartography (MacEachren 1995), are reflected in the way the discipline-topic lines are grouped for those two topics. In the case of expertise or long-term generic schemata, there is an obvious link to other sources of individual differences, which is made in Chapter 5 of this thesis although not extensively explored in psychology: hence the dotted line linking them despite the non-grouping of their discipline-topic lines. The tendency for HCI practitioners to look at either ergonomic design (e.g. Temple, Barker & Sloane, 1990) *or* usability measurement (e.g. Neilsen and Mack, 1994) *or* task analysis (e.g. Whitefield and Hill, 1994), but not to bring them together in understanding specific single contexts (at least in published studies), is reflected by the disparity of the discipline-topic lines, but the close links drawn in this thesis between their implications for digital maps are shown as dotted inter-topic lines.

Psychologists interested in memory have considered mental representations of spatial knowledge in at least three different ways, grouped on the diagram: semantically-influenced 'chunks' (e.g. Gilhooly et al, 1988), reference frames that influence or are influenced by spatial language (e.g. Carlson-Radvansky and Irwin, 1994), and notions of survey- or map-like overviews of an area or array (e.g. Hirtle and Heidorn, 1993). The latter has also been an area overlapping environmental psychology and human geography, and studied by a distinctive cross-disciplinary grouping of researchers, e.g. Gärling & Golledge (1993). Therefore it is linked to both these and to cognitive/general psychology in the diagram. Closely related to it, as far as environmental psychologists/geographers are concerned, is the transfer of knowledge between maps and 'real' large-scale space, which has also been a crucial consideration for cartographers trying to improve map design (e.g. Wood, 1993). Finally, the contrast and potential conflict between system- and map-oriented cognitive models of a GIS has been mentioned by cartography/GIS-oriented authors (e.g. Medyckyj-Scott and Blades, 1990), but not in any other disciplinary context, although this thesis draws a link between it and the problem of usability measurement.

Not all the ideas explored in this thesis are represented in the diagram, however. Among issues discussed later is at least one key issue which arose in the course of the work and became central to it, but which had previously not been addressed in any of the literature (and hence cannot be linked clearly to any existing discipline). This is the extent to which a digital map is understood, by a given user performing a given task, as a *map* at all, rather than just used and edited as a visual array, or as a means to access records of information. Most of the literature to date has assumed that the geographic representativeness of a map is fully interpreted and mentally encoded as such, regardless of the situation in which it appears, provided the display is recognised as intended to be a map. This assumption, rarely explicitly stated, is discussed in Chapter 4, and further explored in the study reported in Chapters 6 and 7.

Although the diagram is quite complex, it serves to illustrate the nonlinearity of the concepts discussed in the next few chapters, which are broadly ordered according to chronology of the research the author performed. Like any thesis, this one has to form a sequential narrative, when often a hypertext document would better represent the links among the many ideas.

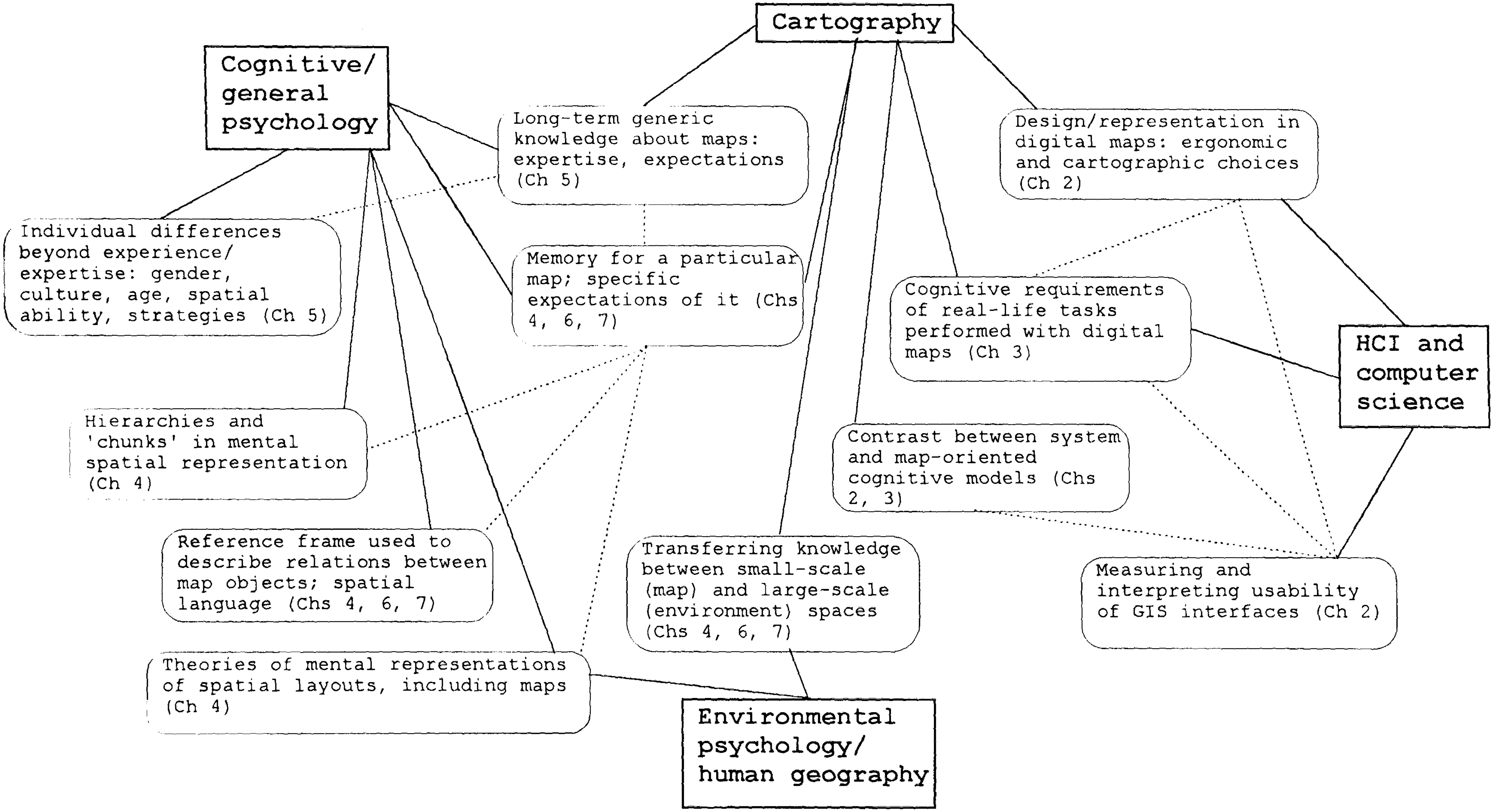


Figure 1-2. Diagram of major ideas and disciplines contributing to this thesis

1.4 Summary

The potential for user-map interaction provided by digital maps, as well as aspects of their design and display, pose new questions for our understanding of map use and interpretation, as well as renewing older issues concerning any map. This has been increasingly recognised by cartographers, geographers, psychologists and computer scientists over the past two decades, but limited empirical work has been performed; approaches to it have been disparate and sometimes based on discipline-bounded assumptions. Issues raised in this chapter relate particularly to the human-computer interaction aspects of digital map use, and to the cognitive models and memory representations developed from interpreting the map itself. These issues are interlinked and cross disciplinary boundaries in many ways, and will be inevitably interwoven in the remainder of this thesis. The overall approach will emphasise real-world tasks and ecological validity over laboratory control; thus Chapters 2 and 3 will be concerned with reanalysis of some real-world data, and with literature considering the nature of map users' actual tasks.

Chapter 2: GIS and USIS: usability of systems and maps

The research reported in this chapter follows directly from earlier work investigating usability issues in geographic information systems (GIS). This earlier project at Loughborough (the USIS Project[2], described further below) had yielded a large amount of usage data which had only been summarily examined by the end of the project, but held the potential to reveal interesting cognitive issues and pointers to further research.

In the late 1980's and early 1990's, the GIS research community conjectured that the use of a computer system handling digital map data would necessitate the use of the same cognitive skills that humans use in navigating through geographic space, and/or in interpreting paper maps, plus skills relevant to the HCI aspects of the system. This combination was assumed to form an interesting research area (as illustrated by the NATO ARW in March 1994, described in Chapter 1). However, virtually no *empirical* evidence had been presented to indicate which cognitive issues are particularly interesting regarding digital map users in real situations. The present author therefore decided to re-examine the USIS data describing real-world GIS use, mostly from Phase II of the project (an 'observation study' consisting of a series of case studies in actual workplaces) to identify potentially relevant issues.

The exploratory reanalysis focused first on HCI aspects of GIS, and subsequently on cartographic aspects of the maps displayed in typical use. The results of these system- and map-oriented reanalyses form the bulk of this chapter. In both this and the next chapter, a particular focus of concern will be any evidence concerning the role of the geographic representation in the digital map display. As discussed in Chapter 1, this has been assumed to be the main issue of interest in digital map use, but here it will be set in the context of the overall tasks and user-system interaction.

---

[2] The author is grateful for discussions and references from other members of the USIS team: David Medyckyj-Scott, Val Byrne, Colin Monckton and Peter Fisher. These helped her to shape and interpret the analyses in Chapters 2 and 3. The interest and co-operation of the participating GIS users is also gratefully acknowledged.

The sections below in this chapter will describe:

2.1 Description of the original USIS Project (not part of the present author's independent doctoral work); summary of the available data relating to HCI aspects of GIS in workplace settings

2.2 Summary of a re-analysis of the initial usability results, focusing on those which appeared both statistically significant and relevant to cognitive aspects of digital map use; implications and directions for further research suggested by the data.

2.3 Summary of a subsequent collaborative project, cartographically analysing the visible digital maps displayed in the USIS observation study videotapes; presentation of overall results of this, and issues raised regarding the methodology and implications.

It should be noted that the data collection and basic usability evaluation from the two funded phases of the USIS Project are not described in great detail, and are not counted as part of this author's original work towards a Ph.D. The USIS Project was not initiated, led or solely executed by this author, although the analyses detailed below have been performed by her alone and at her own initiative.

2.1 HCI data from the USIS project

2.1.1 Initial postal questionnaire survey

The USIS (Usable Spatial Information Systems) project had begun in January 1992, funded by the Economic and Social Research Council as an extension to the Midlands Regional Research Laboratory funding programme, and ended (in funding terms) in August 1993. The project's objective had been exploratory rather than experimental, i.e. it sought to identify key end-user issues for GIS in typical workplaces, rather than testing any prior hypotheses. It had consisted of two phases:

I. a postal survey of GIS users, asking questions on usability of software, hardware, training courses and documentation, based on standard HCI recommendations;

II. an observation study involving visits to actual workplaces and examination of how end-users actually interacted with their GIS, what conditions they worked under, and how usable the GIS appeared in finer detail.

The Phase I survey questionnaire had been designed and distributed by the two originators of the project: David Medyckyj-Scott of Loughborough University's Department of Computer Studies, and Hilary Hearnshaw at the University of Leicester's Department of Geography. It had included initial questions about the respondent's system (software and operating system), tasks, training and experience, frequency of use and organisational environment. The bulk of the questionnaire had consisted of 90 questions on how people found working with their GIS. To give the project a baseline to work from, most of the questions had been derived from the descriptions of seven basic 'usability principles' given within Part 10 ('Dialogue Principles') of the then-draft international standard ISO 9241[3]. These are listed below, with shorthand names used in the project given in italics:

Suitability for the task (*Suitability*): the extent to which the GIS supports the user in the effective and efficient completion of the task;

Self-descriptiveness: the extent to which dialogue steps are comprehensible through feedback from the GIS or help requested by the user

Controllability: the degree to which the user is able to maintain control over the course of the interaction until the current goal of GIS use has been met

Conformity with user expectations (*'Predictability'*): the degree to which the GIS fits the user's experience, education and knowledge of the task

Error tolerance: the extent to which the user can achieve the intended result from the GIS with minimal or no corrective action taking place.

---

[3] Although this standard was not finally published until 1996, the author's then colleague David Medyckyj-Scott was one of the expert panel to which drafts were sent for comment at each stage of the long, slow process of international agreement, so the USIS project took advantage of this 'inside knowledge'.

Suitability for individualisation (*'Individualisation'*): the extent to which users can customise their GIS to their individual needs and skills, for a given task

Suitability for learning 'Learnability': the extent to which the GIS provides understanding and guidance to the user during the learning phases.

The questions had also incorporated the System Usability Scale (SUS), a 10-item scale developed some years ago by the Digital Equipment Corporation as a 'quick and dirty' measure of general usability, which had earlier been shown (Wong and Rengger, 1990) to correlate highly with other, longer, usability questionnaires developed in the HCI field.

All these usability questions took the form of statements about the respondents' GIS. Respondents had to circle a number from 1 to 5 on Likert scales ranging from 'Strongly disagree' (1) to 'Strongly agree' (5). Half the items in the overall questionnaire were randomly transformed to read negatively (e.g. 'I like to use the GIS frequently' ▩ 'I do not...'), to minimise response generalisation. The full list of questions, with explanatory notes, is shown in Appendix 1.

The survey had been distributed to organisations known to be GIS users in the UK, the rest of Europe and Australasia (America, probably the biggest GIS market, could not be covered due to the lack of a US-based researcher to identify potential recipients and distribute the questionnaire). There were 159 usable responses (37% of the number sent). After all responses had been received, extensive analyses (largely performed by the present author) yielded a number of findings, which were variously published and publicised (Medyckyj-Scott et al, 1993; Davies and Medyckyj-Scott, 1993a; Davies and Medyckyj-Scott, 1993b; Davies and Medyckyj-Scott, 1993c; Gooding, 1993; Davies and Medyckyj-Scott, 1994a; Davies and Medyckyj-Scott, 1994b; Davies, 1994).

The survey had revealed some key aspects of GIS usability which were problematic, despite showing a generally positive response overall. In particular, error messages were very poorly rated by respondents, and were often apparently incomprehensible to them. User and system documentation were also given low ratings for their helpfulness and usefulness. Although many GIS lend themselves well to customisation through the use of macro programming, respondents whose GIS had been customised found it no easier to use than those using the 'off-the-shelf' version of the same product; furthermore, many respondents were unable to make simpler adjustments to adapt the user interface to their own preferences. The complexity of many systems appeared to serve only to frustrate the many users who used GIS for relatively simple routine mapping and inventory tasks, since they found response times to be needlessly slow and the interface confusing. Finally, ratings of GIS usability varied enormously not only between different types of software, but also between types of user organisation and between users at different levels of expertise, suggesting that there may be no such thing as an ideal generic GIS user interface.

As a result of these findings, various recommendations for improving GIS design and user support were published (Davies and Medyckyj-Scott, 1994). The results also indicated various possibilities for future research into overall organisational and HCI issues (Davies and Medyckyj-Scott, 1993).

Among the issues not listed at the time was a concern in the present author's mind about unmeasured, possibly psychological, factors which could have affected responses to the survey. Tentative multivariate analyses of the results failed to show any specific variables which strongly predicted groupings among respondents, out of the detailed personal variables collected (such as training length and type, length of GIS and computing experience, frequency of use, organisation and task type, system type, etc.). This problem was not specifically tackled in the observation study except to the extent of anecdotal evidence, partly because the small number of users would make further multivariate analyses impossible, and partly because the project was seen to be about testing systems rather than users. The open co-operation which was obtained in the observation study would not have been experienced had the study included any personal intrusions or psychometric tests. We can only look at secondary, indirect indicators that could suggest such individual differences, since the data in neither phase of the project directly measured cognitive processes.

2.1.2 Field study: workplace observation

The second phase of the project, the workplace observation study, involved visits to workplaces around the UK where GIS were in regular use. The intention of this study was to add more depth of insight into the problems encountered by GIS users, by combining direct questioning and observation with more objective measures of user/system performance. As in the survey questionnaire, effort was made to adopt established and validated techniques of usability evaluation, although in both phases this was inevitably limited by the pioneering nature of the study itself – most HCI studies in the past had evaluated only one system, or systems in only one organisation, or a set of less variable systems (e.g. text editors) which were easier to compare. Impartial, comparative, end-user evaluations in the workplace setting, as opposed to a usability testing laboratory, are relatively rare in the HCI literature, perhaps due to commercial confidentiality restrictions.

The approach taken in the observation study was therefore a combination of evaluation methods, aimed at gathering both subjective and objective data with the hope of comparing and summarising across different products, organisations, application areas and user types. To achieve this, the visits were carefully planned and prepared; each then followed a standardised agenda including structured and semi-structured interviewing, and observation and video recording of typical work tasks being performed using the GIS. The method followed is described below in more detail, to explain the origin of the collected and subsequently reanalysed variables.

In total, 23 visits were made, including two pilot visits to establish the method. Respondents to the earlier survey had been asked to indicate if they would be happy to be involved in further, on-site, research. Of those who responded positively, those selected were chosen so as to maximise the spread of geographical location, type of organisation, type of system in use and the purpose to which it was put. The visits were thus distributed as far afield as Edinburgh and Southampton, with a very broad range of application areas (from local authority planning and highway departments, through utilities and commercial companies, to environmental and other research institutes), and a number of markedly different GIS products (running on various hardware platforms). The total duration of each visit was generally just under 2 hours, but sometimes longer when users chose to spend longer discussing their system. Only one user was observed and interviewed in each organisation; in general s/he was either the person who had previously completed the questionnaire, or was selected by them. They thus tended to be either the person charged with managing the GIS implementation/use, or an experienced and articulate junior staff member. All were regular GIS end-users: this requirement was stressed before the visits took place.

Table 2-1 (next page) summarises the characteristics of the various user sites. Since the focus was on the GIS and its use and usability, personal user characteristics such as age were not collected. Precise timing information about the visits is not available.

Clearance to videotape the users' interaction with the GIS was obtained prior to site visits. Only one organisation refused, though it was visited anyway and other data gathered. It was explained that the videotapes were to help gather objective measures of system usability, rather than to examine its use by the organisation, and strict confidentiality was assured. The users were almost always quite relaxed about the video camera, once it was set up (slightly behind and beside them, to capture the screen without intruding on or recording the users themselves). In general, around 20-30 minutes of interaction was recorded (min=9:05, mean=23:42, max=42:28), and generally terminated when the user completed a set of tasks This and other timing variables are discussed further later in the chapter.

Organisation type	**GIS type[4]**	**User gender**	**Self-rated GIS knowledge**	**Months using this GIS**	**Tasks on video**	**Survey questionnaire completed?**
Utility (water)	V	m	unknown	unknown	Open map; query; digitise	No
Retail	V	m	High	42	Open map; output map; digitise; convert data; print report	In phase I
Local govt. (highways)	V	m	Medium	24	Open map; output map; measurement; convert data; edit map	After visit
Environmental research	V	f	High	unknown	Open map; output map; edit map	After visit
Geo. research	V	m	High	6	Open map; query; attribute search; convert data	In phase I
Local govt. (general)	V	m	Medium	12	Open map; query; output map	After visit
Local govt. (planning)	V(/H?)	m	unknown	unknown	Open map; query; attribute search; output map	No
Geo. research/ consultancy	R	m	High	32	Open map; query; output map; spatial analysis; convert data	After visit
Local govt. (planning)	V	f	Medium	18	Open map; digitise; edit map	After visit
Telecoms	V	m	Low	26	Open map; attribute search; output map	After visit
Regional govt. (demographics)	V	m	Medium	18	Open map; query; convert data	After visit
Local govt. (research)	V(/H?)	m	Medium	66	Open map; query; attribute search; output map	In phase I
Local Govt. (planning)	V	f	Medium	42	Query; output map; digitise; add attributes; measurement; edit map	After visit
Geo. research/ consultancy	H	m	Medium	6	Open map; query; edit attributes; edit map	After visit
Environmental management	R	m	High	18	Open map; query; output map	After visit
Environmental management	V	m	High	2	Open map; edit map	After visit
Utility (electric)	V	m	Low	3	Open map; query; attribute search; digitise; add attributes; spatial analysis; edit map	In phase I
Utility (water)	V	f	unknown	unknown	(no video permitted)	No
Local Govt. (accident analysis)	V(/H?)	m	High	3	Open map; query; attribute search; print report; measurement	In phase I
Utility (water)	V	f	Low	24	Open map; edit attributes; edit map	After visit
Regional govt. (housing)	V	m	High	3	Open map; query; attribute search; digitise; print report; output map	In phase I

Table 2-1. Summary of the 21 post-pilot GIS user visits in observation study

---

[4] R=raster, V=vector, H=hybrid

The method which was chosen to evaluate subjective aspects of the interaction was the Ravden and Johnson (1989) usability evaluation checklist. This consists of eleven sections. Sections 1 to 9 ask the user/evaluator to rate the system's degree of conformance to various aspects generally considered to facilitate ease of use: there are around a dozen questions in each section. The section headings are listed in a table below (Table 2-2), along with Ravden and Johnson's descriptions of their purpose. In total, there are 129 individual questions, plus nine 'overall' ratings of the nine aspects. The questions mostly use a 4-point Likert rating scale where one option has to be selected from *'Always'*, *'Most of the time'*, *'Some of the time'* and *'Never'*, and there is room for comments to be added. The 'overall' rating question is worded almost identically in each section, as illustrated in Figure 2-1.

Overall, how would you rate the system in terms of visual clarity? (Please tick appropriate box below.)

Very satisfactory	Moderately satisfactory	Neutral	Moderately unsatisfactory	Very unsatisfactory

Figure 2-1. Example of 'overall' question at end of section (the words 'visual clarity' were replaced by the relevant section heading each time, but the format was otherwise identical)

Ravden and Johnson envisaged that their checklist would be used in a situation where a new software product was being developed. In such a situation, they assumed, it would be possible for evaluators to be thoroughly briefed on the principles, purpose and use of the checklist, and then actually use the system to carry out realistic tasks before rating its usability. In the present study, however, the systems were already in use, and it was impossible for the researchers to act as hands-on user/evaluators, while the existing users would not have time to be briefed on the principles and use of the checklist or to complete it during the visits. Therefore, the USIS researchers split the questions into those which necessitated a response from a hands-on user (e.g. to judge whether the system's way of working fitted their expectations), and those for which the researchers could later act as evaluators by observing the videotapes (e.g. to state whether colours were used consistently within the user interface).

Heading	Description (Ravden and Johnson 1989)
1. Visual clarity	Information displayed on the screen should be clear, well-organised, unambiguous and easy to read. Sample question: "Does the use of colour help to make the displays clear?"
2. Consistency	The way the system looks and works should be consistent at all times. Sample question: "Is the method of entering information consistent throughout the system?"
3. Compatibility	The way the system looks and works should be compatible with user conventions and expectations. Sample question: "Is information presented in a way which fits the user's view of the task?
4. Informative feedback	Users should be given clear, informative feedback on where they are in the system, what actions they have taken, whether these actions have been successful and what actions should be taken next. Sample question: "Are messages displayed by the system relevant?"
5. Explicitness	The way the system works and is structured should be clear to the user. Sample question: "Is the structure of the system obvious to the user?"
6. Appropriate functionality	The system should meet the needs and requirements of users when carrying out tasks. Sample question: "Is the way in which information is presented appropriate for the tasks?"
7. Flexibility and control	The interface should be sufficiently flexible in structure, in the way information is presented and in terms of what the user can do, to suit the needs and requirements of all users, and to allow them to feel in control of the system. Sample question: "Can the user choose the rate at which information is presented?"
8. Error prevention and correction	The system should be designed to minimise the possibility of user error, with inbuilt facilities for detecting and handling those which do occur; users should be able to check their inputs and to correct errors, or potential error situations before the input is processed. Sample question: "Is it easy for the user to correct errors?"
9. User guidance and support	Informative, easy-to-use and relevant guidance and support should be provided, both on the computer (via an on-line help facility) and in hard-copy document form, to help the user understand and use the system. Sample question: "Is it clear how to get in and out of the help facility?"

Table 2-2. Sections 1 to 9 of the Ravden and Johnson usability checklist

Section 10 of the Ravden and Johnson checklist lists 25 common problems encountered by users of computer systems, which have to be rated as 'No problem', a 'Minor problem' or a 'Major problem'. The researchers added 10 further problems which had been specifically mentioned by respondents to the earlier postal survey, some of which - e.g. bias towards vector or raster data handling (see Glossary), or problems converting incompatible geographic data formats - were specific to GIS.

Section 11 of the checklist asks some general summary questions to establish the user's judgement of the best and worst aspects of the software, aspects which are particularly confusing or irritating to the user, aspects they would like to change, and final comments.

As well as the checklist, responses to the survey questionnaire used in the previous study were obtained for 18 of the 21 users visited (either from having completed the survey previously or completing it after the researchers visited). These provided useful context information about the users' background and experience, as well as extra subjective rating data to complement the checklist.

To complement the subjective data gathered in the checklist and in the survey questionnaire, the videotape recordings of users' task demonstration sessions were transcribed (including timings in seconds) on an action-by-action basis. The level of detail was deliberately chosen to be just above system-specific descriptions; in other words, just above the level of individual keypresses or mouse movements which have been described as 'interfacing responses' (Shepherd, 1989). This decision was based on the study's aim to compare findings across different GIS, which involved different sequences of physical actions to achieve the same subtask goals. Its initial purpose was as an extra descriptive tool, to enable the researchers to summarise the variety and speeds of actions taken by participants in their daily work. Further use of this 'performance' data is described later in this and the next chapter.

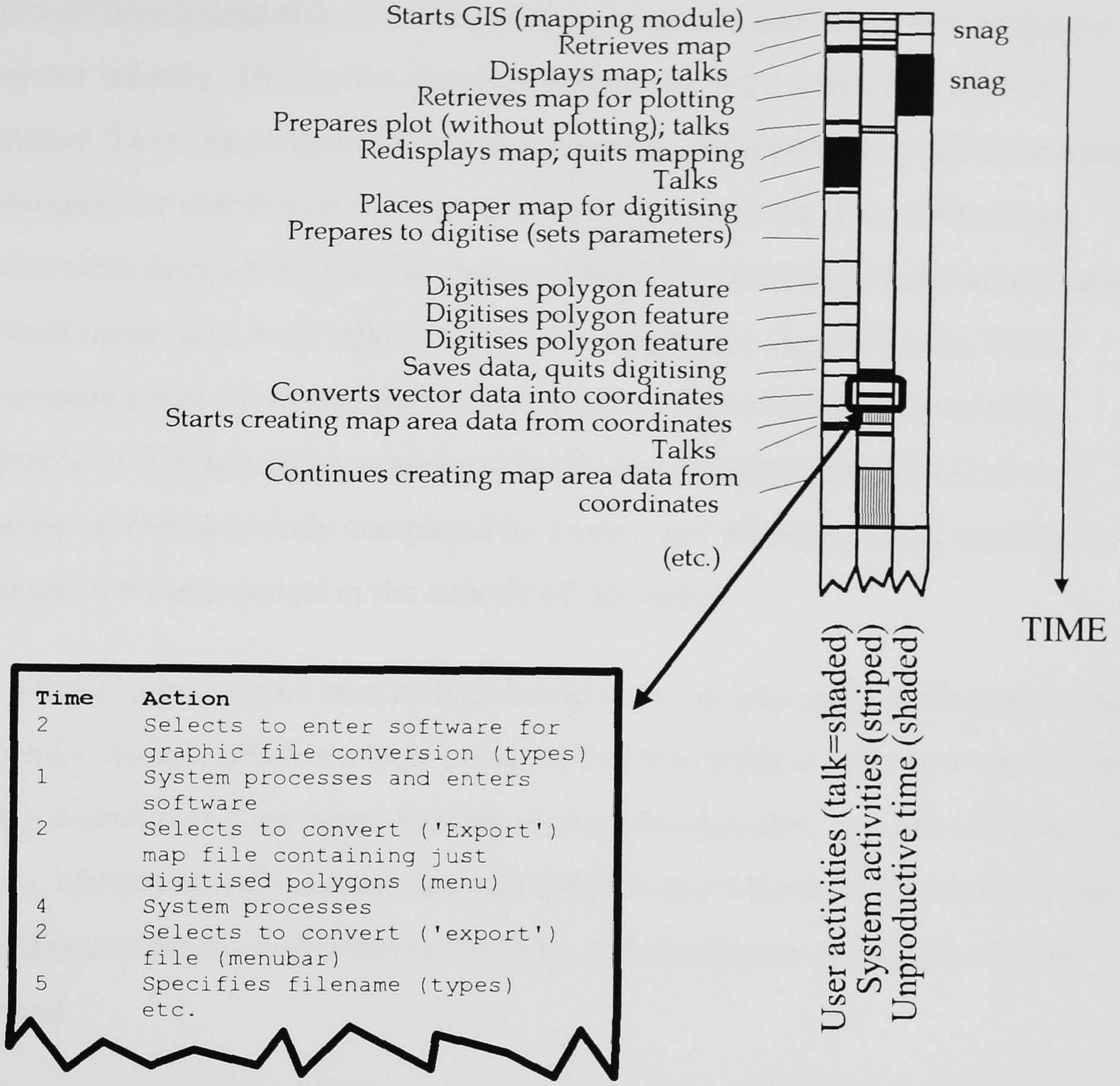
The USIS Project suffered from delays and discontinuities, owing to staff changes at various points in its duration. By August 1993 (the project's end date) the visits had been completed, and the time-consuming and painstaking transcription of the videotapes had been performed by the present author (much of which took hours of frustrating rewinding and pausing due to poor camera focus, brevity of user actions and the lack of explicit information on the GIS screens about what was going on), but little statistical analysis had been performed. The data analysis at that time concentrated on obtaining findings which could complement or contradict those of the earlier survey, and which could point towards an overall understanding of users' work and problems. The final project report for the ESRC, which was accepted as satisfactory (as part of the MRRL's overall report on its work), concentrated on summarising basic findings related to the earlier ones.

After the end of the USIS funding, it was tempting to draw a line under the results and recommendations, and to archive the data. However, after registering as a part-time research student in October 1993, the author felt that the data potentially held further findings. The decision was taken to embark upon a series of systematic statistical analyses of the many variables collected in the observation study, with the intention of identifying results which could prompt further research.

2.1.3 Available variables

As stated above, variables available for further analysis of the observation study data were derived from the Ravden and Johnson (1989) checklist, the users' responses to the survey questionnaire used in Phase I of the USIS project, and the videotaped task demonstration sessions.

The illustration below (Figure 2-1) shows the different levels of analysis considered in the videotape-based data. An overall timeline may be drawn summarising the basic events and activities, part of which is shown to the right of the picture. Any of the list of user activities shown to the left of this timeline, identified loosely at a 'subtask' level, is itself a collection of individual actions as defined earlier; some of these actions are listed in the extracted sequence of notes on the lower left of the diagram, which are typical of the action-based timings and notes extracted from the tapes. The horizontal lines and shaded portions of the timeline show the start and end of activities; the occurrences of 'snags' (see below); periods when the system was processing and the user had to wait until it finished; and 'talk time', where certain users were speaking while no other event was occurring (despite repeated requests not to do so – such time was eventually recorded as a variable in its own right).



Time	Action
2	Selects to enter software for graphic file conversion (types)
1	System processes and enters software
2	Selects to convert ('Export') map file containing just digitised polygons (menu)
4	System processes
2	Selects to convert ('export') file (menubar)
5	Specifies filename (types)
	etc.

Figure 2-2. Levels and types of activity transcribed from the USIS observation study videotapes

The descriptions and counts of low-level actions served as indicators of the degree of repetitiveness of users' tasks, the speed with which they performed them, and the presence of 'bottlenecks' such as particularly slow or constantly-repeated actions. One might expect particular aspects of the system's usability to be associated with these indicators; for instance, one might expect users performing more repetitive tasks to be less bothered by lack of flexibility or limited functionality in the system, or by having to solve complex error situations. One might expect users who manage a high mean speed of performing each action to complain less about response times. One might expect the presence of 'bottlenecks' to be associated with a low compatibility with the user's task needs and expectations.

The timings of actions were also used to calculate performance metrics based on those developed by the ESPRIT MUSiC Project (MUSiC, 1992). MUSiC (Metrics for Usability Standards in Computing) was a collaborative European Union-funded project involving academic and research establishments and companies within the computer industry. The project developed and validated psychophysiological, predictive (based on analysis of system functions), subjective and performance metrics for gauging the usability of software and hardware products. The performance metrics were designed by the UK National Physical Laboratory; these were deemed the most up-to-date, well-validated and easily applicable metrics for the USIS observation study, given that the MUSiC subjective metrics were not as easily adaptable to GIS-specific concerns as Ravden and Johnson's checklist (and they depended on being entirely completed by a user), and physiological or predictive measures were impractical in the context of the study.

The MUSiC metrics used here included *snag time*, i.e. time spent dealing with 'snags' (generally, actions which are later cancelled and thus achieve nothing towards task goals), *search time* (time spent looking through functionality, e.g. browsing through menus, without actually issuing any commands from them) and *help time* (time spent obtaining assistance either from online help, documentation, colleagues or other sources).

The total time spent *not* dealing with snags, search or help is labelled *productive time*, and this can be converted into a *productive percentage* of overall task time to give a proportional measure, comparable across different users and tasks. The use of the MUSiC definition of 'snag time' rather than 'errors' in this context has been discussed elsewhere (Davies and Medyckyj-Scott, 1995): in summary, it avoids the assignment of blame, and allows inclusion of situations where users have to backtrack but no error message is produced.

Overall, including variables derived from the survey, checklist, actions and performance metrics, the data re-analyses involved well over 100 variables[5]. With only 21 users (and with only 20 on tape and only 18 survey questionnaire responses from them), obviously this number of variables must be treated with great caution: the analyses had very limited statistical power and could only be taken as indicators of

---

[5] Unfortunately, in the several-year time lapse since the original study, non-significant data analysis results have become lost, so it is not possible to show full correlation matrices in this thesis.

potentially interesting relationships, not as strong evidence for one hypothesis or another. In each analysis, missing data was excluded on a pairwise basis in order to maximise the information included in the analysis; for this reason, *n* varies in the analyses reported below. It was still felt that quantitative analysis was appropriate since this is (or was at the time) more usual than ethnographic methods in usability evaluation, and facilitated comparisons and conclusions to be drawn across all the case studies. Nonetheless, qualitative data such as users' comments and general working conditions were also carefully noted, and key issues from them will be summarised briefly below.

2.1.4 Analysis strategy

The following sequence of analyses was performed on the USIS observation study data after October 1993, when the author's doctoral research officially began.

1. General overview of data in each of the variable categories, and observation of general trends; examination of relationships between different variables within each category (timings, action frequencies, checklist scores and survey responses).

2. Summarising data for each of the 21 individual case studies; interesting aspects of each user's behaviour and system use.

3. Examination of relationships between variables in different categories, and possible trends indicated by particular patterns of relationships.

Verbal protocol analysis (VPA) was also considered as a possible data extraction method. This is perhaps the most common method of attempting to extract data relating to cognitive processes from videotaped interaction or discussion, and usually involves noting the frequency and timings of occurrences of particular types of utterance that are of interest to the researcher (Ericsson and Simon, 1993). However, VPA stands the most chance of success if the discourse is specifically and exclusively focused on relevant tasks or topics, and if encoding rules can be defined which are almost certain to capture the effects that are of interest, if they exist. It also depends upon the discourse itself being a true reflection of the cognitive processes beneath, rather than a forced attempt by the subjects to tell the researcher what they think s/he wants to hear. The author eventually decided that the discourse data collected in the USIS visits was too mixed and researcher-influenced to be suitable.

2.2 Key Findings

2.2.1 Observations and qualitative data

One aspect of users' systems which had not been covered in the earlier survey was the ergonomic conditions in which the users worked. Although the lighting conditions of users' workplaces were found to be surprisingly poor, causing strong reflections and glare from the screen in most cases, users often had apparently not considered them until the time of the researchers' visit. It might be expected that the lighting conditions could affect users' ability to identify and manipulate spatial data on the screen, certainly worsening any problems in navigating through or editing the displayed maps, but little evidence was seen of this.

In fact, there was little evidence in general of users having problems with the visual or cartographic aspects of their tasks (as opposed to the software use aspects). Users' comments, and statements about the 'best' and 'worst' aspects of their systems, nearly always focused on issues such as command syntax, case sensitivity, error messages etc. Their most common mistakes (according to their own estimates) tended to be typing errors, or mistaken identification of files etc., rather than errors of spatial judgement or navigation, although one user claimed to frequently delete the wrong spatial object, and another said her work often failed quality assurance checks owing to inaccurate placement of data.

The 35 questions in section 10 of the Ravden and Johnson checklist showed some common problems arising for a number of users, while others were more idiosyncratic. The 'top 10' common/major problems were as follows, with the most common/major problems first (this somewhat crude ranking was achieved by assigning values of 0, 1 and 2 respectively to users' ratings of problems as 'No problem', 'Minor' or 'Major', and summing across users):

1. Bugs and crashes
2. Missing functionality

3. Memory/processing power
4. Slow response times
5. File import/export problems
6. Poor documentation
7. Unexpected system events
8. Solving error situations
9. Printing/plotting problems
10. Getting 'lost' within the system

Some of these problems replicate Phase I findings; note also the incidence of items such as "Unexpected system events" and "Getting 'lost' within the system", which may reflect incompatibility between the system's and user's cognitive models.

Similarly, when asked what aspects of the system users found confusing or difficult to understand, they tended to mention file storage structures, command syntax or system jargon rather than spatial or cartographic issues. While the researchers' failure to note serious spatial problems could be ascribed to their unfamiliarity with the data and tasks, the tendency for *users* not to mention such problems does not encourage us to believe they are significant in comparison to the HCI issues arising in GIS. However, the next chapter will explore the possibility that this may be because the system-centred approach apparent in much GIS design, as well as in HCI evaluation, precipitates a spurious emphasis on non-spatial activity irrelevant to users' true task goals (see also Medyckyj-Scott and Blades, 1990).

Of course, it is possible that the users in this study (1) were influenced unconsciously by the tendency for the evaluation materials and researchers' questions to focus on the HCI aspects of the systems, so that the users also focused their attention on these; (2) may have been a self-selecting group, in that people of low spatial aptitude or interest would not obtain jobs working with GIS (since most such jobs require an MSc degree in spatial analysis, or at least some formal cartographic training). Certainly there is abundant evidence that people differ in their performance on and attitudes to tasks which are considered to measure cognitive spatial abilities (e.g. Lohman et al, 1987), although there has always been debate about what these measures should be and how they should be interpreted (e.g. Eysenck, 1953; see also Chapter 5). The present study may have invariably selected users who were highly competent and motivated to manipulate digitally-displayed maps.

Users made various comments about the user interface aspects of their GIS. Although the different tasks and application areas of different users made it inevitable that their preferences would differ to some extent, there was some evidence that differences between them extended beyond this. One user particularly praised the menu structures within her GIS (while admitting that for some tasks she preferred to run macros from outside the GIS environment) and also the feedback alerts checking that a selected user action was really intended. However, at least two other users pointed out that such alerts could be useless as users generally clicked 'OK' without taking a second thought, and one claimed they were 'a real pain in the arse'. The notion of 'user-friendliness' aroused diametrically opposite emotions in different users, as with users of other IT systems.

A noticeable factor which may have affected users' responses was their attitude towards their organisation and system, and the aims of the study. More than one user behaved defensively to some extent, apparently feeling the need to justify or dismiss problems encountered with the software. On the other hand, two (female) users both appeared anxious about how their mistakes or problems would appear to the researchers, and whether their responses would be relayed to their superiors. Such political problems are an unfortunate flip-side to the increased ecological validity of performing studies in the workplace (as you gain the real world, you also take on any real-world desire for deception), and may well have contributed to the unexplained individual factors that had affected the Phase I survey.

2.2.2 *Performance measures*

2.2.2.1 Time-based variables

Variable	**Definition**	**Min**	**Mean**	**Max**
Total task time	time from user beginning to start up GIS to point where user or interviewer signals end of task demo.	9:05	23:42	42:28
Talk time	time (within the total task time) that the user spends talking without doing anything.	0:00	1:45	11:06
Talk percentage	talk time as percentage of total task time.	0%	7.6%	52.6%
Net task time	total task time - talk time	8:58	21:57	39:13
Snag time	time (during net task time) that the user 'spends dealing with snags which arise'*. Snags are when 'the user or the system completely negate the results of previous action(s)'*; i.e. time spent doing an action (and undoing it again, if necessary), with no productive result.	0:00	1:21	6:25
Search time	time (during net task time) that the user 'spends exploring the structure of a system without activating any of the structures'*. E.g. pulling down a pulldown menu and removing it without selecting an option.	0:00	0:11	0:57
Help time	time (during net task time) that the user 'spends making use of any help aids provided'* – includes seeking human advice, accessing online help, reading manuals/cue-cards.	0:00	0:09	1:25
Unproductive time	snag time + search time + help time	0:00	1:41	7:31
Productive percentage	percentage of net task time which is not unproductive time (NPL called it 'Productive Period'*).	59.8%	90.6%	100%
System process time	time (within the net task time) during which nothing happens other than the system processing (i.e. waiting time)	0:50	7:39	17:43
System percentage	system process time as percentage of net task time	9.3%	33.7%	65.7%
Major interface changes	number of times during the task demo that the interface style (method of input, colour scheme etc.) changed over all or most of the screen	0	7.1	30
Major change frequency	no of major interface changes per minute of net task time	0	0.3	1.6
Snag percentage	snag time as percentage of net task time	0%	8.1%	35.3%
Search percentage	search time as percentage of net task time	0%	0.8%	2.9%
Help percentage	help time as percentage of net task time	0%	0.6%	4.1%

* (NPL, Performance Metrics Directory, Document 2 in the MUSiC Performance Metrics Toolkit)

Notes:

1. Times expressed above as minutes & seconds
2. *n* was 20 for all timing variables; one user was not recorded on video.
3. Some variables are 'negative' measures of performance (e.g. snag percentage) while others are 'positive' (e.g. productive percentage).

Table 2-3. Time-based variables in the observation data (extracted from videotapes)

The amount of time spent waiting while the system was processing was generally large, averaging a third of the net task time, and reaching a maximum of nearly two-thirds. This finding fits in with the degree to which users in both phases of the USIS study pinpointed system response times as a major problem with many GIS. Time waiting for the system was still counted as 'productive' time in the definitions of the MUSiC metrics discussed below, but this enforced sluggishness was a major factor, wasting a large proportion of users' time (but not, anecdotally, appearing to cause most users any subjective frustration: they seemed to be accustomed to this pace). It could be that in the years since the study, GIS users may be using more powerful hardware which copes better with the software and data complexities, so hopefully this issue has improved.

In general, users ran into few problems while demonstrating their tasks. The use of the MUSiC metrics ensured a more comprehensive definition of 'problems' than simple error counts, but even so the mean 'productive percentage' (percentage of total task time not designated as 'snag', 'search' or 'help' activities) was 90.6%. However, some may argue that this is too low, since it implies that GIS users spend a tenth of their hands-on time unproductively fire-fighting. The lowest 'productive percentage' was only 59.8%; here the user was unproductive for 40% of the net task time, although of course this does not mean that this situation would still be true over longer periods. By contrast, three users worked with no hitches at all for 100% of the net task time.

Overall, the mean snag, search and help times were 81 secs, 11 secs, and 9 secs respectively, i.e. 8.1%, 0.8% and 0.6% of total task time, respectively. For most users, therefore, little time was spent either obtaining assistance or searching through the system for the necessary option. This should not surprise us, since the users were supposed to be demonstrating their *normal* tasks, and they would be unlikely to need much assistance in or show much uncertainty about tasks they performed on a regular basis. We cannot tell from this basic data whether the users would spend much more time 'searching' the system or obtaining help when performing a less familiar task.

The vast majority of the non-productive time of most users was accounted for by 'snag time', i.e. time spent doing, undoing or redoing actions which constituted 'snags'. The snags encountered by the users varied, but certain problems were encountered by several users despite working with markedly different GIS products. The most common type of snag, encountered by seven of the users, was a failure by the user to specify a name or sub-area of a map or file, which was often already visible on the screen, before attempting to perform an action that required it to have been explicitly selected. In these cases the apparent cause of the problem was often the users' failure to remember the need for an 'object-action' ordering, although occasionally the system was expecting the user to enter the specified name as a parameter to a typed command, and the user failed to do so. One may hypothesise that users can easily forget that what they are seeing on the screen is not understood in the same way by the system; it 'sees' a map only as a set of data, which must be referenced by a name or ID code, whereas the user sees and takes for granted an obviously visible object.

Similar to this problem was one encountered by five users, one of them making the error four times in succession before realising his mistake. These users attempted to perform a search, query or hardcopy plot of the *displayed* map while a different 'base map' area was currently *selected* (as far as the system was concerned), so the wrong result was obtained. Again, this occurred with users of systems which in other aspects displayed very different terminology and actions, ran on different hardware platforms and operating systems, and handled different data formats. Again, it appeared as if the users were guided in these actions more by the map area that was visible on the screen than by the parameters previously specified for the system to use.

Errors such as these may be related to the concept of an 'availability heuristic', first suggested by Tversky and Kahneman (1973) and more recently discussed by Sutherland (1992). Studies have suggested that visible entities, and also those more semantically important to a person, tend to be more 'available' for retrieval from memory than more abstract and less personally relevant entities; what is most easily retrieved then biases one's judgement, decisions and actions. It could be that the inherently visual, spatial nature of GIS tasks, far from causing a difficult problem for users in itself, may be so 'natural' or 'available' to them that they forget that the system still requires somewhat more abstract specifications and commands.

Obviously, the user's task goals are concerned directly with the map representation, not with the system's unintuitive demands and procedures; furthermore the visible data on the screen is inherently more memorable and concrete than the syntactic niceties of the command language or dialog box. These make it inevitable that the map on display will be the entity uppermost in the user's mind, and hence more 'available' as well as more visibly obvious. Such a bias may also be strengthened by (and cannot easily be empirically separated from) the 'picture superiority effect', a consistent finding in cognitive psychology since the 1960s (e.g. Paivio et al, 1968), that pictures of objects are easier to recall than their names[6].

Overall, it suggests that the actions required by the system are not entirely reconciled in the users' minds with the task they are trying to achieve: this issue of defining and describing digital map tasks will therefore be further investigated in the next chapter. The finding may be seen as ironic in the light of some authors' arguments, when pushing the case for incorporating virtual reality into GIS as early as possible, that people "do not connect with maps at a gut level" (Jacobson, 1995, p. 242). Experienced users, it would seem, may almost 'connect' too well: or at least, better with the map than with the computer displaying it.

2.2.2.2 Correlations among the time-based variables

Snag percentage correlated highly, and of course negatively, with the overall productive percentage (Spearman's r=-0.974, p<0.1%), although search and help percentages did not correlate significantly with this overall measure. This illustrates the dominant contribution of snag time to productivity problems in this study. One might expect snag, search and help times to correlate highly with each other if certain users were generally more 'problem-prone' than others. However, although search and help percentages were weakly correlated (Spearman's r=0.474, p<5%), i.e. users who

---

[6] The reasons behind this effect are much disputed, however, and have been suggested to be caused largely by the greater visual distinctiveness of pictures and by task requirements, rather than some difference in 'deeper' conceptual processing (Weldon & Coyote, 1996).

spent more time searching the system to find the required option also spent more time getting online, paper or human assistance, neither search nor help correlated significantly with snag percentage (so such users did not necessarily waste more time cancelling actions or errors). The low power of these correlation analyses make them inconclusive, however.

2.2.2.3 Action-based variables

Variable	**Definition**		**Min**	**Mean**	**Max**
No. of actions (Acno)	No. of discernible, non-combined, actions (by user or by system) seen on video during task demonstration session	All	63.0	140.1	355.0
		User	55.0	117.3	306.0
		System	4.0	26.6	65.0
No. of different actions (Acdf)	No. of (discernible, non-combined) actions which were different from others (i.e. not repetitions of another action)	All	40.0	68.2	156.0
		User	31.0	55.3	127.0
		System	4.0	12.9	29.0
Action ratio (Acra)	Ratio of different actions (Acdf) to all actions (Acno), expressed as a percentage (100xAcdf/Acno)	All	16.1	53.0	76.6
		User	15.7	51.6	76.9
		System	17.6	55.8	100.0
Action occurrences (Acoc)	Mean no. of occurrences per action (i.e. Acno/Acdf; inversely related to Acra)	All	1.3	2.1	6.2
		User	1.3	2.1	6.3
		System	1.0	2.2	5.7
Action times (Acti)	Mean duration of each action (Acno/Net task time)	All	5.7	10.8	20.7
		User	4.6	8.3	13.5
		System	9.7	21.3	74.5
Slowest action occurrences (Sloc)	No. of occurrences of the action that had the longest mean time	All	1.0	1.4	3.0
		User	1.0	1.7	4.0
		System	1.0	1.5	3.0
Slowest action time (Slti)	Longest (mean) duration of any action	All	31.0	102.7	285.0
		User	19.0	50.1	149.0
		System	22.0	91.3	285.0
Most frequent action occur's (Froc)	No. of occurrences of the action that occurred most often	All	3.0	17.1	87.0
		User	3.0	16.8	87.0
		System	1.0	9.8	40.0
Most frequent time (Frti)	Mean duration of the action that occurred most often	All	1.8	7.2	14.0
		User	1.8	5.8	14.0
		System	3.7	18.0	75.4

Notes:

1. Times expressed above as seconds
2. *n* was 20 for all action variables; one user was not recorded on video.
3. Where there was a tie for the 'most frequent' action, that which had the longest mean time was used; where two or more actions tied for the longest mean time, the one that occurred most often was used. Reasoning: this would highlight potential 'bottlenecks'.

Table 2-4. Action-based variables (extracted from videotapes)

The results showed a large degree of variability in the number of actions users performed. With a maximum value that was more than five times the minimum, and a standard deviation of 72.0, this was a far greater variability than that found in the duration of the task demonstration session. This was apparently due to the different speeds at which some users performed some actions (also reflected in the variations in the mean duration of actions). This was extremely fast in the case of users performing simple repetitive tasks over and over again, such as selecting a succession of polygon objects and issuing commands to alter them in some way. Nyerges (1993) suggested that speed of execution of spatial tasks may be one of the greatest factors differentiating amongst GIS users.)

This was partly backed up by a considerably smaller variability in the number of *different* actions performed. This is as would be expected if those users who performed the same actions repetitively also performed them faster than those demonstrating a variety of tasks. It is not surprising, then, that the ratio of [number of different actions] to [total number of actions] also showed wide variability, with a minimum of 16.1%, a maximum of 76.6%, and a standard deviation of 14.6%. This reflects the tendency for some users to perform simple, repetitive tasks throughout the task demonstration session. This was reasonable since the researchers had asked to see typical/normal work with the GIS, and previous studies had shown such tasks to be the norm in GIS use, e.g. Coleman et al (1992).

It was felt to be potentially useful to isolate the most frequently occurring action for each user, and the action that took the longest time (averaged over all its occurrences) to perform. Unsurprisingly, this 'slowest' action was often a period of system processing rather than an action performed by the user. The longest time thus recorded was 285 seconds, or four and three-quarter minutes! The mean 'slowest' single action time (across users and systems) was 102.7 seconds (well over a minute). The mean slowest *user* action time was 50.1 seconds. Considering the level at which individual actions were defined, this could still be seen as a long time. But these slowest actions tended to occur rarely, with a mean occurrence of only 1.4 and a maximum of 3 or 4.

For one user, the most frequently-performed action was performed 87 times during the task session (in this case, this involved pointing to a map feature to select it), while for others the most frequent action only occurred 3 times. The mean frequency of the most frequent action (system or user) was 17.1.

On average, this most frequent action performed took 7.2 seconds, dropping to 5.8 for the most frequent user action. This is surprisingly high: six seconds is a long time to take to perform a single low-level action, and certainly a long time for the one performed most often. One can imagine the tedium involved in repeating a 6-second action an average of 17 times over a 20 or 30 minute period, yet this may be the most common mode of use of GIS: despite their intended use for one-off high-powered analyses, in reality much of the work performed with these systems is (or was then) tedious, repetitive and slow.

2.2.2.4 Correlations among the action-based variables

The duration of the most frequent system action correlated negatively with the 'acoc' (mean no. of occurrences per action) variables (e.g. with 'acoc' for all actions, Spearman's r=-0.573, p<1%): thus where the system processing time tended to be longer for frequent system actions, actions tended to be repeated less, i.e. more different actions were performed. Perhaps the users whose tasks were more varied 'pushed' the system harder in terms of more complex processing which took longer.

The frequency of the most frequent system action correlated negatively but weakly with its mean duration (Spearman's r=-0.444, p=5%) – a situation which is clearly desirable, but which was not significantly demonstrated for the most frequent user or overall actions. The duration and frequency of the slowest system action were also weakly negatively correlated (Spearman's r=-0.453, p<5%). Given the desirability of minimising the occurrence of events that take a long time to perform, it may be unfortunate that there was *no* similar negative correlation between the duration and frequency of the slowest or the most frequent *user* actions (although once again we should add a caveat about the low power of these correlations).

Snag percentage was negatively related to the number of system actions (Spearman's r= -0.522, p<5%), the mean no. of occurrences per system action (r=-0.741, p<0.1%), and the amount of system repetition (r=-0.675, p=0.1%). This suggests that fewer errors seem to have been made when the system performed more of the actions, rather than the user (this is probably not surprising, since a macro combining a series of commands will generally avoid the minor slips and lapses inevitable in human performance).

2.2.3 Adapted Ravden and Johnson checklist

2.2.3.1 Sections 1 to 9: overall and 'total' scores

The graph below (Figure 2-2) shows the mean 'total' ratings obtained for each of the first 9 sections of the checklist, obtained by summing the scores on all items within each section (except the 'overall' ratings, which were on a different scale) and converting them into percentages of the highest possible score (i.e., if all questions had been answered with 'Always', the score would have been 100%, but if all questions had received the answer 'Never' it would have been 0%)[7].

For comparison, the 'overall' ratings (assigned by the evaluator on a 5-point scale) have also been converted into percentages and averaged, and are shown on the same graph.

---

[7] This transformation was used to facilitate interpretation of the data, since interpreting simple means would involve noting differences at 2-3 decimal places, and on a scale whose minimum was not zero.

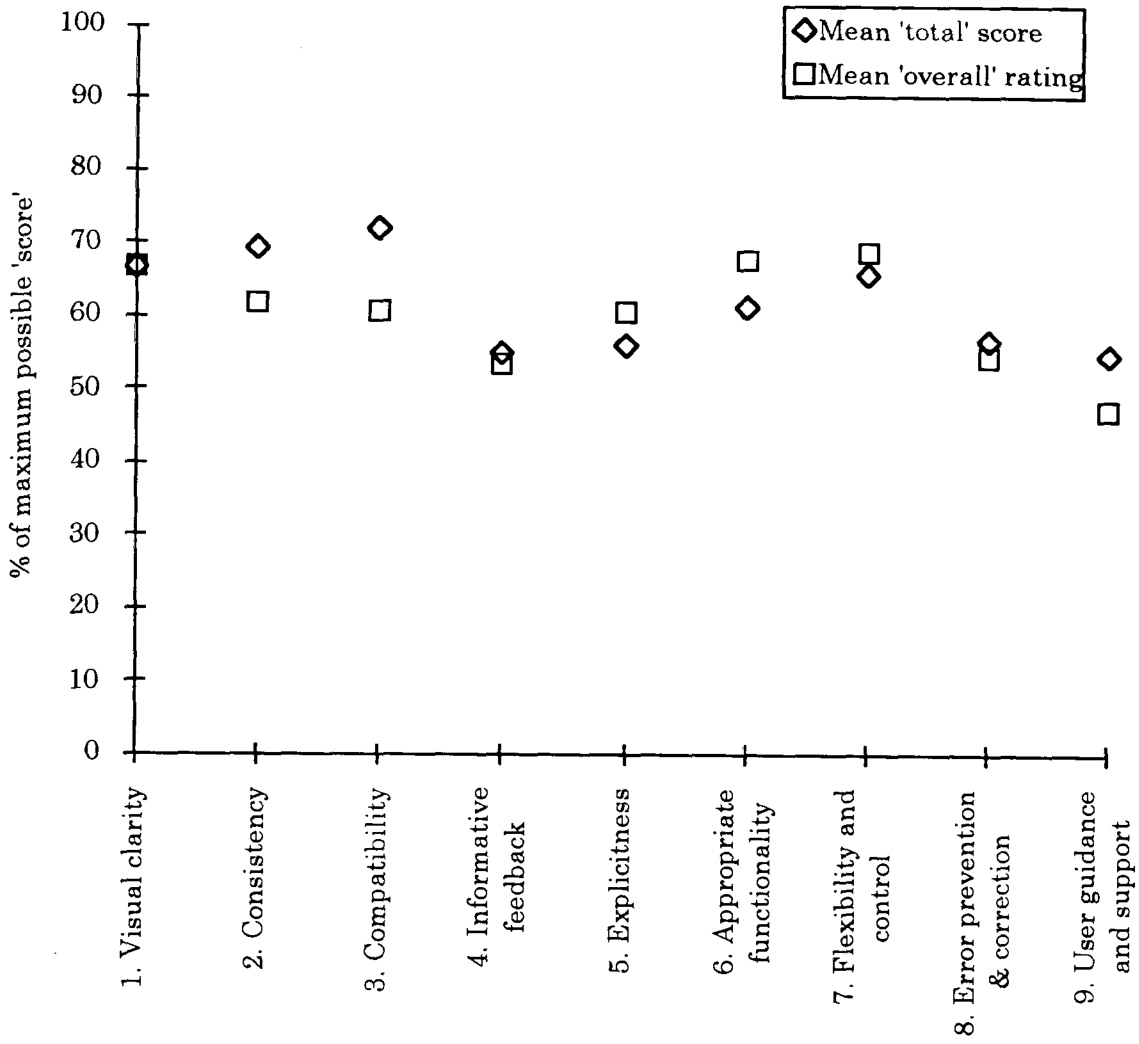


Figure 2-3. Graph showing mean overall/total scores on sections 1-9 of the Ravden and Johnson usability checklist

The most interesting aspect of these findings is their similarity with the overall findings of the Phase I survey (Davies and Medyckyj-Scott, 1993, page 35), in that error prevention and correction, feedback messages and user support, all scored relatively poorly. Also, some aspects which scored highly – flexibility and control, compatibility with users' expectations, and appropriate functionality – replicated findings for the 'Controllability', 'Predictability' and 'Suitability' categories in the earlier survey.

The 'total' and 'overall' scores did not coincide exactly, partly because the 'overall' ratings allowed the evaluator to take account of factors not covered by the individual items in each section. For example in section 9, if online help was not present or not used, it was meaningless to ask *how* it was accessed and implemented, so these items were often omitted during visits. However, lack of online help is in itself seen as a hindrance to usability by most experts, so the 'overall' rating was lower to reflect this.

Analyses were performed to correlate the 'total' scores with other variables. The results of such analyses have to be treated with caution: the Cronbach's alpha reliability measure for each section is shown in the table below (Table 2-5), and can be seen to be very low for certain aspects.

Checklist section	**α (reliability)**	**n of items**	**n of users**
1. Visual clarity	0.6019	16	18
2. Consistency	0.9114	14	20
3. Compatibility	0.6951	16	18
4. Informative feedback	0.8139	18	18
5. Explicitness	0.7241	13	10
6. Appropriate functionality	0.1013	11	10
7. Flexibility and control	0.5128	15	16
8. Error prevention and correction	0.7290	14	13
9. User support	0.8487	11	13

Table 2-5. Cronbach's alpha reliability coefficients for sections of the Ravden and Johnson usability checklist

2.2.3.2 Section 1: Visual Clarity

The results for this section are reported in more detail here because of the intuitive likelihood of its particular importance in the context of spatial data such as maps. Some of the questions in this section had surprisingly low scores. The percentage scores across all the visits, calculated as a proportion of the potential score if every visit had scored 'Always' for all questions, are shown in the table below (Table 2-6).

Question	n^8	%
1. Is each screen clearly identified with an informative title or description?	21	54.0
2. Is important information highlighted on the screen? (e.g. cursor position, instructions, errors)	21	54.0
3. When the user enters information on the screen, is it clear: (a) where it should be entered?	21	79.4
(b) in what format?	21	55.6
4. Where the user overtypes information on the screen, does the system clear the previous information, so that it does not get confused with the updated input?	19	84.2
5. Does information appear to be organised logically on the screen? (e.g. menus organised by probable sequence of selection, or alphabetically)	20	56.7
6. Are different types of information clearly separated from each other on the screen? (e.g. instructions, control options, data displays)	20	81.7
7. Where a large amount of information is displayed on one screen, is it clearly separated into sections on the screen?	21	87.3
8. Are columns of information clearly aligned on the screen? (e.g. columns of alphanumerics left-justified, columns of integers right-justified)	21	76.2
9. Are bright or light colours displayed on a dark background, and vice versa?	21	68.3
10. Does the use of colour help to make the displays clear?	21	68.3
11. Where colour is used, will all aspects of the display be easy to see if used on a monochrome or low resolution screen, or if the user is colour-blind?	21	63.5
12. Is the information on the screen easy to see and read?	21	65.1
13. Do screens appear uncluttered?	21	57.1
14. Are schematic and pictorial displays (e.g. figures and diagrams) clearly drawn and annotated?	21	54.0
15. Is it easy to find the required information on a screen?	21	66.7
Overall, how would you rate the system in terms of visual clarity?	21	66.7

Table 2-6. Mean scores on individual 'Visual clarity' questions in the Ravden and Johnson usability checklist, converted into percentages

The table shows that only about half the time were the displayed maps considered to be clearly identified, highlighted, drawn and annotated. Only about two-thirds of the time were they seen as easy to see and read, or sensibly coloured and contrasted. These findings, corroborating many comments in the cartographic literature bemoaning the lack of good design in many GIS map displays (e.g. Green, 1993), were the impetus for a further analysis of the maps using design principles collected

---

[8] Missing values, causing decreased *n*, occurred where questions were found to be irrelevant or unanswerable for a given system, e.g. where 'overtyping' was not possible or where menus were nonexistent (such as with command-line rather than menu/dialog-box interfaces).

from various cartographic and ergonomic studies, which will be described later in this chapter.

2.2.4 *Survey questionnaire responses*

The graph below (Figure 2-4) shows the pattern of mean responses (by usability category) received from the 18 observation study participants who completed the survey questionnaire (shown in full in Appendix 1); the mean responses from all earlier (Phase I) respondents are also shown for comparison. [Note: *n* was not always 18 for the scores shown, because most respondents tended to omit some questions.]

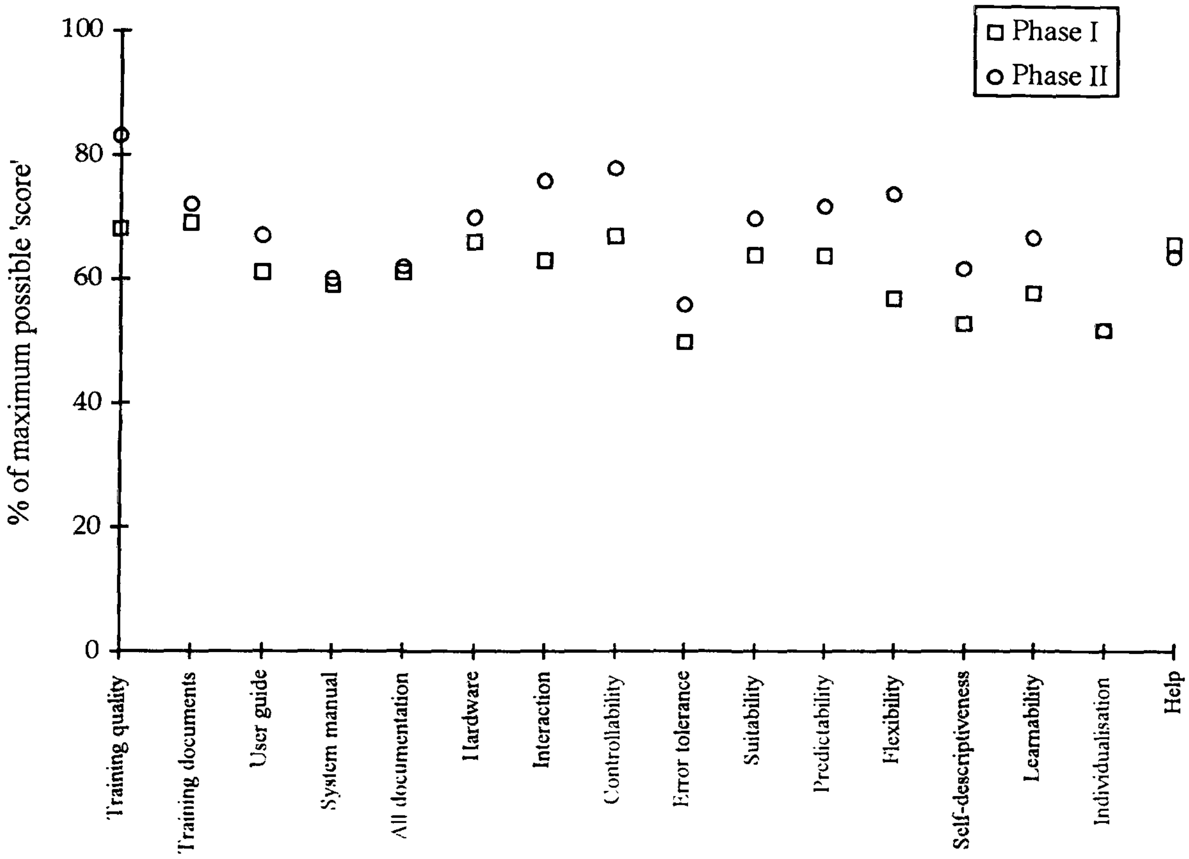


Figure 2-4. Graph comparing mean survey questionnaire scores for the earlier USIS postal survey and the later observation study

The graph indicates that the relative strong and weak points of the Phase II users' GIS, according to their subjective ratings, followed similar patterns to those of the Phase I respondents' systems, suggesting that the Phase II users comprised a reasonably representative sample of the overall GIS user population (assuming that the original survey sample can itself be assumed to have been representative: see Davies and Medyckyj-Scott, 1993). However, the graph also shows a general tendency for the ratings to be higher for the users visited in the observation study, and the next section discusses possible reasons for this in more detail.

Correlation analyses between the question categories shown in the above graph revealed fewer significant correlations than the same analyses did in Phase I. Nevertheless, certain patterns of relationships were found which reflected similar tendencies to those found in Phase I: notably, patterns of weak correlations between the Suitability and Predictability categories and other usability aspects (again suggesting a key role for measures reflecting cognitive compatibility of some sort).

2.2.4.1 'Old' versus 'new' respondents

The trend towards more positive responses in Phase II was investigated further through a series of comparative analyses between 'old' and 'new' respondents. Group comparisons between users who completed the questionnaire during Phase I (including those who were later visited in Phase II as well), and those who only completed it after being visited in Phase II, showed that the groups differed significantly in that the 'new' respondents were significantly more positive than the 'old' had been about training quality (Mann-Whitney U[9]=390.0, n=116 'old' + 11 'new', p<5%) and interaction (Mann-Whitney U=550.0, n=141 'old' + 12 'new', p<5%), and a number of individual survey questions.

However, chi-squared cross-tabulation analyses to check the characteristics of the two groups showed that the 'new' group were also more likely than the 'old' to be maintaining data for 'present' use (rather than analysing data collected in the past or performing future prediction tasks such as site selection; Pearson's test of χ^2, P=15.9, n=82 'old' + 16 'new', p<0.1%). The 'new' users were also more likely to be maintaining data that represented networks (of cables, pipelines etc.), followed in likelihood by cadastral data (e.g. property boundaries), whereas the 'old' group showed a greater trend towards the use of data representing vegetation, land use or other 'coverages' (Pearson's test of χ^2, P=11.4, n=108 'old' + 15 'new', p<5%).

---

[9] Unlike a parametric t-test, to the best of the author's knowledge unequal group sizes are not a problem for the Mann-Whitney U test, which explicitly addresses them (see e.g. Siegel, 1956).

This indicates some potential discrepancy in representativeness between the two groups; it also offers one explanation for the more positive survey ratings given by the 'new' users, since the results of Phase I had shown a tendency for users dealing with cadastral and network data, and performing largely up-to-date maintenance tasks with that data, to be generally more satisfied with various aspects of GIS usability. It is also possible that the 'new' and 'old' users were using different software or under different conditions – for example, 'new' users may have had more access to operating systems running a graphical user interface, facilitating certain aspects of interaction.

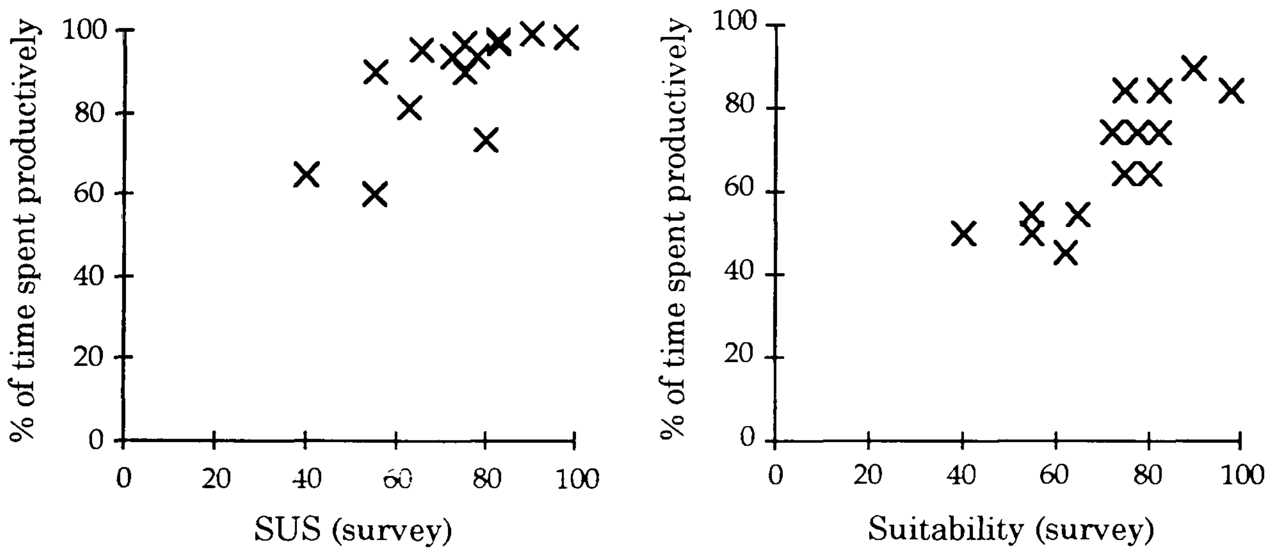
2.2.5 Relationships between variable types

Out of the many correlations run among the variables, only those which were significant, relevant, and convincing (when graphed) will be discussed here. (As stated earlier, some non-significant analysis results have since been lost).

The time users spent in talking during the videotaped interaction session (despite being asked not to) was recorded as a variable in its own right, and included in data analyses in case it showed any trends that could be related to user or system characteristics. The percentage of total task time that users spent talking to the researchers correlated weakly with the 'total' scores for sections 4 (Informative feedback: $r=-0.443$, $n=20$, $p=5\%$), 5 (Explicitness: $r=-0.462$, $n=20$, $p<5\%$) and 6 (Appropriate functionality: $r=-0.526$, $n=20$, $p<5\%$) of the checklist.

The simplest explanation for this group of findings is to assume that some users were so keen to ensure that the researchers understood their system and tasks, that they could not help explaining what was happening at times when they felt that this would not be clear from the GIS display. Such a situation arose more often for some users than for others, and these were the ones who felt compelled to explain what was going on. Other possible explanations would include the possibility that individual differences in user characteristics were reflected both in their natural talkativeness and in their tendency to feel that the system's user interface was neither explicit nor suitable for the tasks at hand; however, since the 'total' scores reflect the responses from both the evaluator and the users, this is not a sufficient explanation.

Another set of interesting findings consists of significant relationships between the productive percentage (based on the proportion of task time not spent dealing with snags, accessing help or searching for functions, as described above) and some of the subjective scales in both the Ravden and Johnson checklist (9 subscales) and the original survey questionnaire (17 subscales). Productive percentage correlated weakly with the 'totals' for section 1 (Visual clarity: r=0.446, n=20, p<5%) and section 3 (Compatibility: r=0.503, n=20, p<5%). The snag percentage, which as discussed above formed a major contribution to the overall proportion of non-productivity, also correlated weakly with the Compatibility 'total' (r=-0.447, n=20, p<5%). The correlations of the productive percentage with the SUS and 'Suitability' variables of the survey questionnaire were more impressive (Figure 2-5):



r=0.807, n=14, p<0.1%

r=0.861, n=16, p<0.1%

Figure 2-5. Graphs illustrating two strong correlations between subjective and objective usability measures in the USIS observation study

The right-hand graph, showing the 'Suitability' variable, reflects summated responses to statements such as 'I find the GIS unnecessarily complex', 'Tasks can be performed in a straightforward manner' and 'The screen always displays information in the way I expect'. Such items are obviously related to the cognitive compatibility between the user and system models, and so this is again strong evidence of the importance of this in successful GIS use. The strength of the two correlations is impressive, given that the survey was completed by the user at a separate time from the visit, and the productivity metric was derived from the researcher's analysis of the visit video while unaware of the user's survey scores.

Although these relationships are based on a small sample size and could be random 'glitches', a pattern emerges which, as in general HCI, emphasises the importance of the cognitive models on which the user and system base their behaviour during interaction. The visual clarity correlation suggests a link also to the importance of the visible maps, as discussed earlier. Overall, we have broad (though statistically weak) evidence that a suitable cognitive model underlying the interaction is an essential factor in successfully performing GIS tasks, and that the visual aspects of the digital map may feature significantly in this model.

However, a note of caution should be sounded: there were not as many significant correlations as expected between the checklist 'total' scores (explained earlier) and the survey usability categories (which are detailed in Appendix 1). Certain specific correlations would be expected since they were based on similar usability-related concepts: for instance, section 2 of the checklist (Consistency) would be expected to correlate well with Predictability; section 3 (Compatibility), section 5 (Explicitness) and section 6 (Appropriate functionality) with Suitability and Predictability; section 4 (Informative feedback) and section 8 (Error prevention and correction) with Error tolerance; section 7 (Flexibility and control) with Flexibility and Controllability, and section 9 (User guidance and support) with User guide, System manual, All documentation and Help. Yet *none* of these correlations in fact reached significance.

What happened? Various factors could be involved, including the low statistical power of the multiple correlations, the time lapse between the survey and visit for some users, the different circumstances in which the questions were asked, the different level of detail of the questions in the survey and checklist, and the fact that the checklist reflected ratings by the evaluator as well as the users themselves. It should also be borne in mind that every question included in the checklist and questionnaire asks about a slightly different aspect of usability, presumably on the assumption that some will be more relevant than others to any given system, and our attempts to link them together conceptually are fairly crude and with varying reliability. This does not detract from the general point that certain subjective problems were observed in actual second-by-second performance, and that these seemed in general to be connected to the suitability of the cognitive model of the system's visible and functional features. Correlations which *were* weakly significant included those of section 3 (Compatibility) with the general ease-of-use categories SUS (Spearman's r=0.538, n=14, p<5%) and Interaction (r=0.541, n=17, p<5%), reinforcing this point.

A few *negative* correlations also arose. Section 7 (Flexibility and Control) was negatively related to Training quality (r=-0.691, n=16, p<1% – see graph below, Figure 2-5), making it appear that the users who had had the best training were less able to choose their way of working with the system. Put another way, the users who knew the system functionality and were most aware of the various possible ways of achieving their goals had found their training *less* satisfactory, perhaps because it had failed to teach them enough about the system's potential, or because it had proved difficult to teach the conceptual basis of a system which allowed a variety of paths to the same goal. Some anecdotal evidence from users' comments suggested that training had indeed been less satisfactory for systems which were more complex and flexible. However, the possibility of this being due to a random Type I error, given the large number of correlations being run, cannot be ruled out, and the graph is not entirely unambiguous.

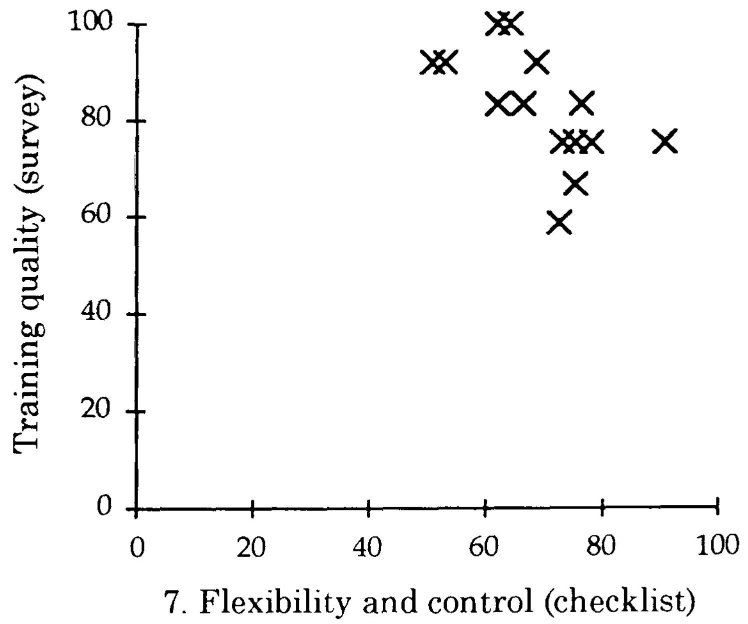


Figure 2-6. Graph of subjective measures of quality of training received survey questionnaire against system 'flexibility and control' (Ravden and Johnson checklist)

Finally, Individualisation was negatively related to section 5 (Explicitness: r=-0.892, n=8, p<1%). The low *n* should be noted here: as with the respondents in Phase I, few users were able to answer the Individualisation questions in the survey questionnaire. Those who were able to do so had a strong grasp of the functionality and may have performed some customisation themselves. There was some evidence (in Phase I as well as here) that the Individualisation category tended to be negatively correlated with other aspects of usability: it related more to the functional power of the system, which seems to have been strongest in the least usable GIS.

In addition, 'acraall' (ratio of 'different' actions to all actions, overall) correlated weakly and negatively with Error tolerance ($r=-0.487$, $n=17$, $p<5\%$), suggesting that the problem of error handling (shown in both phases of the project to be among the most serious weaknesses of GIS) is less severe for GIS users who perform relatively simple, repetitive tasks and who are therefore less likely to encounter unusual error situations.

2.2.6 Summary of findings and implications

1. There was little anecdotal or objective evidence of problems with the spatial or cartographic aspects of users' tasks. This was so even though extremely poor lighting conditions sometimes affected the visibility of the GIS displays, and despite great variations in the type and design of map displayed. Users also failed to mention spatially-related problems when asked about their most common mistakes and the worst aspects of their systems. However, some aspects of 'visual clarity' suggested poor design with regard to the digital maps displayed.

2. The most common (measurable) 'snags' involved errors whereby the users 'forgot' that data which was visible on screen or easily pictured in their minds was known to the system only as a virtually abstract entity that had to be correctly specified in order to perform system functions. This may suggest an 'availability' error in users' cognition during interaction, perhaps due to the immediacy and visual nature of the map and its greater relevance to the user's task goals.

3. Subjective measures from the checklist and questionnaire showed similar patterns of response to those found in the USIS Phase I survey, with error tolerance, feedback and user documentation again being the most poorly-rated areas. Similarly, ratings of the degree of difficulty regarding a list of usability problems

implicated the same aspects as in the survey, especially incompatibility between users' and systems' cognitive models.

4. Related to this, several relationships among variables indicated an apparently central role of cognitive compatibility to successful user acceptance, speed of working and avoidance of errors. There was some evidence that the visual aspects of the system were particularly relevant to this compatibility issue.
5. Users differed diametrically in their user interface preferences and desires. Users' specific situations, and their political attitudes regarding their organisation and system, and regarding the study itself, appeared to affect responses and may have similarly affected responses in the USIS Phase I survey.
6. System processing occupied on average around a third of the duration of the task demonstration sessions, illustrating the poor response times of many GIS. There was some evidence that slower system processing was related to users performing non-repetitive and perhaps more sophisticated tasks.
7. Users who encountered more 'snags' during interaction were not necessarily those who sought more help or spent more time searching uselessly through system functionality. There is thus no evidence for a generally 'problem-prone' user. Fewer snags occurred where the system, rather than the user, performed a greater proportion of the task.
8. Extreme variability in the number of individual actions performed by users appeared to be largely due to differences in speed of working, which was particularly high for users performing more repetitive tasks (who also spent longer demonstrating their work, and appeared more content with some aspects of usability). Such repetition appeared to be common among the users visited. It was not necessarily true that the most frequent actions were among the shortest in duration, although this would have been desirable.
9. There was some evidence that training courses had been less satisfactory where the GIS were more complex and flexible in their functionality. Functional power was indirectly linked to poor usability, as in the USIS Phase I findings.

10. Users' tendency to talk (despite being asked not to do so) was strongest where both they and the evaluator found the system's display to be poor at clearly illustrating what was occurring during interaction.

2.2.7 Further investigation

Both phases of the USIS work, and the subsequent analyses performed by the author, focused largely on 'traditional' HCI issues. This was partly due to the remit of the original project, and partly inevitable given the deliberate use of established system evaluation materials: the survey questions were based on guidelines in the ISO international software ergonomics standard IS 9241 – itself based on established HCI expertise – and the Ravden and Johnson checklist and MUSiC performance metrics were broadly based on the same body of knowledge. Within these limits, the work nevertheless managed to indicate a few cognitive aspects of digital map use which seemed worthy of further research:

- The apparently central role of cognitive 'compatibility' between user and system models of the task at hand, affecting both the productive and subjective aspects of usability, needed further clarification. The need for an integrated cognitive model of GIS use, incorporating individual differences in user cognition, was stressed by Turk (1992). However, other authors have suggested that this may be impossible for complex information systems, based on the limited current state of knowledge of users' actual cognition even with simpler systems (Grant and Mayes, 1991), and that we need a more systematic understanding of users' actual tasks, before we can identify the cognitive processes likely to be relevant to those tasks; a generic model would be too broad to be of use. This issue of understanding and analysing tasks will be explored in Chapter 3.

- Many snags in the observation study seemed to be due to apparent 'availability' errors, and the tendency for performance to relate to cognitive compatibility and visual clarity. It could be that the inherently spatial nature of GIS work, rather than creating a cognitive burden in itself, actually interferes with users' ability to cope with the complexities of the software because it is itself seductively task-relevant, and distracts from the computational interaction. It was felt it would be important and interesting to examine *which* visual aspects of the map may be significant in aiding users' tasks, both from a cartographic (see below) and a psychological (Chapter 4) perspective.

- The unexplained factors apparently affecting response patterns in the questionnaire survey may be due to individual differences in attitude, cognition and/or personality, which in turn may also affect users' actual use of GIS (a possibility which can currently only be supported by anecdotal evidence). Individual differences among users, and possible dimensions and constructs underlying them, will be considered in Chapter 5.

2.3 Cartographic quality and the observation study data

The work to date has failed to illuminate the supposed 'problem' of the spatial/ cartographic element of GIS tasks. This not only reflects the measurements used, but also a general dearth of evidence within the observation study videotapes of any real problems on the users' part in navigating round the displayed data. The users in general appeared to struggle not in becoming 'lost' within the *map*, but within the *system*. This may not be so surprising, given that some authors, e.g. Downs and Stea (1977), have argued that in fact humans are astonishingly *good* at spatial manipulation, interpretation and memory. Nevertheless, ratings of visual clarity did vary in the observation study, and appeared to be related to both subjective and performance measures of the user-system interaction. What is it about a digital map which may help or hinder users' performance and sense of comfort or ease? Do we already know, from previous studies? Can metrics or checklists for a map, measuring its quality or ergonomics, be derived in the same way as for a computer system?

This was investigated further in a collaborative project between the MRRL and cartographers in the Dept. of Geography at Loughborough University, funded internally at Loughborough and managed by this author, which examined the cartographic quality of the maps displayed in the observation study videotapes. This will be reported briefly below (although, being collaborative, it should be noted that only the data analysis *after* the videotape evaluations was solely the author's work, the rest having been collaborative although led by her[10]). As mentioned above, a more psychological approach, looking at how users mentally process the spatial representation of a map, will be taken in Chapters 4-7.

2.3.1 Cartographic quality: sources for checklist[11]

Although cartographers have been performing studies of map effectiveness and use for some thirty years, no straightforward objective checklist or measurement method has been devised for measuring what could be described as 'map usability'. In HCI, perhaps due to the involvement of computer scientists and the perceived need for prescriptive and evaluative tools, many such checklists and tools have been developed, but cartographers perceive their discipline to be as much a craft as a science, and have often been reluctant to produce generic guidelines that would restrict the individual map-maker.

With the advent of GIS, previously static maps have become ephemeral and editable objects with which the user can interact, as described in Chapter 1, and the different visual medium and task demands of these maps create a situation in which cartographically untrained people view displays which are not carefully designed products, but temporary visualisations of selected data layers.

Despite the availability of alternative ways of accessing and displaying the data held in a GIS database (e.g. alphanumeric searches, statistical summaries, report listings), the map is still the dominant method by which the user both interacts with the spatial data and views the results of that interaction. The USIS data, and other studies (e.g. Lee, 1995) show that the majority of GIS users do display and/or print maps in the course of using the system. Yet "despite this appreciation of the need for a map graphic for the display of the results of analysis, there seems to be little awareness of the need for the good design of these maps amongst GIS users" (Lee, 1995, p 34). Green (1993)

---

[10] Literature gathering and checklist derivation were a joint effort by the present author, David Medyckyj-Scott, Erica Milwain and Peter Robinson. Checklist formatted and completed from the videotapes by Erica Milwain and Peter Robinson, both qualified and experienced cartographers with some basic GIS knowledge. Data analysis by present author only.

[11] Thanks to David Medyckyj-Scott for helping to develop the ideas in this subsection.

also commented on the fact that little thought appears to be given to the map, and listed poor layout and poor choice of colour, font and symbols as common characteristics of digital maps. He argued that the maps produced as a result of some query or analysis may thus often fail to state the message in clear cartographic language, and may as a consequence affect decision-making performance. They may even convey false impressions about the facts contained in the displayed data.

It was observed as early as 1979 that map design must be reconsidered in the context of digital maps (Anderson and Shapiro, 1979). More recently, Visvalingam (1992) wrote

> For GIS to be effective the geographic data upon which the analysis is based relies on accurate spatial representation. Such accuracies, while not absolute, will often be determined by the requirements of the spatial analysis and not the subsequent need for cartographic integrity when displaying results. However, to efficiently communicate the results of the analysis, sound cartographic principles have to be applied to the maps. As a consequence, GIS map output should embody the principles of generalisation and cartographic design. [p. 51]

Both Lee and Green suggested that the problem of poorly designed digital maps is the result of the use of GIS by individuals with limited, if any, knowledge of cartography. Even if they do have some basic cartographic knowledge, many GIS provide no more than the basic graphic design tools and control mechanisms to allow the user to produce effective maps. Other GIS do include more comprehensive graphic design tools, but these can be difficult to use and much of the functionality difficult to access. In defence of the vendors it needs to be noted that, while some of the mapping functions required are relatively straightforward to implement, others are much more complex. Yet even the simplest of cartographic rules were rarely being incorporated into GIS at the time of the USIS study: e.g. the default values for mapping functions and symbol tables could have followed general cartographic practice, but often appeared to have been decided on a computer programmer's whim (as one or two of the visited users remarked).

To evaluate the USIS maps from a cartographic perspective, the author and colleagues tried to uncover a suitable existing checklist or set of guidelines for cartographic quality in digital maps. No such checklist had been published. Hopkin and Taylor (1979) provided a human factors checklist for aviation maps, but at quite a high and general level, and primarily aimed at human factors principles. One other checklist formed part of a design guide for environmental maps (including risk maps) developed for the New Jersey Department of Environmental Protection and Energy (Monmonier, 1994). It covered data, design variables, format, and communication effectiveness, but had not been published and was too specific for the variety of map applications seen in USIS. An extensive literature review was therefore undertaken, to gather relevant experimental results and expert views which could be incorporated into a checklist for us to apply.

Worth (1992) commented that "Although cartography and the core concepts of conventional static map design have probably existed for 2,000 years, it is amazing how difficult it is to find them listed somewhere". This was certainly the case in the present study, which entailed extensive literature searches. While many papers and texts bore titles *suggesting* they aimed to define key mapping conventions for digital maps, often the papers did not actually do this.

It was felt important to try to identify conventions and recommendations that either applied specifically to the digital map context, or were clearly universally applicable to map design in any context. This meant excluding some authors' recommendations about background colour, where these were made with paper maps in mind rather than the perceptual or technical demands of cathode-ray or LCD screens. Recommendations from the ergonomics/human-computer interaction literature were included as well as those from cartography - largely because the cartographic literature on digital map design principles was so sparse - but again only where the context of use was cartographic maps rather than other types of information display.

Another area which was largely excluded concerned recommendations for statistical maps (maps depicting the relative distributions of phenomena such as unemployment or disease). These tend to be more commonly produced for presentation on paper than for on-screen work. Statistical maps were not produced at all within the USIS study (Davies & Medyckyj-Scott, 1996).

As with any literature review, the process of identifying and extracting criteria for the checklist may have been subject to some subjective bias. However, the process was performed by four researchers, meeting regularly to discuss and revise the lists of criteria they had unearthed. It was hoped that idiosyncrasies were thereby kept to a minimum. Another potential source of bias, and one harder to overcome, was access to copies of papers: where it was not possible to locate a known paper or book through accessible British library sources, the source could not be scanned for recommendations or conventions. The table below (Table 2-7) lists some of the major sources of criteria for the checklist, and illustrates the various backgrounds from which those sources were drawn.

Subject area	**Author(s)**	**Context**	**Recommendations used in the checklist**
Cartography	Bernhardsen, 1992	Cartography in GIS	Zooming; system limitations on design; indicating inaccuracy; symbols; indicating data sets
Cartography	Brown, 1993	Cartography in GIS (considering display hardware)	Screen flicker; WYSIWYG; foreground/ background colours; patterns; lettering
Cartography	Cuff and Mattson, 1982	Thematic map design	Lettering
Cartography	Keates, 1989	General cartography	Accuracy; scale; contrast; symbols; colour; patterns; lettering; generalisation
Cartography	Makkonen and Sainio, 1991	Cartography in GIS	Colour
Cartography	Monmonier, 1993	Cartography general and in GIS	Patterns
Cartography	Robinson et al, 1984	General cartography	System limitations on design; accuracy; scale; legend; indicating data sets; patterns; lettering; generalisation
Cartography	Worth, 1988	Cartography in GIS	Legend (not needing to refer to it)
Ergonomics	Luria et al, 1986	Colour displays	Colour
Ergonomics	Shneiderman, 1992	General software ergonomics	Response speeds
Ergonomics	Young and Miller, 1991	Colour displays, varying viewing distance (displayed plan maps)	Symbol sizes; colour
Computer graphics	Kirsch, 1994	General review of computer graphics	Presentation
Computer graphics	Ware, 1988	Colour displays of univariate maps	Colour
Computer graphics	Widdel, 1990	Colour displays (esp. maps)	Subjective checklist (aesthetics, potency, quality)
Military	Davis and Swezey, 1983	Land battle map displays	Information organisation; highlighting; symbols; colour
Military	Knapp and Moses, 1982	Land battle map displays	Highlighting
Military	Knapp et al, 1982	Land battle map displays	Panning; symbols; colour
Military	Potash, 1977	Map design (esp. for use in the field)	Colour
Military	Moses and Maisano, 1978	Route selection	Panning
Aviation	Remington and Williams, 1986	Helicopter map displays	Symbols
Aviation	Spiker et al, 1986	Helicopter map displays	Colour
Maritime	Eaton, 1993	Factors considered for maritime navigation system	Screen brightness; zooming; clutter

Table 2-7. Major sources of checklist criteria for digital map design

2.3.2 Constructing and using the checklist

The methodological processes of constructing, piloting and coding the checklist will not be described here, since they formed a lengthy collaborative piece of work (which, although written up, has unfortunately remained unpublished due to the authors all leaving their posts at Loughborough). The content and structure of the checklist will be briefly summarised; the full version of it is given in Appendix 2.

After peer review and piloting had caused some revisions of content and wording, the version of the checklist finally used in the evaluation consisted of the sections shown in the table below (Table 2-8). No attempt was made to restrict the number of questions in any section, or to judge which aspects deserved closer attention than others. Instead, the distribution of questions directly reflected the amount of attention and hence the number of recommendations made by authors in the literature. This is itself an interesting phenomenon: for instance, there had been much greater focus on symbology, particularly in military-based research, than on name labels or scale.

Checklist section			**No. of items**
Part I: Context	1. General		12
Part II: System	2. Screen/monitor		3
	3. System usability:	Response times	6
		Information organisation	11
		Navigation	12
		Interaction	6
		Highlighting	12
		Hard copy	1
Part III: Map/task	4. Description of map(s) and reason for choice		2
	5. Task suitability:	General	7
		Information availability	2
	6. Cartographic variables:	General	8
		Presentation/layout	3
		Supporting information: accuracy	5
		Supporting info: scale, orientation	9
		Supporting info: legend	5
		Map projection	1
		Symbols	39
		Colour: general	8
		Colour: background	10
		Colour: strong colours	8
		Colour: light colours	16
		Patterns	10
		Names	19
		Generalisation	7
Total			232

Table 2-8. Sections of the cartographic design checklist

After completion, the first draft of the checklist was sent for peer review comments to selected individuals who had previously written about cartographic issues in GIS mapping[12]. The checklist was then revised, and two randomly-selected videotapes were used to pilot its use within the planned evaluation context. After further minor revisions, based on the experience of the two evaluators within the pilot, the videotape evaluation began.

2.3.3 Using the checklist to evaluate videotaped map displays

2.3.3.1 Materials and method

As explained earlier, each videotape included approximately 20 minutes of recorded 'normal' work by the GIS user under study, who had been asked to work on typical tasks (preferably those that would have been performed had the researchers not visited). One of the two qualified cartographers in the team viewed the recorded interaction on each videotape and completed a copy of the checklist. For two of the tapes, chosen at random, both evaluators separately completed the checklist in order to evaluate cross-rater consistency.

A number of methodological issues arose from the experience of applying the checklist:

- The **confusion in focus** between the visible map, and the system functionality which supported it; the evaluators found themselves both making judgements of the visible screen and checking through the video for evidence of supporting system functionality. For this reason, we concluded that a future version of the checklist should be split and deal with system-related and map-related questions separately. Map-related sections could be used to answer Green's (1993) suggestion that most maps produced by GIS tend to be poor due to lack of skilled cartographic input. System-related sections could contribute towards answering his and Wood's (1993) additional assertion that software was often not intended for cartographic presentation and thus does not facilitate good design.

---

[12] Comments on either the general study or the specific checklist were obtained from John Lee, Michael Wood, Mark Monmonier, Alan MacEachren, Mahes Visvalingam, Robert Taylor and Chris Board, all of whom had published papers and/or books on evaluative aspects of cartography. The author and her colleagues are grateful to them for their help.

Nevertheless, the need to consider both aspects must continue to be stressed when evaluating the design and usability of digital maps.

- **Balance and length of checklist**: the evaluators found that for most tapes approximately two hours were required to work through the checklist, and that extensive winding and rewinding was necessary to find portions of recordings that showed different aspects of the digital maps and of users' interaction with them. It became clear that for the checklist to be practically applicable in future map use settings, some items would be candidates for deletion: those which were difficult to answer, failed to discriminate between maps since they were almost bound to be present or absent, or created unreliability between evaluators due to ambiguity or controversy (see 3). In total, 172 questions were given the same response for more than half (i.e. 11 or more) of the tapes, leaving only 40 evaluative questions which showed real variety among the systems and maps under scrutiny (but see below).

- **Inter-rater response reliability**: for two of the twenty videotapes (randomly selected), both evaluators made a separate evaluation (time did not permit duplication of the evaluation for all tapes, but it was felt that these two would suffice to indicate the least reliable questions in the checklist). In total, of the 220 Part II and Part III questions (the evaluation questions about system functionality, map design and task suitability), the two evaluators gave different responses on 45 questions for one videotape (tape A) and on 50 for the other (tape B). (In other analyses below, only one set of results was used, that of the first evaluator to analyse each of the two tapes; this happened to be a different evaluator for tape A than for tape B.) Responses which were within one point of each other (e.g. "Always" versus "Mostly") could arguably be disregarded. This reduces the number of 'significant' differences on Tape A to 42, and on Tape B to 44. Reasons for these differences are listed in Table 2-9 below, along with ways the team identified for improving consistency in future studies.

Reason	**Ways to improve consistency**
Different interpretations of question	Reword to reduce ambiguity
Different degrees of understanding how system works	Ensure evaluators are specifically system-literate
Basing response on different portion of video	Agree part to evaluate; evaluate 'live' not from video
Subjective judgement differences (e.g. 'easy', 'clear')	Reword to indicate approximate standard expected
Different amounts of effort to find answers	Prior guidance on time to be spent on each question?

- ***Table 2-9. Causes of inter-rate inconsistency in applying the map design checklist***

- **Mode of usage of checklist**: The situation in which the checklist was used in the present study was prone to some obvious problems. Any aspects of the map or system that had not been captured clearly on tape could not be effectively evaluated, although it could be argued that these were also aspects which would not be much easier to evaluate in a 'live' evaluation where a GIS user was present. The temporal nature of the video recording necessitated moving backwards and forwards through the tapes to find answers to questions which, in a 'live' evaluation, could be answered by directly manipulating the system or asking the user. Finally, in a 'live' evaluation the evaluators would be present (they had not taken part in the USIS visits), and would have a richer experience of the system and the user interview than could be obtained from a videotape.

2.3.3.2 Results

As stated above, there was a large number of items which failed to distinguish sufficiently between different maps to be of much use in rating them. However, for some of these, the response given tended not to be a null or neutral response, but 'No/never' (implying the absence of a desirable property). These could be seen as representing the least frequently applied design principles recommended in the source literature; those which received a 'No/never' response for at least 25% of the tapes (and were considered potentially relevant to the users' applications, by the team of evaluators) are shown in Table 2-10 below.

Checklist item	**n of 'No/never' responses**
Is there a scale bar?	18
Can the user easily ascertain where s/he is on the map? (e.g. use of 'viewports'?)	14
Where the user's attention is to be drawn to a particular point on a display filled with information, is a moving or flashing cursor used?	13
Has the user had sufficient cartographic training to perform his/her task?	13
Is there an indication of orientation (of the map)?	13
If colour coding is used, does it avoid possible confusion by red-green colour-blind users?	13
Is all information required for the task present in the work area (screen) at all times (e.g. legend)?	12
Is black used as a background colour where the screen map does not need to match a paper map?	12
Is the speed of zooming in sufficient? (less than 4 seconds)	11
Does the layout avoid windows overlapping the map?	10
Are patterns (textures) used?	10
Are special areas of the display distinguished using either brightness/contrast, colour coding or a surrounding box?	9
Are high priority messages and codes highlighted using either brightness/contrast, different character sizes, colour coding or a surrounding box?	9
Is white used as the background colour where the screen map has to match the paper map?	9
If a hard copy is printed or plotted does the map on the screen look the same as the hard copy (WYSIWYG)?	8
If different layers of information can be selected for display, are they clearly named?	7
Are windows tiled to avoid overlapping one another?	7
Are command or data entry errors highlighted by brightness/contrast, different character sizes, or colour control?	7
Where specific information has been changed or is about to be changed, is this highlighted using either brightness/contrast, different character size or colour coding?	7
Is highlighting of any kind restricted to only one or two types of object at all times?	7
Is there any clear indication of the order of selected information?	6
Where related information is in separate windows, is that relationship obvious?	5
Are unusual values highlighted using either brightness/contrast, different character size or colour coding?	5
Is the map on the screen in a form familiar to the user?	5
Are the colours used aesthetically pleasing?	5
If the user has to search and identify objects, are the most effective colours used for the most important 'targets' (i.e. red is the most effective, then blue, yellow, green, black and white)?	5

Table 2-10. Map design principles apparently not followed for digital maps observed in the USIS observation study

Perhaps the most surprising issue, for anyone trained in cartography, is the apparent disinterest of GIS users in traditionally vital cartographic components such as a scale bar, legend and orientation indicator (a north arrow). This contradicts the findings of Lee (1995), whose postal survey of GIS users found that at least two thirds of the respondents claimed to include a scale bar and/or legend in their maps. This can be explained in a number of ways, e.g.: (1) Lee's survey respondents may have been thinking of maps they *produced* on paper or for on-screen *presentation*, rather than the day-to-day working displays; (2) most respondents with any cartographic training would know that such components are generally deemed necessary, and could have felt obliged to state that they correctly included them. Lee himself, however, pointed out that in working digital maps, with an effective audience of only one person viewing the map, such items may not be necessary.

Against the issues highlighted in the above table should be set the fact that in all 20 tapes, the evaluators answered 'Yes/Always' to questions as to whether the GIS and the map scale used were appropriate to the user's tasks, and whether a medium blue background (considered the worst for digital maps) was avoided. The map representation and background colour were deemed appropriate to the task overall in 19 of the 20 tapes. In 18 of them, the screen, colours used and layer selection capabilities were considered to be suitable, and users had been suitably trained in the specific tasks and had sufficient 'domain' knowledge. None of this should be too surprising, given that these were mostly experienced daily users. The two cartographers who performed the evaluation wrote comments such as "visually disturbing", "resolution extremely poor when zoomed in", "white cursor when not moving blends in with all other white lines", "house symbols virtually fill the screen when zoomed in", "not very elegant"; but they generally felt that despite clutter, aesthetic ugliness and strange colour/font choices, the maps mostly appeared "just about" clear enough to use.

Many of the highly specific recommendations included in the checklist, such as issues of whether panning was continuous, whether symbols subtended 8mm and were proportionally sized, whether redundant coding was used and whether letters were used in labels in preference to numbers, received null answers for every one of the tapes, either because the evaluators were unable to discern the answers or because they were deemed irrelevant to the display under scrutiny. Since many of these precise recommendations had been derived from military studies, in which rapid target identification within highly-specific battlefield displays is often crucial, perhaps it is hardly surprising that they seemed irrelevant to the typical Ordnance Survey-based street map displayed in a local council planning office.

Overall, then, the attempt to apply generically prescriptive map design knowledge to the map displays recorded on the videotapes was of limited success: it appeared that many of the recommendations, despite being touted as broadly applicable by their authors, could not be sensibly applied to the common GIS application contexts of local authorities, environment agencies, utilities and commercial companies. This was as true for some detailed laboratory-based ergonomic studies as for military, aviation or maritime findings. Above all, the interviews with the users themselves demonstrated an apparent lack of concern with map design issues, and the analysis above of their errors and problems showed little performance detriment in map-related actions.

These findings, and the earlier 'Visual clarity' ratings from the Ravden and Johnson checklist, create a paradox: on the one hand, there was some evidence that overall visual clarity was related to usability and to effective (snag-free) user performance, but in general was not a complaint for the USIS participants, but on the other hand some aspects within both checklists scored quite poorly despite having been identified by ergonomists and/or cartographers as key recommendations for good design. It is clear that the situation is more complex than "does the map design matter or not?": subsequent chapters (especially chapters 3, 4 and 6) will look more closely at the impact of the specific task on how map users encode and interpret the visual information. They will also examine the relevance of users' prior knowledge, abilities and expectations in overcoming any deficiencies in the clarity of the map representation.

2.4 Summary

Analysing the USIS data from both the HCI and cartographic point of view suggested limitations on the ability of generic prescriptions, both ergonomic and cartographic, to relate to users' real-life problems and needs. The importance of considering the task context became obvious, in different ways, in both the analyses described above: the usability metrics suggested a central role for cognitive compatibility between task and system which appeared to affect both user satisfaction and objective productivity. The cartographic study suggested that many prescriptions about 'good' cartographic design may be simply irrelevant for on-screen, working, digital maps, and that even more specific recommendations can be irrelevant if aimed at military battlefields instead of local council offices.

We need to go beyond these prescriptions and to look more closely at digital map interpretation and use, not least because this is more likely to explain at least what *should* happen if GIS were better designed so that their computational procedures did not interfere with the real task. The nature of this 'real task' and its component activities will be examined in the next chapter; subsequent chapters will focus on the cognitive processes involved and on factors which may influence them.

Chapter 3: Working with digital maps: tasks and issues

The interactions between a user and a digital map have the potential to be highly complex, with many sophisticated interactions and analyses which cannot be attempted with ordinary paper maps. However, the USIS Project found that for most GIS users, daily operations are restricted to basic tasks such as zooming and panning, viewing and printing/plotting, digitising objects and adding 'attribute' information to them, and searching for specific information or for digitising errors. Thus the usability problems reviewed in the last chapter occur even though people are using digital maps at a fairly unsophisticated level (in terms of the system's analytical capabilities). Nevertheless, users must presumably still invoke some 'expert' knowledge of the system, the map it displays, and the real world the map represents. As shown in Chapter 2, this knowledge apparently allows users to interact successfully with the digital map, generally with few problems or errors, despite the unusual characteristics of its 'space'.

This chapter will examine the actual tasks that digital map users perform, using both HCI and cartography perspectives, to demonstrate the different ways in which they can be understood. In particular, the chapter will seek to show the distinctive characteristics of even basic GIS work, as opposed to most text-based computer work, and will consider the user's interaction not just with the commands and menus of the *system*, but with the *information* contained within the map. The latter point leads into subsequent chapters, which will move away from a 'system and function' perspective to examine the cognitive processes of understanding the displayed map information. The initial focus on real-world tasks and applications will hopefully enable the reader to relate the later, more academic, analyses back to practical reality.

Why take this approach to digital maps? One answer is that nobody else has done so. Typically, a leading researcher at the 1994 NATO Advanced Research Workshop was heard to state that he was interested in the use of GIS to facilitate visualisation and decision support for geographers and professional planners, and was "not interested in the technicians". Indeed, most GIS research either performed or proposed to date has been aimed at geographic visualisation, complex analyses and spatial decision-making, not at more basic handling of geographic data. Reginald Golledge (Golledge, 1995) has argued that we must understand the errors people make in their use and understanding of basic spatial concepts or 'primitives', since these errors may be propagated in handling more complex spatial constructs, and as mentioned in Chapter 1 the identification of these 'primitives' was discussed at the 1994 NATO ARW. However, even Golledge's paper and that discussion were ultimately concerned with the uses of GIS for visualisation, rather than basic spatial information retrieval or information management.

From the point of view of funding bodies, and of researchers whose primary interest stems from a geographic background, this focus on advanced visualisation is inevitable and appropriate. It is, after all, this visualisation and modelling capability that sets GIS apart from cheaper digital mapping systems (Frank, 1993), and provides their potential for decision support. However, we seem to know little enough about the cognitive activities involved in basic usage and reading of paper-based maps, let alone digital maps that permit a whole new dimension of interaction and control. It would seem that progress in understanding cognition in this area needs to begin with the frequent, essential, everyday level at which users tend to interact with digital maps, regardless of any intention to view the data at some point from a more analytical perspective. It would also seem likely that this level of interaction and understanding is the one most common to all application areas, and which is most likely to be offered to an even wider audience in any publicly-accessible map-based information system.

Having made this argument, it seems pertinent to admit that there is no clear dividing line between the 'everyday' and the 'advanced' usage of GIS. One user visited by the USIS researchers spent most of his time digitising electricity cables, but also ran frequent network analyses to check that all his cable data had been successfully and appropriately connected. A network analysis is one of the more advanced, powerful features offered by a GIS, but this user was treating it as a trivial data validation activity, since he had only to select a single command for the system to perform it for him. Similarly, in a well-constructed, highly usable mapping system for public access, such as in a travel agency or route-finding system, it is easy to imagine a one-click implementation of sophisticated GIS-level functions such as finding the shortest route between two towns, or spatial correlation to identify easily-climbed hills accessible from public transport stops.

The point to realise here is that even the most sophisticated *functions* within GIS may be reduced to simple mouse-clicks in a system designed for simplicity and usability, yet the actions performed by *users* are probably always based on similar cognitive constructs and processes, since these depend upon being able to understand and manipulate the map by using the GIS functions. Nevertheless, it is also natural to ask what characteristics of a given digital mapping *task* may sometimes cause it to require different or greater cognitive effort by a GIS user. The need to identify such characteristics was strongly emphasised at the 1994 NATO ARW (Nyerges et al, 1995).

The remainder of this chapter will describe the following approaches to understanding digital map users' tasks:

- data from the USIS survey and observation study, analysed according to various distinctions developed in the HCI domain;
- a reanalysis of some of the same data, developing concepts of 'work' and 'enabling', i.e. deciding the purpose of individual actions and deriving some measures intended to reflect some kind of productivity;
- review of digital map tasks as viewed in the cartographic literature, drawing links between concepts and distinctions and attempting to bridge the gap between the HCI and cartography approaches.

3.1 Tasks in the USIS survey and observation study

The work described in this section has been reported in detail elsewhere (Davies, 1995); only the aspects and findings relevant to the present thesis will be described here.

It had been suggested by previous authors (e.g. Nyerges, 1993; Mark and Frank, 1990) that producing a taxonomy of digital map-related tasks could provide a structure for efforts to improve GIS design and training. In the absence of any data, other than the USIS work, initial thoughts about such a taxonomy (e.g. Albrecht, 1994) were largely speculative or based on what *systems* could do rather than what users *did* do. The present author therefore conducted an analysis of the USIS data, from both the survey and observation study, in order to summarise the common activities which digital map users performed[13].

3.1.1 Survey

As part of the wider attempt to produce formal methods and analysis schemes for usable system design and evaluation, task analysis formed a significant strand of HCI research in the late 1980s and early 1990s (Diaper, 1989). Much of this work, however, assumed that an independent evaluator would be assessing users' tasks in detail, rather than depending on their reports. However, in the present case the only direct task data available is from the observation study, where less than two dozen users' tasks were sampled (each on a single day, for only around 20 minutes each) in the workplace observation study. Therefore this data was supplemented by first re-examining the earlier USIS survey data, in which 152 respondents wrote free-text answers to the open question "What tasks do you currently do with your GIS?"

Since the answers to such a deliberately open question were likely to vary in level or granularity of description, and since they were also being set alongside the checklist and video evidence from the observation study which were likely to differ again in granularity, Rasmussen's (1986; Rasmussen et al, 1994) five-level 'Means-Ends Abstraction Hierarchy' for task descriptions was adopted to aid classification.

This scheme allows tasks to be considered at levels varying from 'Physical form' at the bottom (what actually happens in the physical world) to 'Purpose' (organisational and high-level objectives). These levels were interpreted by the author based on

---

[13] At least, in 1992/3 in Europe and Australia: there is no obvious reason why these should have changed since then, or should differ substantially elsewhere in the developed world.

Rasmussen's descriptions as shown below, and they are given here along with an example from a typical GIS application domain.

Purpose: the reasons for the system's design or purchase, in terms of organisational mission and goals; e.g. *Regional water and sewage distribution.*

Abstract function: priorities and goals relevant to the system's more specific purpose within the organisation; e.g. *Running and maintaining pipeline network.*

Generic function: descriptions of the 'functions' within the system, irrespective of how they are performed; e.g. *Data storage and updating, producing fault reports.*

Physical function: actual activities of the user/system (in terms of how they control the system and what they do to perform the generic functions); e.g. *Digitise pipeline features, search and edit attribute records, print out report.*

Physical form: the appearance, location and physical movement relevant to particular actions; e.g. *Drag digitising tablet to draw line; type name.*

The reader will notice that, as Rasmussen et al discussed (1994, pp. 41-42), each higher level acts to answer 'Why' something is done, with the level below it describing 'What' is done in some way, and in turn the level below that describes 'How' it is done. Thus the top level is entirely concerned with 'Why' (so far as analysing a system is concerned: it does not explain what the system or user does to contribute to the goals, nor how they do it), and the lowest one entirely with 'How' (i.e. descriptions at that level cannot tell us what the actions are leading to or why they are occurring).

After analysing the responses as discussed below, the levels of description which emerged to a greater or lesser extent from the survey, the observation study checklist and the observation study video data are shown in Table 3-1.

	Survey responses	**Task frequency checklist**	**Video of work session**
Purpose	✓		✓
Abstract function	✓		✓
Generic function	✓	✓	✓
Physical function	✓	✓	✓
Physical form	✓		✓

Table 3-1. Levels of description in the Rasmussen 'Means-Ends Abstraction Hierarchy' applicable to data sources from the USIS studies

As can be seen, despite overlaps, the task descriptions given in the survey were generally not at the same low level of detail elicited in the observation study. Nevertheless, as shown in the graph below (Figure 3-1: derived from broadly categorising into 'generic function' descriptions where possible), they still illustrated the tendency for unsophisticated tasks to be the norm in people's work: only half the respondents claimed to perform any kind of analysis activity.

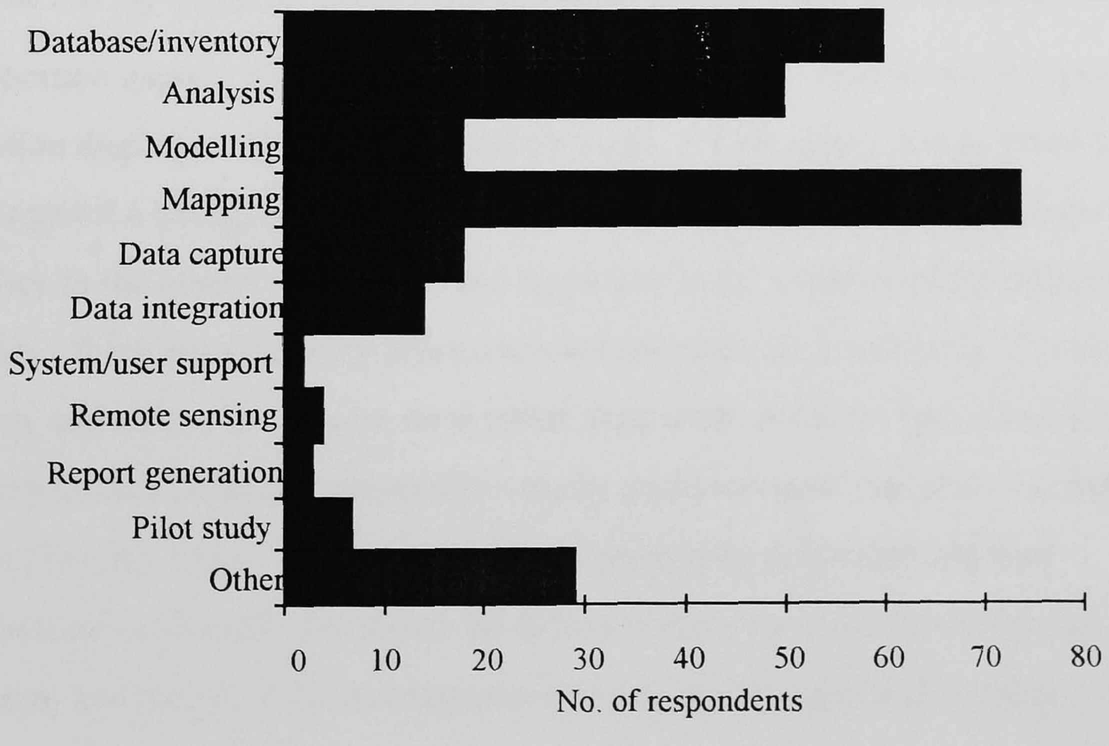


Figure 3-1. Types of digital mapping task reported by USIS Phase I survey respondents

Another interesting aspect of the survey responses was the way that 88 (58%) chose unprompted to explain the *content* of their information within this 'task' question: i.e. the application area in which they worked and/or the type of information represented in their maps. Of those 88, 60 (68%, or 39% of the total respondents) gave enough detail to determine the precise content of their data. These are listed, in two columns of descending frequency, in Table 3-2 (which adds up to more than 60, since some users mentioned storing more than one type of item in their GIS).

Information content	n	**Information content**	n
Census and other socio-economic data	9	Fires	2
Land cover and use (general)	8	River and sea flood areas	2
Water mains pipelines + street furniture	8	Soils	1
Land ownership and charges	7	Mineral/ chemical occurrences (not mines)	1
Geology and/or underground water sources	5	Coastlines	1
Flora + fauna distribution, vegetation	4	Elevation (height above sea level)	1
Planning applications and constraints	3	Streetmaps, buildings, urban planning	1
Road accidents	3	Grounds maintenance	1
Agriculture	2	Electrical distribution cables	1
Forestry	2	Cellular communications	1
Mines (mineral extraction etc.)	2	Bus stops and routes	1
River and sea flood areas	2	Highways	1
Climate	2	Emergency vehicle command & control	1
Crimes	2	Electoral boundaries	1

Table 3-2. Types of geographical features mapped by USIS Phase I survey respondents

One important aspect of this table is that while some of the most common types of information displayed in respondents' digital maps, e.g. pipelines, would often be viewed against a background of a portion of an urban streetmap - the map type seen most often in the observation study, and employed in the experiment described later in this thesis - there are also many other users whose maps are at the scale of regions or countries, and whose focus is on rural rather than urban areas (in fact, considerably more than in the subsequent observation study, partly because the latter was based solely in the UK). These users' knowledge of geographical features and their representation is probably dependent on different expertise from those viewing streetmaps, and may involve less opportunity to apply concepts and learning from everyday life.

For example, in Figure 3-2 below, once we know that a series of similar blocks situated alongside a pair of parallel lines are consecutively-numbered semi-detached houses in a single street, these are probably easier for most people to visualise and interpret than topographical features marked by contours. Thus the number '150', as a house number, will lead most people to some sensible inferences in the left-hand map (e.g. the house adjoining it is either 148 or 152; clearly this is a fairly long street even though only a few houses are currently visible). This is easier than in the right-hand map, where we may not be sure of the unit of measurement, or whether the contour 'outside' the marked one is higher or lower, or overall what kind of topographic feature is being represented. To a trained reader of topographic maps, however, especially a non-British viewer unused to our suburban semis, the map on the right may communicate more information and be easier to visualise. This issue of specific expertise will be returned to in Chapter 5.

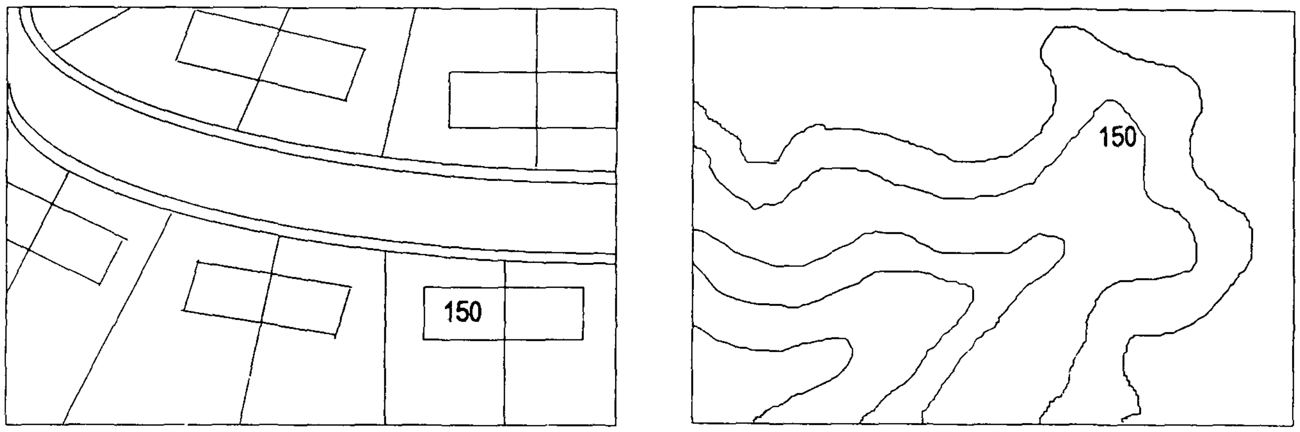


Figure 3-2. Map viewers' specific expertise will increase familiarity of different types of map feature/type, and e.g. the deduced implications of the number 150 in the two contexts shown

The demands of different map types are important to bear in mind, but will not be dealt with thoroughly in the remainder of this thesis, much of which for simplicity's sake will focus on streetmap-scale displays (which were the most common map type in the USIS observation study).

3.1.2 Observation study

Within the observation study, the structured interview component of each visit included completion of a checklist of 48 common GIS activities/tasks. The users were asked to state how often, if ever, they performed the tasks in the checklist. In the absence of an existing taxonomy of GIS tasks, as discussed above, the checklist had been designed by members of the project team (Davies, 1995 has more details). After piloting the checklist, it was felt that it was adequate to enable the construction of task 'profiles' of the users. The researcher completed the checklist by discussing the user's frequency of performance of each item.

The most frequently-performed tasks among the 21 users included (roughly in decreasing order of frequency): opening, closing and saving map/data files; starting and quitting the GIS software; displaying a map on the screen; editing, updating and integrating map data; and browsing around it. Other tasks which tended to be performed frequently included querying attribute data by clicking at a point or feature on the map; starting up the computer on which the GIS ran; plotting maps and printing reports onto paper; spatial data capture (digitising, drawing or scanning in spatial features for adding to the digital map); selecting and searching by specific attribute values; converting data between different formats (e.g. where some data was obtained from a different GIS or an outside supplier); and generating summary statistics and reports on-screen.

Overall, these tasks were largely fairly routine and basic (both from the user's and the computer's perspective), and were largely focused on handling a map rather than dealing with computational functions or with text/numeric details. As such the checklist showed broad agreement with the range of tasks actually performed by users during the visits, as shown below.

3.1.2.1 Tasks demonstrated

An approximate guide to the activities recorded in the task demonstration session (for the 20 users recorded on videotape) is given in Table 3-3 (see Glossary for descriptions of some of these activities).

User	Startup	Map retrieval	Query (point etc.)	Search attributes/polygon	Zooming in	Panning	Zooming out	Printing report	Map presenting/plotting	Digitising features	Editing attributes	Analysis	Data conversion	Measurement	Map creation from coordinates	Editing features
1	o	•	•		•	•	•			l,s	•					
2	o	•						•	o	p		•	•		•	
3	•	•	•		•				•					•	•	•
4	•	•			•				•							•
5	•	•	•	•									•			
6	o	•	•		•				•							
7	•	•	•	•	•				•							
8	•	•	•		•				•			•	•			
9	o	•			•	•	•			p						•
10	o	•		•	•				o							
11	o	•	•		•								•			
12	o	•	•	•	•		•		•							
13	•		•		•	•	•		•	p	•			•		•
14	o	•	•		•	•	•				•					•
15	o	•	•		•				o							
16	•	•			•	•	•									•
17	•	•	•	•	•	•	•			l,s	•	•				•
18	o	•	•	•	•		•	•						•		
19	•	•	•		•	•	•				•					•
20	•	•	•	•	•	•	•	o	o	p						

Table 3-3. Basic tasks recorded in each USIS observation study videotape

Notes: •=whole activity recorded; o=only partly recorded.
In digitising column, p=polygon, l=line, s=site (single point object)

A key point here is that although the most common tasks generally involved manipulating the visible map, this would not necessarily *have* to be the case (and wasn't for every user). A GIS can be used extensively with little reference to the maps; it is possible to query, analyse and generate reports with the maps only functioning as illustrations peripheral to the task. An example of such use was seen in a visit to the marketing analysis department at the headquarters of a major supermarket chain, where the focus was on generating sales predictions of various types of food product based on socio-economic data (from the national census) and on geographical position of the shops. Here production and manipulation of a map was a relatively minor and infrequent task, and mostly a means to the end of performing calculations and illustrating textual/numerical reports.

Even in more locally-oriented applications like local authorities, it would be possible to perform many tasks (such as searching or querying) in less map-oriented ways, but in general the map *was* the predominant means of interaction between the user and the information (but not between the user and the *system*, which generally had to be instructed via menus, buttons, typed commands and dialog boxes). The central role of the digital map itself was therefore qualitatively confirmed in this data.

Analysis of the videotapes showed the maps to be present on the screen most of the time, and used as the main means by which queries were specified. This was true both for relatively junior digitising staff who input and edited data, and for managers performing queries or obtaining paper plots from them. This centrality of the map within the task, for most users, should be contrasted with its limited appearance in the HCI-oriented usability measures reported in the previous chapter. The map appears to be what users were *trying* to use, but they may have had to spend too much time dealing instead with the system.

3.2 Analysing and evaluating tasks

The author then decided to perform a further, more detailed analysis of the observation study video data. The main basis for this can be summarised in an observation by Gould (Gould, 1994) that "what people *say* they do with GIS is not necessarily what they *actually* do - or at least not the whole story. Generic level tasks can be divided into dozens of subtasks which are no less important to recognise." In other words, the *nature* of individual actions may tell us something about the way the system requires users to handle their tasks.

As shown in Chapter 2, the analysis of usability variables in the USIS observation study indicated a worryingly large number of repetitive but slow actions in achieving users' tasks. There was also an impressively strong correlation between the amount of time spent productively (basically, in not correcting errors or looking for functions/assistance) and the degree to which the users rated their systems as 'Suitable' for their tasks (in terms of complexity and relevance of the actions and functions performed). After those analyses were complete, the author felt it would be interesting to try to identify where the actions being performed were perhaps superfluous, unintuitive in terms of the task goal (which is broadly assumed to be handling spatial information), and/or ripe for being automated in some way to simplify users' work.

Most task analysis procedures developed or used within the HCI community (see Whitefield and Hill, 1994, for a summary of them) have been purely descriptive, in that they break down the task into actions without trying to judge which actions are directly of use and which are artefacts of the system design (see also Kirwan and Ainsworth, 1992). However, one HCI paper (Whitefield et al, 1993) made an initial attempt to develop an analysis that was more prescriptive, in distinguishing 'work' actions (those which actually achieve the intended goal, generally defined as transforming some object or element in the work domain) from 'enabling' actions (those which prepare for the performance of the actual work). The author decided to examine the action-by-action data already extracted from the observation study videotapes, to attempt to apply this work/enabling distinction.

The analysis which resulted has been reported in detail elsewhere (Davies, 1998). In the event it involved far more methodological and philosophical considerations than might be expected at first sight. Whitefield et al had only applied the analysis to a few simple text editing tasks: a GIS is a very different beast from a basic word-processor. Not only is the content and structure of the data obviously different - a spatial array rather than a linear narrative - but the system also acts as the chief store, retrieval mechanism and analyst for the information as well as its editing tool. Furthermore, the cartographic complexity of handling maps, and the ways in which textual attributes can be linked to visible spatial features, make it inherently more difficult to define clear low-level task goals and to decide which actions 'do work' and which merely 'enable' it.

There were also other issues to consider: in particular, are 'enabling' actions necessarily bad or wasteful? Various literature, in the past century of contemplating and analysing human work, has suggested that strict notions of 'productivity' are often inappropriate through failing to allow for human growth, fulfilment through learning and sense of control, etc. This was discussed further elsewhere (Davies, 1998); the basic conclusion was that we can try to identify 'enabling' tasks but should be careful in our diagnosis of their superfluity. Certainly, as with the problem of 'search' time in the MUSiC metrics used in the last chapter, we should be careful that we do not denounce browsing through information as an unproductive 'enabling' task, when it is an important step towards a user's understanding.

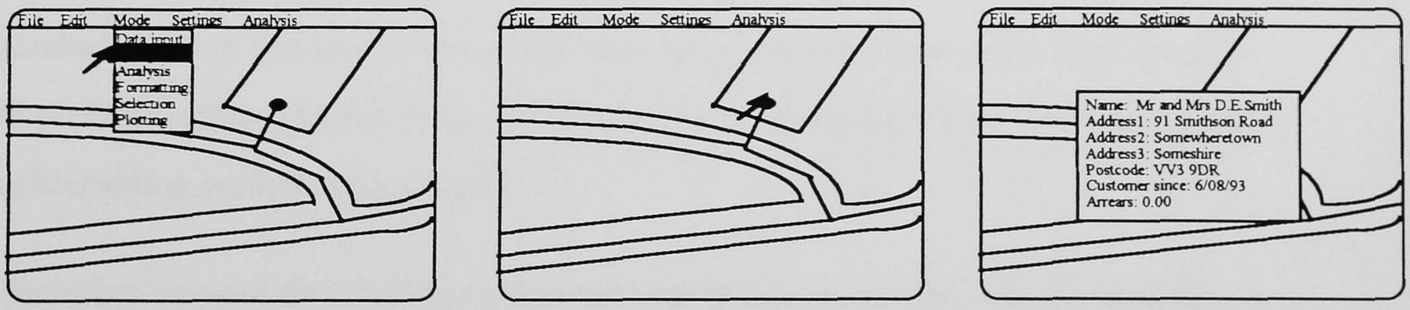
The present author also felt it necessary to distinguish goal acquisition behaviours (checking what the user's goals actually are, i.e. what needs to be done) from enabling behaviours within the performance of the task: Whitefield et al had counted them as enabling behaviours while recognising that this was not always appropriate. In addition, while Whitefield et al had focused on tasks with a word-processor which was already set up and running, with the file loaded and present on the screen, many of the tasks performed by users in reality are what the present author decided to term 'general enabling' - enabling actions in that they perform no actual work, and general in that they are not specific to any low-level goal of transforming the information content. Examples of general enabling are starting up the computer, starting the software and loading the map file, prior to performing a series of edits or queries. Thus Whitefield et al's original work/enabling distinction was supplemented with two extra categories, goal acquisition and general enabling.

Other decisions included the issue of non-actions, such as the time periods discussed in Chapter 2 where the user waited patiently while the system performed one slow action (or a series of actions, e.g. from a programmed macro script). These were counted as enabling, since the user wasn't performing work but was still stuck in front of the machine without the task having been completed.

Ten of the original tapes were randomly selected for the re-analysis: this was considered to be sufficient for testing out the analysis method, and to suggest obvious areas where the amount of 'enabling' rather than 'work' was seriously disproportionate. The aim was not to produce statistically significant numerical analyses, but to demonstrate general tendencies within the tasks performed by the users. The results showed that the proportions of the different types of action varied enormously between users. This was true whether the number of actions, or the total time taken for them, was considered.

Just one example should suffice to illustrate this: six of the ten users performed a point query task at some point in the session, in which they retrieved a set of textual attributes for a displayed object by pointing the mouse or digitising tablet at it. A typical sequence of actions within this task is illustrated in Figure 3-3:

Opens software, retrieves map etc.	General enabling
Checks paper map (to find next set of attributes to be checked)	Goal acquisition
Requests to perform query	Enabling
Points to an object within map (and system displays attributes)	Work



Name: Mr and Mrs D.E.Smith
Address1: 91 Smithson Road
Address2: Somewheretown
Address3: Someshire
Postcode: VV3 9DR
Customer since: 6/08/93
Arrears: 0.00

Figure 3-3. Actions taken to perform a 'point query' with a typical GIS, and how they may be classified according to Whitefield et al's 'work' versus 'enabling' distinction

The first occurrence of each of these was compared across the 6/10 users who performed it[14]. The relative times spent on enabling and on work actions within this first, simple, query task are shown in Figure 3-4.

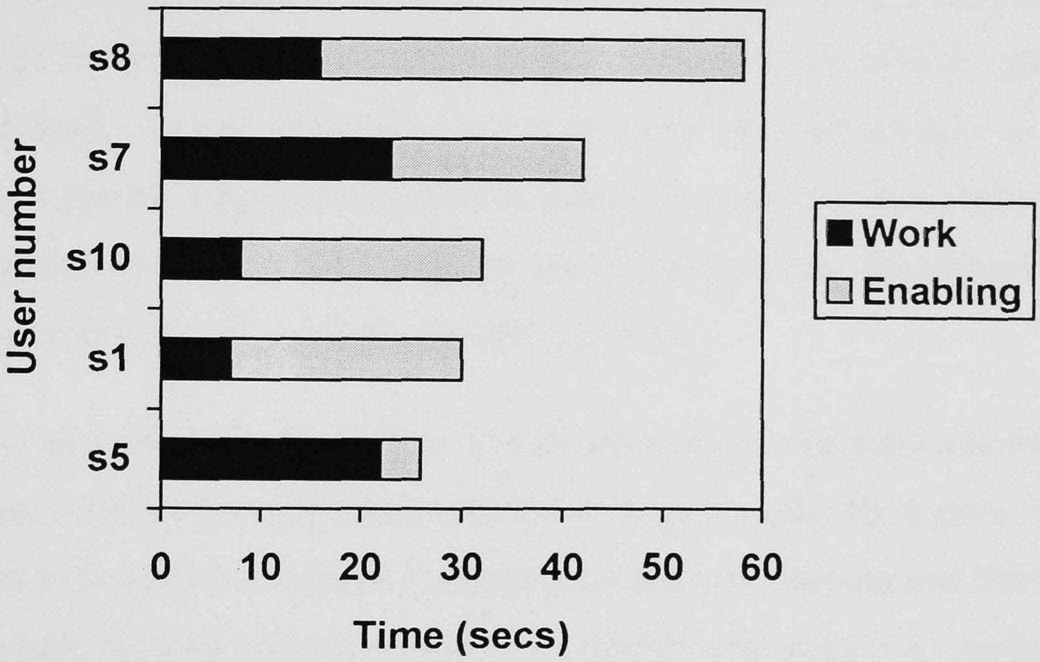


Figure 3-4. Time spent on the first point query performed, and proportion of 'work' and 'enabling' time, by different GIS users

It can be seen that the proportions of apparent 'work' and 'enabling' varied greatly, but did not particularly reflect the overall time taken (arguably the most important measure of task efficiency). There was no evidence from the subjective or objective metrics that users who apparently had a higher 'enabling' proportion found their systems harder to use (and if there had been, small sample sizes aside, such metrics will probably always be much less effort and more informative than performing the work/enabling analysis in any case).

It therefore seemed that, besides extra methodological problems with the analysis (discussed in Davies, 1998), there were unlikely to be any further useful generalisations that could be made about the work/enabling distinction within the varied GIS and tasks in the observation study. Perhaps it is reasonable to conclude that any attempts to be more prescriptive about systems will always need to be

---

[14] With such a query, quite often the user has to set the system into 'query mode' first as part of the task, and can then point at a succession of features to perform a series of queries. Thus taking the first occurrence meant including any specific setting-up actions for the task, as well as the defining mouse-click.

tailored to the specific task and system context. In any case, a given action could actually be classified differently depending on the context in which it was performed, even by the same user and on the same system.

Zooming in, for instance, could be performed solely to make visible a detail that needed to be checked (work), or could be done to define an area to be plotted on paper (enabling), or could occur just after the map was retrieved and prior to a series of tasks on a specific region of it (general enabling). Zooming in could even occur as a goal acquisition behaviour, e.g. where the user did it to check what another user had left unfinished in previous digitising/editing work.

Yet ideally, we could argue that zooming is an unnecessary additional task which with a paper map, where we would simply lean closer or grab a magnifying glass, would be so trivial as to be performed without conscious awareness. Haunold and Kuhn (1994) used a similarly detailed analysis, following the GOMS task breakdown method (Card et al, 1983), to show that digitisers working on the national cadastral map of Austria spent around three-quarters of their working time performing routine actions, of which zooming and panning were the most frequent (and in that particular system, were very clumsy, involving opening a special window to select the type of zoom, then waiting for the system to redraw). According to Haunold and Kuhn, the labour hours spent in zooming/panning alone was thus costing the Austrian project a million dollars. Furthermore, each zoom/pan action, whether or not we see it as crucial 'work' or mere 'enabling', required explicit mental preparation on the part of the user (because of the clumsy way it had to be performed) and thus arguably disrupted the task while also forming part of it.

This suggests that the user's cognition in the context of their task is more crucial to defining what is going on in it than a detached analysis like Whitefield et al's. If the users' main interest is in the map, and if they prefer to zoom in to specific scales while working because that level of detail suits their understanding of the space that is represented, then identifying this need and optimising its fulfilment is perhaps more feasible and important than objective measures of the overall speed or effort, or of proportional 'work'.

After all, if a digital map is zoomed in too far (e.g. in Figure 3-5 below), a user may only see a few lines on the screen and not be sure which spatial feature each line 'belongs' to (e.g. the west wall of one building or the east wall of another), or what the feature is (a house, a church, or even perhaps a road or other feature if colour coding fails to distinguish them). If a specific area is extremely familiar however, perhaps after hours of tedious and precise digitising, the user may be able to identify a particular part of a street simply from the exact curve of a single unmarked line on the screen.

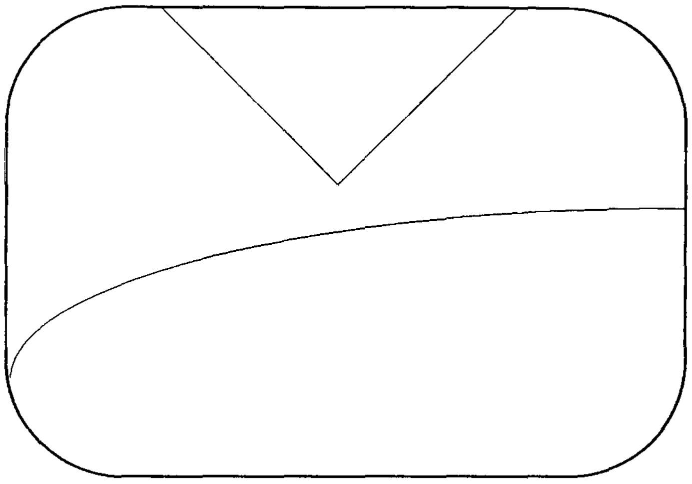


Figure 3-5. When a digital map is 'zoomed in' too far, no identifying features may be visible and the user will depend on specific familiarity and/or general expertise to identify objects

For a complete novice user, of course, some lines, shapes and text abbreviations may fail to be identifiable at all as representing real geographic entities. The actions a user will take will surely depend very much on the extents of both generic and specific familiarity with such maps, and on the degree to which geographically identifying the features is even relevant to the task at hand.

Finally, it was concluded that the work/enabling analysis still seemed to omit some types of behaviour: the actions necessary to clear up, save things and put them away after a task, i.e. *dis*abling behaviour. Overall, it became clear that a lot more methodological work would be needed (not of sufficient interest to the present author) before the work/enabling distinction could safely be applied to the low-level interaction between the user and any system as complex as GIS.

3.3 The cartographic perspective

We will now turn to considering how cartographers have described and classified users' tasks, i.e. the interaction between the user and the digital map from the perspective of the information rather than the system. Since viewing a digital map is not entirely distinct from viewing any other map type, we can start by drawing on cartographers' attempts to classify the low-level tasks which general map use tends to involve.

McCann (1982) extracted 18 task components from a task analysis of map-reading tasks, based partly on her own observations and partly on a review of the cartographic literature: this work was part of a report for the Canadian military, and the task breakdown reflected the tasks which military personnel might perform with a map. Some of the tasks therefore either reflected 'live' use of a map (e.g. for orienteering), or uses of a map for understanding topography of rural terrain (neither being common applications of digital maps). Others, such as determining absolute reference locations or extracting distribution patterns, can be performed automatically by a GIS if the user issues appropriate analysis or query commands, so these are also no longer strictly tasks (in the sense that the user does not perform them, but views and interprets the results). The tasks that remain relevant in the context of digital maps are:

1. Symbol detection
2. Symbol interpretation
3. Pattern detection and interpretation
4. Determine the attributes (non-spatial characteristics) of map features
6. Comparison of symbols or patterns
7. Annotation, updating, highlighting
11. Determine relative location of map features
12. Search for map features
15. Inferring unportrayed features
17. Visualisation

Even in the very different context of the GIS user in an office, most basic operations in viewing and understanding a map will correspond to the task components put forward here. However, as discussed by various authors (e.g. Schlichtmann, 1991), these tasks do assume that the map is correct; examining a map for errors (by comparing it to other maps or data, or to real-world views) is another potential task, and one which is quite common among GIS users (e.g. when checking that information has been digitised correctly). We might add that with a digital map one can change its whole appearance, in ways that are not possible once a map is printed on paper, and such edits or transformations are additional to McCann's list (albeit related to her category of 'Annotation, updating, highlighting').

It is important to note how different this list is from the earlier summaries of what USIS users were doing with their GIS, which inevitably focused on the physical actions performed with the system and the visible display. However, both lists would be recognisable to a GIS user as being 'what they do'. The McCann list focuses on users' *goals* and *cognitive* processes, whereas the earlier analyses focused on *what users have to do with the system* to let them achieve these. The gap between the cartographic standpoint and the reality of GIS design, which focuses on a computer system with complex generic functionality to control geometric representations, is thus obvious. Perhaps with more 'intelligent' user interfaces, as wished for by other authors (e.g. Turk, 1990; Haunold and Kuhn, 1994), future GIS functionality will be focused on people's real interest in geographic information, rather than on what Whitefield et al might have summarised as a gigantic 'enabling' process.

A broad approach to understanding tasks for GIS, which to some extent synthesised these two aspects of cartographic goal and system function, was the geographic information usage taxonomy put forward by Calkins and Obermeyer (1991). While focusing largely at the level of organisational implementation (so less detailed than the approaches discussed so far), Calkins and Obermeyer considered, separately, the characteristics of uses of geographic information, and the characteristics of the data and system functions. However, there seems to have been some intention on their part to link the two together in practice: the taxonomy was meant to be used to structure surveys of GIS use in organisations, and clearly the authors were expecting the relevance of particular GIS characteristics to depend on the overall purpose of users' tasks. The purposes they identified were (pp.346-7):

a. to build or maintain an inventory of significant data (possibly responding to a mandate)

b. to monitor a process, such as the issuance of permits where the objective is to insure compliance with appropriate regulations or procedures; to keep track of a series of planned events over a predictable time period

c. to browse through a geographical database
   (i) structured (goal oriented) browsing, e.g. environmental modelling
   (ii) unstructured, e.g. spatial exploratory data analysis
   (iii) looking up specific facts

d. to summarise geographical data
   (i) simple counting
   (ii) measurement (i.e. area or perimeter of a polygon)

e. to analyse mapped or spatial data, using the geographical domain as the organising principle where the problem is well defined
   (i) map overlay analysis
   (ii) network analysis (route finding, districting, optimal site selection using network distance, etc.)
   (iii) terrain analysis models
   (iv) spatial interaction models (based on a zonal structure)
   (v) flow analysis through a network (stream network models)

f. spatial decision support systems to assist in reaching a decision in situations where the problem is ill-defined

They then identified these primary functional GIS capabilities as potential necessities for supporting use:

a. display of graphic data
b. query and display
c. counting and direct measurement operations
d. map overlay (map combinations)
e. network algorithms
f. spatial models

It should be noted that these higher-level descriptions of task and function *seem* to fit together quite logically, but we have already seen that actual *implementation* of GIS requires the user to waste a lot of time on actions not directly contributing to the purposes outlined above.

Another important point, and perhaps a justification all by itself for the focus solely on the digital map in subsequent chapters of this thesis, is the recognition in Calkins and Obermeyer's work that the geographical domain of the map is used "as the organising principle" for the data stored in the system. In other words, the local authorities, utility operators, environmental agencies, retailers and communications companies visited in the USIS observation study were choosing explicitly to use the visible map as the *medium* through which their data was organised, classified, searched or browsed, and displayed or output. So the user's grasp of the digital map representation - both of the basic geographical symbolism and their organisation's own data objects superimposed upon it - becomes the crucial cognitive aspect of the task. The *system* could be improved, or users can obtain help from online, paper or personal sources to enable them to cope with it; but the *map* is the information itself, over which they supposedly have complete control.

The importance of matching maps to cognitive aspects of tasks has been illustrated in various military studies (e.g. Harwood, 1989), in experiments where helicopter or aeroplane pilots perform realistic tasks with a simulator. In the late 1980's there were a number of such studies, either manipulating the visual display or studying individual differences between pilots. Harwood's one focused on emphasising either the spatial relations between objects in the environment, or between them and the pilot himself, by using a display which was either 'north-up' (north at the top of the screen) or 'track-up' (top of the screen is the horizon in the direction the pilot's heading). Harwood's conclusion was that pilots performed better on north-up maps overall, for general navigation and accurate object location, but the track-up map was better for orienting oneself when lost in an unfamiliar area, and therefore displays should be configurable but with north-up as the default. The complexity of what people do with even these highly simplified airborne displays illustrates the importance of considering how task interacts with interpretation and design choices, and this has also been thoroughly demonstrated with regard to route navigation displays on land (Burnett and Joyner, 1995).

Focusing on classifying these visual aspects of digital map tasks, Knapp (1995) examined a subset of a classification scheme of visual operations developed for scientific visualisation. She determined that out of this scheme, the four primary visual operators where maps are concerned are 'Identify', 'Locate', 'Compare' and 'Associate', and argued that all digital map tasks (as opposed to GIS system-oriented tasks such as file management etc.) could be considered in terms of these four operations, which in turn could be clarified according to spatial, temporal and single/multiple characteristics. For example, the operation 'Identify' was described by Knapp as follows (p.367):

> Definition: to ascertain the definitive characteristics of
>
> Spatial identification
>
> Singular:
> - length
> - surface area
> - volume
> - shape
> - irregularity
> - orientation
> - midpoint
> - slope
>
> Multiple:
> - min, max, range
> - average, variability
> - pattern of distribution
> - layout
> - distance
>
> Temporal identification
> - change
> - extent
> - sequence
> - movement
>
> Associative identification
> - distinguish
> - categorise

In other words, the list shows the ways in which 'identify' is one part of interpreting map objects and their spatial or semantic relationships. It should be clear to the reader that all of McCann's map tasks, or at least those visual tasks that involve viewing the map without altering it, can be fitted into one of Knapp's four operations, as follows:

- Identify: symbol detection, symbol interpretation, visualisation
- Categorise: determine the attributes of map features
- Locate: search for map features, infer unportrayed features
- Associate: pattern detection and interpretation, comparison of symbols or patterns, determine relative location of map features

This scheme has potential use in analysing a given visual problem-solving map task, such as asking users to describe a route through the represented space, if such a micro-level analysis was useful (e.g. in training digital map users to perform highly specific tasks).

However, the reader will by now realise that a key variable distinguishing different tasks, and also distinguishing Knapp's scheme from Calkins and Obermeyer's (besides differences in their level of analysis), is the degree to which the user understands the representativeness of the map in relation to the 'real' world. Writing from a geographer's standpoint, as many commentators on GIS have done, Calkins and Obermeyer seem to have assumed that geographical understanding was a precondition for using the information. Most activities identified by McCann and further broken down by Knapp, however, such as symbol detection and visual search, could arguably be performed to some degree by a map user who had no idea what the symbols meant, or even that the display was meant to be a map. Only visualisation (e.g. during wayfinding or decision-making) implies a complete mental translation of information about a scene, from a map to an image of how a street (say) would look from a particular perspective, or the reverse process. Other operations may require only simple visual scanning of a display which could arguably be meaningless, although some will require a knowledge of items' semantic labels/ categories. One's knowledge of the geographical visualisation represented therein may not be crucial to most digital map operations.

This could explain why cartographic principles of representation, examined in Chapter 2, seemed not to impact greatly on users' concerns or performance in the USIS observation study: whilst they had to recognise and edit the visible geometry of the data, they did not necessarily need to be constantly aware of its geographic symbolism. Having a legend present, for example, or being able to rapidly spot 'target' object types, may have been unimportant factors in many users' day-to-day work.

This leads us to ask what kinds of knowledge and processes *are* generally involved in viewing and interpreting a digital map, and how the user's task and other factors determine or influence them. The next chapter will review the literature regarding spatial cognitive processes, Chapter 5 will look at the effects of individual differences such as expertise and ability, and the study in Chapters 6 and 7 will examine these factors' interaction with task constraints.

3.4 Summary

This chapter has focused on the tasks performed by digital map users, as indicated by the USIS data and by the cartographic literature. Eschewing the frequent tendency to see digital map use as interesting only insofar as it is used in advanced visualisation, analysis and decision making, the focus on day-to-day map use activities shows that the spatial information content of the map (rather than the system-related issues and actions discussed in the previous chapter) appears to be central to most users' conceptions of their tasks and goals. An analysis of where actions appear to contribute to these goals ('work'), or not ('enabling' etc.), suggested that few clear generalisations can be made across different systems or applications, since the proportion of time spent on directly 'useful' actions varies greatly. It is argued that expertise, creating familiarity with the task and the map/system, tends to overcome such issues, and the cognitive processes contributing to this need to be better understood. Cartographers have tended to consider the geographically representational aspects of map-based tasks as the key means by which to distinguish them, but it is suggested that many of these tasks may not require extensive geographic understanding or visualisation to be performed successfully.

Chapter 4: Cognitive processes with digital maps

This chapter will review the literature from psychology, cartography and related areas, concerning the cognitive structures and processes involved in viewing, interpreting and remembering a digital map. The chapter will concentrate on these aspects:

1. ***Seeing and 'reading' digital maps***: the unique perceptual and cognitive constraints on viewing a digital map, compared with paper maps and other types of visual information.

2. ***The structure of users' map knowledge***: the likely nature, features, structures and distortions within the user's mental representation of the map, in the light of theoretical and empirical developments in cognitive psychology.

3. ***External factors influencing spatial understanding***: how the mental representation of space influences or is influenced by external aspects such as task, cross-scale transfer of knowledge (e.g. using a map to locate oneself in real, navigable space), and perceptual salience.

4. ***Reference frames: interpreting relations between objects***: how we appear to utilise certain hypothetical structures to determine and describe where objects are in relation to one another, in relation to the map or screen, and in relation to ourselves.

The chapter ends by formulating tentative hypotheses about how people's interpretation and mental representation of a typical digital map may be influenced by the tasks they are made to perform with it and/or the language used in describing it. These hypotheses will form the main starting point for the experimental study described in Chapters 6 and 7. Chapter 5 will meanwhile take its cue from the spatial cognitive processes discussed in this chapter, to consider how individual differences may arise in these processes which may be related to age, gender, experience, expertise and other factors.

4.1 Seeing and 'reading' digital maps

GIS, and other systems handling digital maps, may be seen as a special case of complex systems[15], where the spatial representation depicts not just a small real-world layout such as a roomful of equipment, but a real-world space too large for users to completely view *except* via maps (or scale models, or aerial photographs). Thus digital maps are an interesting case not only of human-computer interaction, as the USIS work examined, but also of the translation of real-world geographical knowledge into a small-scale digital representation. The user can interact with the data and change it, unlike a paper map, and many of the latter's cues and conventions are absent. David Mark (1993) argued [p.58] that "The fact that this knowledge of the graphical world of the screen or paper must be combined somehow with knowledge of the geographical world that the GIS or other software represents, is perhaps what makes user interfaces for GIS an especially interesting and challenging topic for theoretical and applied research."

One obvious problem in this interaction is the physical limitation of the screen, which typically has a diagonally-measured size of between 14 and 21 inches, and a resolution of a few hundred pixels across and down. The world is round, and the geography on its surface is continuous, not packaged in discrete curved-edge-oblong areas. Therefore, as with displays of long text documents (e.g. in a word-processor or digital library application), users are severely limited in the proportion of the information that can be seen on a screen at any one time. As with many text displays, this can be controlled to some extent by 'zooming' in and out, leading to a trade-off between visible resolution and available quantity of information. This trade-off holds up to a point, but not beyond that point: when zoomed in too far, clarity tends to be poorer since users can no longer smoothly read text or view the whole of a shape; when zoomed out too far, the display reaches the point where the user simply can't

---

[15] A term usually applied to computerised *process control* or monitoring systems in factories, power plants etc. where the positions and/or status of various equipment are shown graphically on a single display. Although this has a 'real-time control' element which is not always true of GIS applications (but can be, e.g. in emergency services), they are both *complex systems* allowing many operations on a graphical representation of many real-world objects.

see certain text or shapes (and the GIS may be set up so that some types of feature and text label deliberately vanish at such scales).

The restricted field of vision afforded by the screen may also affect the way we visualise the information beyond it. We who have become used to the computer age must not take for granted the ability to accept this unnatural limit on vision, as discussed in Chapter 1: distortions in our understanding of the overall space appear to arise from the near-oblong limitations of the computer screen.

Apart from these distortions, we still do use our expectations and knowledge to 'see' the data as a continuous map, and we know or assume it extends beyond the visible screen. Within the visible display, according to established findings in visual perception and cognition, we tend to group objects perceptually and in memory to help us understand and recall them. These groupings may not just be according to physical salience, e.g. colour, distance etc., but also by similarity of purpose or type, as discussed in the next section of this chapter. Thus the visual design of the map impacts on how information is conveyed to the user.

We saw in Chapter 2 that cartographers and ergonomists, especially in high-risk areas such as aviation and defence, have studied such design variables in great detail and tried to make prescriptive judgements suited to specific tasks and contexts. Clear differentiation and meaningfulness of colours and symbols, avoidance of clutter in the display, and appropriate use of emphasising techniques such as boldness, red or flashing, all impact on what users notice and how fast they spot important targets. As we concluded, such considerations are less critically relevant in more sedate environments such as utility or highway maintenance offices. However, it is obvious that visual design always makes *some* difference to users' perceptions. Indeed, as we suggested earlier in Chapter 2, the visual design of the map may even serve to 'distract' users from the commands and procedures necessary to interact with the GIS which displays it.

Visually, the data displayed in a GIS can seem very different from a paper map. It will often consist of arbitrarily-coloured lines and points on a blank background which may be any colour, even black (which was quite common in the USIS case studies).

The disparity between this type of representation and the real-life appearance of the buildings and roads is easily as great as in a traditional paper map such as the Ordnance Survey series. We might wonder if the user always sees the display as a 'map' at all. To what extent do users need to keep the display's 'true' representational nature in mind when handling it, i.e. the extent to which it corresponds to real houses and streets[16], and to what extent can the display be treated as if it was its own arbitrary little world of coloured lines, odd geometric shapes and scraps of text?

This problem of map interpretation has been considered as follows (Pick and Lockman, 1983, pp. 219-20):

> The fact that the term reading is applied to extracting information from maps is probably due to the belief that maps, like text, represent information in a highly encoded form. However, maps do in many ways bear a formal, i.e., projective, similarity to the thing they represent. The formal similarity between maps and the spatial layout represented, in general, is not so great as that between pictures and the aspect of the world they depict. And we would not ordinarily speak of reading pictures... Somewhere, say between pictures and maps we begin to speak of READING the display.
>
> To the degree that there is considerable formal similarity between the representational display and the thing being represented as in the case of photographs, the perception of the represented object or layout should not pose any special problem... However, the fact that information in maps is more encoded than in photographs, in the sense of bearing less formal similarity to the part of the world being represented, results in map reading being a more formidable task than picture perception, even for adults.

MacEachren (1995) suggested that the degree to which a user considers a visual display to be a 'map' will affect the representations they apply, i.e. the precise influence of expectations on their 'reading' of the presented information. The concept of 'map', MacEachren admitted, is a fuzzy one, partly culturally-defined, and to some extent 'radial' in that central prototypes exist, around which less obvious examples of 'mapness' are grouped. (For example, for British people an Ordnance Survey map from the 'Landranger' 1:50,000 series may be seen as a definitive 'map'; a raster image of the same area showing vegetation types or land use might be less readily identified as such and hence further from the 'central' prototype).

---

[16] Most common GIS applications involve maps of urban streetscapes. This research focuses on such maps for this reason.

This then is a key issue - the extent to which 'reading' the map involves building an extended representation, beyond the immediate visual and geometric interpretation of the display. If map *reading* is to any extent 'formidable', and is unnecessary in order to perform certain tasks that depend on viewing the map (such as editing its geometry), maybe users don't always fully 'read' it at all (just as proofreaders may not entirely absorb the content of a text).

Laurini (1995) described how GIS users have to handle errors in and updates to map data, performing edits and transformations which have no equivalent meaning in the real geographic space (even if sometimes they are prompted by an event in that space, such as a building being knocked down or sold, a boundary being changed or an entity being reclassified). Laurini described these operations and concerns as "the integration of different representations of reality and mental models" - the user's mental model is of the *data*, which is partly a geographic representation and partly a source of abstract information (such as details of ownership, organisation into layers, colour coding and symbolism reflecting non-geographic entities such as building material types). Some examples of arguably non-'geographic' GIS user operations given by Laurini are:

- alphanumeric updating (which can be used to change the spatial co-ordinates of a point, as well as for changing text attributes)

- correction of previously-generalised lines or shapes, to reflect more accurate knowledge

- integration of buildings or land parcels into combined entities within a cadastre (e.g. where someone buys unfenced land from a neighbour, or two houses are knocked into one, albeit with no visible change in the real outward appearance of the area).

- mixing two layers (e.g. where two companies merge their databases, or two types of cable/boundary/wall are to be treated as a single data layer).

The last two of these can result from real-world changes, but do not have to; the first two normally reflect better knowledge of the real world but not a change in how it actually is. In none of the cases is it helpful to imagine what the depicted landscape actually looks like, and a digital map user (possibly working from amended paper plans) may simply alter the geometry without trying to visualise the 'real' space.

However, as we will see below, most studies of map interpretation and use have assumed that we *do* see the map as a map, and that its visible geometry is only a means to an end in all map tasks. It should be borne in mind in the remainder of this chapter that the 'spatial representations' we posit in users' minds are not necessarily entirely geographical, and may indeed have very little content that looks like the represented spatial reality.

On the other hand, cognitive psychologists have very often tested people's spatial cognition not with symbolic representations of space in maps or pictures, but instead (and somewhat ironically) on either meaningless small-scale abstract geometric arrays of dots and shapes (e.g. Logan, 1995), or else unrealistic arrangements of 'real' objects in 'real' (haptic, i.e. reachable) space, such as unrelated everyday objects placed arbitrarily on a tabletop (e.g. Diwadkar and McNamara, 1997). Generally the focus of such studies is on our ability to perform mental rotations or transformations, which are only one small part of a map user's tasks (except in orienteering, wayfinding and visualisation: as explained in Chapter 3, these are either advanced or mobility-dependent tasks, and not the most common or fundamental aspects of digital map use). Such research will therefore not be reviewed in this thesis, since it tells us little about basic map understanding.

If their tasks *do* in fact encourage them to keep in mind the representational nature of the map, one would expect that users' prior knowledge of the 'real world' should lead them towards certain expectations about how that world operates, and what and where things should be found within it, which they would then apply to their interpretation of the visible display. This issue of geographic knowledge is one aspect of the role of expertise and experience, which will be considered further in Chapter 5.

4.2 The structure of users' map knowledge

One prolific author on the subject of GIS use, David Mark, has hypothesised a number of times that the key issue in GIS is the mental transformation the users have to attempt between the large-scale 3D geographic space represented in the map and in their minds, and the small-scale 2D objects shown and manipulated on the computer screen (Mark, 1993; Mark and Freundschuh, 1995). At first sight this seems an obvious statement about geographic information use - as previously and more elegantly illustrated in Tufte's 'flatland' problem (Tufte, 1990) - but the present author had reason to doubt it when thinking about the tasks of the users visited in the USIS observation study. As argued above, it is unclear whether everyday GIS use involves any such visualisation (whether 2D, crudely 3D or totally 3D). There was no evidence in the USIS data to suggest that Mark's visualisation problem made any noticeable impact. In any event, the process is impossible to perform fully for a user who has never seen, at any scale, what the represented town or region 'really' looks like, which may often be the case.

Of course, the USIS data was not originally intended to specifically address this question, and it could be argued that many visualisation 'failures' would not be revealed by the data since the researchers themselves were not familiar with the 'true' state of the entities represented on-screen. Nevertheless, non-involvement of spatial visualisation (mental imagery) processes on the part of the user is a reasonable hypothesis, given that such visualisation is not *necessary* for many everyday GIS tasks to be successfully performed.

What is necessary for these tasks is knowledge of certain *rules* of geographic space. For instance, the oft-quoted 'first law of geography' states that no two objects can occupy the same point in space at the same time; thus in a GIS it makes no sense for a user to create say two houses 'on top of' one another on the map. (In cases where two or more properties do occupy the same area in two dimensions though not in three, e.g. in a block of flats, a way must be found for the 2D 'space' of the GIS 'map' to represent this, e.g. by allowing for multiple attribute records to be linked to a single map object, or for several adjacent point objects to be placed within the same polygon area with separate text labels.)

Similarly, the GIS user will know from everyday experience of 'real' space that sections of roads do not appear in isolation: even a cul-de-sac is linked at one end to a continuous network of roads. It would thus be easy to spot the error where a discontinuous line segment occurred in a digitised highway network (one USIS user was in fact performing this task)[17]. But these 'rules' of geographic space, – and similar 'rules' about the 'space' of the map, the graphical environment and the computer screen – can probably all be represented in propositional form rather than as imagery in the user's memory.

Recent studies of memory for spatial location (in non-map contexts, e.g. Lansdale, 1998), have suggested that both propositional *and* visual-perceptual processes contribute to recall, such that the former produces an approximate encoding of the object's location (roughly equivalent to "a bit to the left of the white line") while the latter produces an exact recall of it (if it can be retrieved at all). We should be careful, therefore, in pushing the 'spatial versus propositional' distinction too far: spatial information and/or imagery may well be involved. Nevertheless, we still have little reason to assume that such a 'spatial' element would always be a geographically interpreted (or interpretable) one, rather than simple memory of a red square next to a purple triangle.

Initially, therefore, the author had considered postulating in this thesis that mental visualisation of real geographic space was not involved in GIS tasks, but further reading and reflection altered this view. There is a general consensus nowadays in psychology not only that mental imagery exists, but that some form of spatial mental representation is usually formed as part of a general mental model when performing a spatially-relevant task (see e.g. Johnson-Laird, 1983; de Vega and Marschark, 1996). Thus the most likely mental representation adopted by the user will be a mixed one, with some geographically related aspects that may be propositional or spatial or even

---

[17] New object-oriented military GIS are now recognising the usefulness of such rules in adding automatic 'intelligent' constraints to a digital map (Fletcher 1999): "A tank knows it is a tank - and will not allow itself to be plotted in the middle of a lake unless, say, it knows there is a pontoon bridge beneath it." [p.38] Hopefully this will enable future users to focus more on the meaning of their tasks, and less on the computer's intervening abstractions.

both. The next two subsections will explore the likely content and structure of such a representation, in the light of current theories of visuospatial cognition.

4.2.1 Cognitive maps, schemata and mental models

In geography, and environmental psychology, there has long been a tendency to talk about geographic knowledge being held in people's minds as a series of 'cognitive maps'. This map metaphor implies something like an aerial survey plan of a given area, perhaps built up over time from a combination of visiting it or seeing photographs (declarative knowledge), moving through it on specific routes (procedural knowledge), and/or viewing maps of it (configurational knowledge). The resulting mental representation is seen as akin to a traditional map, albeit with certain key differences such as a focus on personally relevant landmarks, and an absence of some data (e.g. for streets that have never been visited).

The term 'cognitive map' is still used in discussion of people's navigation and environmental behaviour, but various studies have shown it to be quite unlike the physical artefacts we call 'maps'. Besides the fact that affective aspects (attitudes to and preferences for particular places or environments) do not easily fit into the 'cognitive map' concept (Spencer et al, 1989), people's knowledge does not appear to be flatly two-dimensional. Both propositional and spatial information appear to be integrated in structures which show some hierarchical properties, judging from the patterns observed in people's recall and errors: in particular, the use of landmarks for navigation and orientation appears to depend on hierarchically organised 'clusters' of items rather than a layout based on true Euclidean geometry (Hirtle and Jonides, 1985).

It has also been suggested (Peterson, 1987) that there is a parallel between the construction of people's mental spatial structures and the construction of a cartographic map, especially a digital one, not in terms of accurate representation but in terms of its *inaccuracies*: people 'generalise' detail, ignore some types of geographical feature when only certain others are of interest, link descriptive information to pictorial imagery, etc.

In general, however, most of the work on cognitive maps has focused on a person's developing knowledge of a specific area or place, novel or familiar, based on various sources of knowledge or experience. Maps are of course only one of these sources, and there is no real evidence to directly link studies of our long-term knowledge of our home towns or university campuses with the type of mental representation constructed by (say) a GIS user viewing a map area, especially a novel one. Thus there may be parallels between wayfinding-experience-based cognitive maps and digital maps, and hence presumably between wayfinding-experience-based cognitive maps and digital map *users'* cognitive maps, but the different visual and task experiences of these two contexts lead us to expect many differences between them.

Indeed, there is already some evidence of such differences, in the systematic errors found in our long-term representations of familiar 'real' environments. Tversky (1993) summarised the ways in which studies have shown people failing to accurately recall distance, direction, large but not personally-relevant landmarks, alignment, angles and irregularities, when considering familiar areas and places. These errors, which largely reflect a tendency to simplify and to focus on personally useful attributes, are likely to be reduced when the map in front of us is focusing our attention almost exclusively on the precise geometry of the space concerned, and reminding us of its actual measurements. Of course they could affect our recall of the map area at a later date, but this is an irrelevant issue in situations where the data can simply be summoned to the screen again. The cognitive mapping literature is therefore of limited relevance to the tasks of GIS users.

In cognitive psychology, on the other hand, much of the work on visuospatial cognition has focused on situations where a person is told about some everyday objects (most commonly via language, i.e. single sentences or longer descriptions of visual scenes or events, but sometimes visually). This then has to be used to make deductions, e.g. about the veracity of further statements, to gauge the effect on performance of variables such as the degree of mental rotation required, the number of objects involved, or the presence of misleading suggestions in the test statements/questions. Such tasks require the user to visualise or otherwise encode the spatial aspects of the initial information, memorise it, and then interpret it to solve the problem. The situation of the GIS user differs from this experimental scenario in obvious ways: for a start, users are dealing with a *still visible* spatial representation which is largely symbolic rather than pictorial, and rarely have to read or listen to a verbal description of it. In both cases, however, people interpret the information using some knowledge from long-term memory about the physical world.

One view of the way people organise and use this prior knowledge is the notion of *schemata* (which dates back empirically to Bartlett, 1932, and theoretically to Kant, 1787/1934). Given a particular topic or scenario, people's mental information about it can be matched to a hypothetical mental structure or schema, consisting of links between pieces of information, images and propositions, where the structure often has some hierarchical properties (e.g. certain generic information is likely to be recalled first, and then may lead to more specific recollections).

The concept of schemata is closely linked to a later one, that of *mental models* (e.g. Johnson-Laird 1983, 1996). A schema, in the original sense, was taken to be a generalised structure about how a given situation would tend to operate; a mental model is often taken to include more specific information about the current situation and the actual objects and events within it. In addition, mental models are assumed to underlie specific mental images, holding some kind of 3D spatial representation of a space as well as embodying abstract predicates. Obviously the two concepts are not incompatible: one's mental model used in a problem or activity could be seen as incorporating elements of one's generic schemata of other relevant situations. In fact, more recent views on mental representation (e.g. Pinker, 1998) do not distinguish between general and specific types of structure. This broader overall concept will be assumed below.

4.2.2 Representations and processes

The notion that these long-term schemata or models are used to direct and encode generic *spatial* experience dates back at least to Neisser (1976). However, psychologists remain uncertain about whether these schemata or models are built and applied only via explicit awareness and attention, or whether they happen without our awareness or control. It has recently been suggested (de Vega et al, 1996), somewhat unsatisfactorily, that our long-term general knowledge may *always* influence current processing 'automatically and potentially without subject control or awareness' [p. 214] - the notion of cognitive penetrability. If this is true, then no matter what we make a user do with a digital map, their understanding of the geographic space might always affect their behaviour, especially their selective encoding of its information into memory, even where this knowledge is not useful to the task in hand. This seems somewhat counterintuitive. However, as de Vega et al pointed out, it is not clear that cognitive penetrability holds in all situations, and more work is needed.

A complementary concept to cognitive penetrability is implicit learning - the apparent finding that when exposed to a visual array that contains an inherent structure, even though people's task with it avoids drawing attention to that structure and they remain not consciously aware of it, it is somehow absorbed and applied automatically in further behaviour. If this applied to maps then we could imagine, say, that people using an unfamiliar digital map would always 'absorb' the geographical organisation of its entities (e.g. houses along a street) even if they were only asked to count lines or edit geometric shapes. However, recent work (Wright and Whittlesea, 1998) has shown that implicit learning is probably not a passive absorption process, because the relationship between the implicitly learned information and the task at hand affects the degree of learning - hence some active cognitive process is intervening to select what is 'implicitly' as well as 'explicitly' learned. Although the experimental study in Chapters 6 and 7 of this thesis was performed before the author became aware of the implicit learning debate, it will be seen that it also partly investigates whether this task-induced influence on what is 'implicitly' learned (i.e. without actual prompting or being required in order to do the task) holds true in the context of viewing a digital map.

Drawing on the psychological literature, MacEachren's (1995) book on map use and cartography proposed an information processing model of map-based mental visualisation, in which both imagery and propositions (rules/statements, as discussed earlier) would be included. MacEachren's model considered the two processes of 'seeing that' something was in a map, and 'reasoning why' in order to categorise and interpret the information. These two processes are assumed to complement each other and reflect a combined 'bottom up' and 'top down' approach to map understanding, rather than assuming an entirely schema-driven or perception-driven process. MacEachren's main interest, as for many cartography and GIS-oriented authors, was primarily in using maps for visualisation, and so interpretation of the geographic features was assumed to always take place to some extent, although, as will be discussed in Chapter 5, type and level of cartographic expertise was expected to influence the schemata applied to a given map image.

MacEachren's model was also an advancement in that it considered both the specific mental representation of a specific known geographic area, and the more generic schemata that hold a person's broader geographic knowledge: "I contend that the structure of visual descriptions derived from viewing maps will be based upon both general and specific map schemata (the latter resulting from expert knowledge or interpretation of legend information)" (MacEachren, 1995, p.49).

MacEachren's "specific schemata" would be better framed, in the eyes of cognitive scientists like Philip Johnson-Laird, as mental models (Johnson-Laird 1983, 1996); the term 'schema' tends to imply an overview or general plan rather than specific details. Johnson-Laird's recent arguments (1996) that mental models may even be a third form of mental structure *in addition to* propositions and images, and incorporating abstractions such as negations, fit with MacEachren's model even though MacEachren labelled them 'schemata'. MacEachren's model also appeared to deliberately include both the visual geometry of the map (e.g. considering a rectangle), and the geographic interpretation of it (the Missouri river, in the example he depicts), and to argue that inferences about *both* levels seem to draw on fundamental propositional and image representations.

Presumably he would also, like authors who have considered processing of pictorial stimuli (Boyce and Pollatsek, 1992; De Graef, 1992), argue that these representations determine not only *what* one expects to find in the map, but also *where* it should appear, i.e. the spatial relations between objects (e.g. houses generally don't appear in the middle of the road). Thus it might be suggested that the schema or model at least has a direct influence on the spatial *reference frame* used in interpreting the visual display, which will be discussed further later in this chapter.

Although MacEachren, like other recent authors, distanced himself from the old notion of 'cognitive maps' based solely on map-like mental imagery, he added [p. 176] that equally "...there is something intuitively uncomfortable about trying to explain graph or map understanding while relying exclusively upon propositional structures". This agrees with our earlier tentative conclusion that the knowledge that people bring to the map viewing task includes both spatial imagery and propositional 'rules'.

Meanwhile Wood (1993) discussed the way in which the 'top-down' application of prior schemata must combine with 'bottom-up' perceptual processing of the visual display, in order to reach an understanding of the space that is portrayed. For example, a pair of parallel lines may be seen and identified as a road, then combined with other visible elements according to known rules (such as that buildings tend to be beside roads, not on them) to identify the road as a country lane with a village along one part of its length. However, citing a study by Griffin (1983), Wood suggested that there is a variety of potential strategies for interpreting a given map, and that different users may adopt different ones; the role of expertise and individual differences in this strategic choice will be discussed further in the next chapter. This suggests that we must be careful to avoid trying to pin users' behaviour down to one specific sequential or parallel process in a given scenario, without identifying the type and content of prior information which is applied.

Another point made by Wood is that our 'deeper' schemata or models, split by MacEachren into 'propositional, image and procedural' representations, may be conceptually-based and not actually separable into these categories: hence information may be retrievable as *either* imagery *or* propositional *or* procedural information, as appropriate. This implies that we may not be reflecting the ultimate mental representation if we distinguish 'propositional rules' from 'spatial models' in considering the application of prior knowledge to map interpretation tasks.

The past decade has seen increasing favour for this concept of non-modality-specific long-term memory representations (i.e. *not* separate for spatial and propositional information). This succeeds some 20 years of popularity for a 'dual-coding hypothesis' in cognition, by which propositional/ linguistic information was assumed to be encoded entirely separately from spatial imagery, although the two could be associated together where they related to the same entity. Over the past ten years Marc Marschark (e.g. Marschark et al, 1987; Marschark and Hunt, 1989) has been a major proponent of the alternative notion that although some aspects of processing probably are separate and different for visual and verbal information, the underlying memory representation is in fact the same for both. Studies suggest that the long-established superiority in our recall of 'concrete' words (things we can picture) over 'abstract' words may be due to the distinctiveness of concrete words and our ability to relate them to each other, not because they produce an extra image in long-term memory which is absent for 'abstract' items. Increasingly, the propositional/imagery distinction is being seen to apply only to processing of information, not to its long-term memory storage.

We may still want to make the proposition/imagery distinction, however, when we examine the role of language in spatial tasks, since language is inherently propositional even when it is making statements about space and hence does not *present* information in the same way as a visible map (even if it may be ultimately encoded in the same way). The relationship between space and language will be discussed further below.

To summarise: long-term generic cognitive schemata, and specific mental models of the map and its represented space, are apparently used to interpret a displayed digital map. The map itself is also only a representation of the 'real' space, with its own restricted interpretation of 'real world' rules and behaviour, but the user's mental representation does not have the same flat, survey-like structure. In other words:

1. real-world experience/education is encoded into generic, long-term mental schemata or models, which may include spatial and propositional information or may reflect a deeper, non-modal, representation;

2. real-world measurements are recorded by cartographers as a digital map representation which translates real geography into a geometric array of symbols;

3. using the digital map implies interpreting the second representation by making use of the first.

The next section will examine how the structure of our mental representations, and hence our interpretations of maps and other visual displays, appears to be affected by various factors beyond the straightforward Euclidean geometry of the 'real' geographic space the map represents.

4.3 External factors influencing spatial understanding

As mentioned earlier, long-term cognitive 'maps' or models are often assumed to exhibit some hierarchical tendencies in their organisation, and to otherwise differ from the relatively exact geometric reproductions of space in 'real' maps. This may appear, initially, surprising, when the world in front of our eyes does not seem inherently hierarchical in any obvious way. By 'hierarchical', however, we mean not strict family tree-like structures, but more flexible representations in which certain kinds of information are remembered more readily, more accurately or more quickly than others, and are grouped together according to non-spatial attributes such as semantic similarity. In addition, some key pieces of information may also act as cues to facilitate recalling 'deeper' information that is less immediately important. This section describes the evidence for this hierarchy concept, and examines what is likely to influence digital map users' behaviour and spatial understanding.

The best-known writer on the hierarchical and apparently distorted nature of our spatial representations, as mentioned earlier, is Barbara Tversky (e.g. Tversky, 1992). The main evidence concerns our tendency to group entities together in our memory such that when we recall them, we recall members of the same group as being closer together and more similar than members of two different groups. We also tend to judge spatial relations such as "...is west of..." on the basis of an object's position relative to reference points or groups, rather than on an absolute frame of reference.

Much of the work on this type of distortion, and on the apparently hierarchical nature of spatial cognition in general, has focused on people's errors in recalling learned maps or known environments, or in solving problems based on presented maps. According to Hirtle and Heidorn (1993), the spatial constructs we build of any given geographical area reflect the aspects of space which are of most importance to us (depending on our tasks and general interests). For instance, certain types of landmark may be encoded more strongly than others, so that distances between them are recalled with some kind of bias. The 'route distance' required to travel between two points causes us to misjudge the 'absolute' ('as the crow flies') distance.

These biases reflect our underlying priorities and beliefs about spaces at that scale. Their failure to reflect Euclidean geometry illustrates that the 'knowledge' we bring to bear on any new space (or spatial representation) is not just objective geographical rules[18]. Instead, we display biases and priorities which may reflect evolutionary requirements for navigating around our environment (McDonald and Pellegrino, 1993), among other generic factors connected to the map's visible appearance, our perceptual and cognitive biases, and the task(s) we are trying to perform.

The three types of long-term knowledge which may affect the structure of our mental representations are listed below. Note that some factors are relatively objective in origin, and tend to correspond to real phenomena, yet often they could still create a subjective bias in the interpretation of a map.

- certain ***'rules' of real space***, discussed above, may have been developed despite lack of formal teaching: e.g. what Golledge (1995) calls "the first law of geography" (that no two objects occupy the same space at the same time); that houses are built beside roads, not in them, that rivers are generally only crossed at bridges (or fords or tunnels, less commonly), that railways tend to be straighter than roads (at least in most of Britain), that roads always connect to at least one other road, etc. Some of these 'rules' may be incorrect assumptions in some contexts (e.g. when Americans visit Boston or a UK city and expect streets to be configured around rectangular blocks). Other rules, e.g. about the flatness of maps failing to accurately reflect the earth's curvature and hence distorting distances and directions (i.e. the issue of *projection*), may not be learned or may be forgotten, causing inaccurate interpretations in some situations.

---

[18] McDonald and Pellegrino (1993) point out a potential flaw in some experiments on memory for maps and spaces, in that our *encoding* of a map into memory could actually be distortion-free *if* all the errors found in memory experiments occurred during the recall process. The evidence they reviewed suggests however that both the perceptual-cognitive encoding and the later recall are involved in the errors that people produce. Thus the predisposition to bias is still likely to be relevant to situations where memorisation is not involved.

- additionally, certain ***cartographic conventions*** may have been learned about common types of map; e.g. most of the displayed features would be visible above ground if one visited their location (except for specialist maps such as in geology), roads of a given type tend to be portrayed as parallel lines with a constant width and colour coding even when they actually vary in width, buildings are normally shown according to their approximate ground-plan size and shape, without showing other detail; contour lines are often omitted from navigational maps such as road atlases. Again, some assumptions about these conventions may be wrong (e.g. if a subject assumed that a blue line was a river when it was a motorway).

- different ***conventions for digital maps***: the user may have learned that with digital maps, colour choices (especially with Ordnance Survey data, in the UK) may differ radically from those of paper maps (including paper OS maps). Similarly, different combinations of features may tend to appear ('layers' of which can be shown or hidden under the user's control). There may also be other expectations, such as the possibility of extra data which other users have added rather than the mapping agency, and the likely degree of up-to-date accuracy.

In addition to these long-term knowledge types, our focus of attention in viewing a digital map may also be influenced by more external factors. As with our knowledge of 'real' space, certain map features and structures will be more perceptually *salient* than others. This may reflect a map user's current or usual task in viewing maps (e.g. route navigation, decision making), the relevance of needing to transfer knowledge from the small-scale map to the real space it represents, and other perceptual or cognitive biases, such as the human tendency to encode the vertical dimension more strongly than any other, and to refer to and process it more easily (Shepard and Hurwitz, 1985). When making decisions about a map, these factors will affect task performance; one would thus expect them to also impact on the mental representation which is derived during the viewing/interaction.

The subsections below will therefore discuss the influence of the task, of the map scale in relation to the space represented by it, and of the map's 'salient' visual characteristics. The discussion of the latter will lead into consideration of another cognitive concept: reference frames, which are hypothetical structures used in our interpretation of the spatial relations between visible objects. The chapter will conclude with a discussion of the potential relationship between the types of mental representations or models discussed above, and this notion of reference frames.

4.3.1 Relating representations to tasks

It is important to note again at this point that the present research is concerned with the development of a specific mental representation of a (usually) fairly unfamiliar displayed map, *and* with the influence of the user's existing long-term schemata about such maps, but *not* with the user's ability to memorise and recall a single map area from memory, which is where much of the literature on map use has been focused. Memorisation does not appear to be the most significant aspect of most GIS users' work (or indeed most map use tasks in general, as pointed out (Scott, 1987) in criticisms of studies by Thorndyke and his collaborators). For GIS users, the map is usually present throughout their tasks. Spatial *interpretation*, by inferring and applying appropriate knowledge, is the key.

Consider a relatively straightforward map task: identifying a specific object, relation, or route, on the basis of a verbally described criterion (e.g. "find and describe the shortest route from the school to the station"). The user must translate between the propositions of this instruction and their spatial understanding of what is on the screen. The objects representing the school and the station will need to be found, and the lines lying between them will need to be understood as roads, pavements, walls, alleys, etc.

As stated earlier, in such a task both the instruction and the visible map are interpretable only because users' schemata will hold certain 'rules' about how geographic space works (e.g. you can't walk over walls or houses, so routes have to go round them), and perhaps some schematic spatial knowledge of what it looks like (even if the user has not visited the site). Note also that the map designer's, task setter's and user's assumptions about the space and its 'rules' will all need to coincide to some extent, for successful interpretation to take place. Finally, the route-finding instruction will create a different focus of attention in the user than, say, a geometry-editing or site selection task, and we might expect them to learn and infer different things about the map as a consequence.

4.3.2 Transfer of knowledge: scale and perspective

Above we discussed the way that map interpretation utilises long-term knowledge of the 'real' space represented by the map. It is reasonable to ask whether the vast differences in scale between the two might affect this. Mark and Frank (1996) assumed, as Mark has also stated elsewhere (Mark, 1993), that cognition regarding the small-scale space of a map or computer screen is fundamentally different from cognition in large-scale 'real' space (often referred to as 'environmental cognition'). In fact, Mark and Freundschuh (1995) identified at least five 'types of space' based on other authors' typologies, although they admitted that the same concepts and metaphors tend to be applied at the different scale levels. Nevertheless, they still claimed that different kinds of spaces "are conceptualized in different ways" [p.26].

No other authors seem to have adopted this viewpoint, however. As an alternative perspective, McDonald and Pellegrino (1993) viewed maps not so much as small-scale spaces in their own right, as means of 'secondary learning' about a large-scale space. Even when considering large(r)-scale spaces, though, they argued that scale is not easily defined in rigid categories of 'large' and 'small' - e.g. paths within a single room can be seen all at once, but involve bodily navigation, and the cognitive representations developed may or may not be orientation-free (as they should be if the space was small-scale, according to some authors).

Similarly, Montello (1993) explicitly discussed scales of space in the context of the difference between environmental cognition and smaller-scale spaces such as maps, the main focus of most research on this point having been the degree to which learning an environment from a map is 'orientation free' compared to learning from accumulated experience of moving about within the actual or 'primary' environment. Learning from a map (a 'secondary' information source), may or may not free a person's cognitive map from being orientation-dependent, so that they can solve problems about routes etc. from a variety of perspectives[19]. Thus scale is seen as merely a contributing factor in perspective, i.e. ability to orientate oneself regardless of position or direction faced.

This focus on 'primary versus secondary learning' is not of great interest in the context of general digital map use, since most office-based digital map users do not have the chance to learn about the represented space from 'primary' experience (though some users do, e.g. in the military, surveying and/or wayfinding contexts). They also have only a limited (or, often, no) need to apply their 'secondary' learning later in the 'primary' space, e.g. when examining photographs, visiting or visualising that space in their work. Yet the degree to which they can achieve this *anyway*, despite only *expecting* to perform basic tasks with the map itself, might indicate the extent to which their mental representation of the map reflects geographic understanding of it, despite the limitations of the map's geometrical symbolism, its orientation-free perspective and its 'secondary' nature. This will be one focus of the experimental study described in Chapter 6: if Mark and Freundschuh were correct and the map space is conceptualised differently (at least by users not expecting to have to transfer between it and 'real' space), digital map users given a surprise post-test about the 'real' space should be unable to perform it.

Recently, a 1998 workshop explicitly considered the role of scale in spatial cognition of geographic information (Montello and Golledge, 1998). However, this workshop largely raised confusing questions without answering them: certainly, it did not offer any clarity regarding the role of scale in viewing digital maps.

4.3.3 Salience in the map display

This subsection will consider the question of *salience* in the map display: what makes some visible objects stand out more than others, and what do we learn about the visible relations among them?

---

[19] It should be noted that this orientation-free learning probably only applies to maps showing an aerial, i.e. overhead, view of the environment. Maps using this orientation, allowing consistent scale and projection, became standard in the UK in Tudor times (Harvey, 1993), but many still choose to take an angular perspective instead to illustrate prominent features; some good examples are shown in Tufte (1990).

A distinction between what could be seen as 'perceptual' and 'cognitive' salience in map interpretation dates back at least to the classic 1976 book *The Nature of Maps* (Robinson and Petchenik, 1976), which drew heavily on Piagetian concepts to distinguish between object relations based on spatial aspects such as proximity, and those based on objects' 'intrinsic character' [pp. 114-5]. The two authors argued that all types of map employed by cartographers were based on a greater or lesser emphasis on one of these two types of relation between objects in space, and the strength of each type of relation.

In more psychological terms, we could loosely label Robinson and Petchenik's so-called 'spatial aspects' as 'perceptual' characteristics, and consider people's processing of them as 'bottom-up', since they rely on basic sensory processing rather than interpretation of the objects' nature. The same authors' notion of 'intrinsic character', we could similarly consider as involving 'cognitive', 'knowledge-based', or 'top-down' processing, since it involves retrieving interpretative knowledge from stored schemata as discussed earlier.

The next chapter will consider the extent to which these two types of process may be over-simplified descriptions of people's processing during GIS tasks, and the role of individual map expertise in determining their relevance to a given situation. For now, suffice it to say that despite criticisms in more recent cognitive psychology of the notion of 'levels of processing' implied by distinguishing 'top-down' from 'bottom-up', the concept is certainly intuitively appealing when considering symbolic representations like maps, which are considerably more than the sum of their geometric parts. The notion that expertise influences whether and how we apply a 'higher-level' understanding to the map is almost a truism in one sense, since a total map novice would be unlikely to immediately 'see' a red square as a house or a purple line as a road.

It is not only expertise that affects the way in which people choose to solve problems concerning maps, however. MacEachren (1995) reviewed some evidence from studies of simple perceptual discrimination tasks of unfamiliar stimuli, where apparent 'top-down' influence *within the experiment* (e.g. from wording the question slightly differently) could be seen to influence subjects' responses. Indeed, MacEachren seems to argue that 'top-down' processes are more immediately relevant when viewing novel map displays than in viewing real-world scenes, because the latter are so dense and complex that some initial 'bottom-up' filtering is needed first just to make sense of things [p.43]:

> It may be that behaviour in the environment requires more reliance on bottom-up processes as a first sort... In the case of information graphics [maps], however, the problem context for vision is considerably restricted and it is logical that our visual-cognitive processing system can take advantage of this to make better use of expectations in directing where we look or what features we attend to.

So we should certainly not be surprised at evidence of some allegedly 'top-down' processing, i.e. some involvement of prior schemata-based knowledge, even in the most rapid and basic perceptual tasks.

4.3.3.1 Visual search studies

Meanwhile, partly in the course of examining the 'top'/'bottom' distinction and an alternative possibility of some form of parallel processing, various experimental studies of rapid *visual search* tasks have highlighted certain ways in which salience affects performance with a spatial display. Such studies tend to involve subjects having to rapidly identify the presence or absence of an item among a display, or similar simple and rapid tasks. Some relevant findings to our consideration of maps are:

1. Shapes whose outlines are incomplete or open, rather than complete and closed, are processed more slowly (Elder and Zucker, 1993) - this could be significant when viewing and interpreting a map area so zoomed-in that only parts of shapes such as buildings were visible, or where it included incomplete shapes such as partly-digitised objects or semi-enclosed property. However, genuinely incomplete shapes are unusual in maps, though one (an open-fronted barn) will be seen in the digital maps presented in Chapter 6. Arguably, one of the 'rules of geographic space' discussed earlier must be that in real life, shapes are rarely incomplete: even the open-fronted barn is only incomplete in terms of outer walls (which happen to be the lines shown in the maps in Chapter 6), but when seen from the air its roof would appear as a complete oblong. Meanwhile the incompleteness of shapes at the edge of the screen, or at the edge of a paper map,

would be interpreted by most Western adults as a complete shape whose remainder was simply not shown, but was present, and would probably not be especially salient.

2. Theeuwes (1993) has argued that so-called 'top-down' guidance of visual searching can only affect the general guidance of attention to a specific area of the display - when analysing the contents of that area, only physical properties are processed in a visual search task. We could hypothesise that this could be true for the type of experimental task used in most visual search studies, where the only 'top-down' information relevant to the task is the relatively shallow and meaningless task-specific knowledge induced by the experimenter - generally the display is of simple geometric shapes, dots or letters of the alphabet. It is unclear whether we can or need to make the same assumptions about visual searching in a more meaningful context such as a map.

3. Boersema et al (1993) and Wolfe (1994) experimented with more realistic visual scenes than the traditional sparse computer-generated displays. The two studies indicated that when the display was a real-life scene or aerial view, searching for a target was significantly slower than when the scene was a computer-generated image. This was independent of the effect of the number of similar distractor objects present in the scene, suggesting that the continuous and more meaningful nature of the real-life scene *itself* caused a distraction or impedance. One might expect a map, which is less continuous and less immediately meaningful than a recognisable picture, to cause less distraction (but perhaps more than a traditional visual search display, since some higher-level processing may interfere).

4. Dark et al (1996), studying responses to words (rather than shapes etc.), showed that both the semantic meaning of a word *and* its spatial position on the display affected the likelihood of its being processed (in tasks where time did not permit processing of the whole display). This suggests that at least for linguistic stimuli, semantic 'salience' can affect even very rapid and artificial task performance. Even if we should not generalise between spatial and linguistic processing (because, as discussed earlier, they appear to be quite separate, and some recent

neuropsychological evidence backs up this distinction), it still suggests, again, that some underlying structures of 'meaning' can affect rapid perceptual processes.

One might argue that the immediate 'goal' of the task (e.g. making a response to an arbitrarily determined shape or feature conjunction), which is all that seems to be implied when most visual search theorists discuss 'top-down' processing, is indirectly influenced by long-term knowledge. Hence that relatively low-level goal is altered such as to affect the task performance. But it could be possible that long-term knowledge more directly affects some tasks, by somehow strengthening the perceived degree of relevance or irrelevance of the visible features/objects. This could be tested by making both the types of stimuli being searched, and the reasons for their search, more relevant to longer-term 'real life' knowledge (the experiments mentioned above which tried to use more realistic visual fields still utilised very simple and arbitrary targets). This does not appear to have been tried yet within the visual search paradigm, although it would be fraught with methodological difficulties in retaining the same tight level of control as in simpler studies.

In any case it should be borne in mind that the above findings all did arise out of this experimental paradigm. It is not necessarily the case that a user scanning a digital map is performing a time-constrained visual search task for a 'target' in the same way; certainly, there is rarely a demand to perform such a task as rapidly as possible in non-military/navigational use. A rapid visual search paradigm would therefore not be appropriate in research attempting to maintain a degree of ecological validity regarding 'realistic' digital map use.

Nevertheless, overall the various biases and limitations suggested by the visual search studies mentioned above may well indicate a general cross-task tendency towards, for example, paying greater attention to Gestalt-related features such as fully-closed shapes, and more importantly to objects with greater semantic salience (meaningfulness).

4.3.3.2 Encoding biases and distortions

Much of the work on perceptual and cognitive salience, beyond the visual search studies, has focused on people's recall of displays, pictures or maps. As stated earlier, the present research aims to examine immediate interpretation of, rather than memory for, digital map displays, and so memory studies should be treated with caution. However, evidence points to some errors and distortions occurring at the point of *encoding* information, as well as at the point of retrieval, suggesting that some distortions found in memory studies can point us to likely types of salience during viewing of a map.

Barbara Tversky's (e.g. Tversky 1992, 1993) work was mentioned earlier in the context of hierarchical biases in underlying representations. Her studies also examined some of the ways in which our *perceptions* of space at various scales tend to be distorted. On visual displays, symmetry and figure-ground distinctiveness are key perceptual attributes which are both rapidly noticed and strongly recalled, sometimes to the extent of imposing greater contrast/symmetry than originally existed. For example, curves and other shapes are likely to be remembered incorrectly as more symmetrical than they actually were. This was shown to be a perceptual issue, rather than a retrieval error, when subjects were asked to draw a curve while it was still in front of them, and showed the same bias. Furthermore, it appeared that task-related factors could influence the degree of this bias, if the experimenter explicitly called attention to the symmetry or asymmetry of a curve.

The orientation of figures is also likely to be interpreted as being closer to the horizontal or the vertical (or, in graphs, to a 45° line) than they really are. Items tend to be interpreted in groups, and are then assumed and remembered to be closer in distance and similarity to other members of the group than to non-members, even when this is not the case. Items aligned with the vertical dimension are interpreted fastest and recalled best (Shepard and Hurwitz, 1985): a finding which may be linked to our upright posture as humans. Hayward and Tarr (1995) have suggested that even our everyday linguistic descriptions of space reflect these fundamentally salient factors.

The biases and selectivity in our interpretation may also help explain why work by Golledge (1992, 1995) has found people to be surprisingly poor at grasping some of the basic concepts used in geographic theory, such as 'nearest neighbour' or accurate orientation. Golledge put this down to a lack of basic spatial understanding, which seems a little implausible: how would we deal so well with the world around us if we couldn't make ourselves know where objects were in relation to each other, or which were nearest to us? An alternative and more plausible explanation, as discussed earlier, is that people have a different and non-Euclidean spatial understanding that is (presumably) based on what one generally most needs to know, and that simplifies spatial interpretation to avoid cognitive overload.

To summarise this section: the nature of the user's task, and certain perceptual characteristics within the visible map, are likely to influence our interpretation of it as are the underlying knowledge representations discussed previously. Some effect of the map's small scale and symbolic nature, implying 'secondary learning' of the real space it represents, may also be possible.

4.4 Reference frames: interpreting relations between objects

Above all, an overriding salience factor seems to be the imposition of a *reference frame*, i.e. an overall structure that helps people to define the relative positions of objects and the relationships between them. Imagine a picture containing a number of objects: say, the cover of the Beatles' *Abbey Road* album, where the four band members walk across a zebra crossing. Whether we think of a given Beatle as being in front of another, or to the left/right of him, depends on whether we consider the people in the photograph as having their own fronts and backs (an *intrinsic* reference frame), or just view the overall cover with an *extrinsic* (relative to the observer) or *deictic* (relative to the picture) reference frame. We then use different language to describe the object relations: 'in front of' rather than 'to the left of', and 'in the distance' or 'behind' rather than 'above'. Thus people's use of spatial language reflects their chosen reference frame, and can be used as an indicator to infer its nature. This section will examine reference frames, and their influence on our perception of the visible display, in more detail.

Both Tversky's work on spatial memory, and Gordon Logan's work on spatial attention (Logan, 1995) have suggested that the reference frame against which people mentally encode relative distances and positions can be influenced both by cognitive and perceptual processes: e.g. by semantic usefulness (to the task, or in general) and by the provision of a grid within the display being viewed (to provide a local deictic reference frame). Logan studied the salience of spatial relations between items among simple stimuli (dots, stars, letters and numbers), in some detail, to determine the relevance of salience and of reference frames in interpreting even these basic displays. As with the rapid visual search paradigm, the fact that such cognitive factors come into play even in such a situation should make us expect such influences in the more realistic situation of viewing and interpreting a map. It is therefore worth examining Logan's studies in more detail.

Logan's chief interest was in considering how human visual attention is directed between from one object to another, rather than to one single stimulus among distractors (which has been the focus of much other work on attention). In order to tell someone where to look in relation to the current object of attention, linguistic cues such as 'left' or 'above' refer to a relation between the objects which may be either 'basic' (not really a relation at all, since the target is described without reference to other objects, but defined as such by Logan), 'deictic' (in Logan's definitions, specifying relations in terms of another object, but depending on the observer's perspective - e.g. 'in front of the tree'), or 'intrinsic' (i.e. considering that the current object itself has an intrinsic 'front', 'top', etc. - e.g. 'behind the Town Hall').

Logan's experiments involved showing a simple display of just a few blobs, letters or digits, and making subjects perform tasks such as looking at a target beside which a cue word or indicator appeared, and making them state the colour or identity of the stimulus at the cued location. Logan argued that his results suggested that people adopt some kind of reference frame when dealing with deictic or intrinsic relations, but not with 'basic' relations; he based the latter conclusion on an experiment in which the stimuli were the digits 1,2,3 and 4, and the centre cue was one of the four digits. 1 was at the top of the diamond-shaped layout, 2 was on the middle right, 3 at the bottom, and 4 on the middle left. 1 and 4 were consistently responded to faster than 2 and 3, despite none of Logan's other experiments showing a bias to the left or top.

An obvious explanation is that subjects *did* apply a reference frame of sorts, based on the ordering of the numbers that started at the top and ended on the left, and these two locations were therefore more salient to them than the right or bottom - people were not simply switching attention directly from the centre to any old position, but were being guided by this number ordering.

Thus it seems reasonable to assume that reference frames of some sort, based on semantic knowledge about the objects themselves or at least on some kind of configurational understanding of the display, are involved even in relatively meaningless displays such as Logan's. They must then certainly be involved in the interpretation of map displays, which should involve much greater semantic richness.

Evidence from one study (Venturino and Geiselman, 1992), where subjects attempted to match radar displays to a 'real' scene, suggests that certain aspects of the display format could adversely affect their ability to match spatial relations accurately. In particular, if the spatial relations appeared distorted in some way then the matching task became much harder for subjects (though this is hardly surprising). In this study, however, the displays were very sparse radar returns, and the task was a discrimination task between spatial alternatives, so the only measure was yes/no accuracy and it was difficult to know how much real-world interpretation subjects made beyond the mental transformation between orientations.

Hirtle and Heidorn (1993) applied to maps the same distinction made by Logan, between deictic and intrinsic relations. They provided a variety of evidence suggesting the importance of such object relations in people's understanding of a 'space': e.g. people tend to recall distances between places in terms of route distance rather than Euclidean distance, suggesting a type of deictic 'observer bias'. They also tend to describe the location of objects relative to the intrinsic characteristics of other objects, rather than relative to some absolute grid or frame (even in situations where such a frame could easily be applied, such as in American-style city block patterns). Lloyd (1993) also cited studies suggesting that people learned more details of a map where sections of it were presented showing all the types of features present at once, than where presented with separate successive 'layers' of data, because again they could establish more patterns of object relations in the more crowded and varied display[20].

---

[20] This appears to contradict cartographers' assumptions that displays should be kept simple if people are to be able to interpret them sensibly, and that it is a novice mistake to display too many variables or feature types

Mark and Frank (1996) pointed out that *some* assumptions about object relations are already built into any GIS, in the data model adopted (e.g. the way that objects are represented as either continuous or disjointed lines, as points or areas, and as polygons or separate boundary segments). Thus it would be sensible, according to those (and other) authors, to base such representations on a coherent theory of spatial relations, which in turn should reflect or at least pay attention to the way in which people interpret them. For example, one immediate implication of Hirtle and Heidorn's work, in this regard, is that people might find it easier to interpret and discuss a display in which the fronts of buildings and other objects were identified, to facilitate use of intrinsic relations between them.

4.4.1 Reference frames and language

Some work has already specifically looked at the role of spatial language in the cartographic context, regarding the relations between objects. Mark and Egenhofer (1994) asked people to state their agreement or otherwise with statements such as 'the road crosses the park', with respect to various map representations of a line intersecting a polygon. Although in theory phrases such as this indicate only a topological relationship, in reality subjects took it to imply that the distance covered by the road while in the park is most or all of the distance across one of the park's major axes: in other words, that the road goes more-or-less straight through the *middle* of the park. Thus language is used to imply more subtle qualitative relations between objects than may be at first apparent, suggesting that the 'deictic' versus 'intrinsic' distinction may be an oversimplification in classifying people's understanding of relations between objects.

This linkage between language and interpretation suggests that the obvious methodology to use to investigate people's understanding of spatial relations is one that draws on their descriptions of visual displays. One such study (Edwards, 1991) involved some small-scale and informal investigations of people's strategies when trying to interpret essentially abstract line drawings. Edwards' subjects tended to try

---

simultaneously in a GIS. However, the (slightly unrealistic) task of memorising the whole geographic layout of an area, which requires enough visible features to enable encoding of relative positions, is different from cartographic tasks such as interpreting thematic variables.

to name the whole drawing, or parts of it, rather than splitting it into shapes and using shape descriptors. Prior knowledge, or suggestions from the experimenter about what the figures represented, led subjects to use more semantically-rich language based on those suggestions, and to split their description of the drawings into different segments from those perceived when no 'top-down' information was given.

Caution must be taken in placing too much reliance on Edwards' findings, since his study was informal, unstructured and used very few subjects (one of his drawings was viewed by only two people). More rigorous investigation of people's descriptions was undertaken by Denis (1996), whose interest lay in the structural interpretations underlying people's descriptions of visual scenes (and other people's representations built up from listening to those descriptions). In one study, 79 people were asked to describe a very simple and crude 'map' of an imaginary island; verbal protocol analysis suggested that one or another specific level of description were adopted by most subjects, but that a great variety of descriptions were still elicited owing to the (deliberately) unconstrained nature of the task. Where an additional instruction was given, such as specifying that the description was to be given to someone who had to make a parachute jump onto the island, there was far greater consistency between different subjects' descriptions - in particular, they were more likely to pick the same starting point (a meadow that would be suitable for landing in).

Denis's studies focused more on the strategies people choose in giving complete descriptions of areas or routes, rather than on their understanding of the stimuli themselves (which were made very obvious, consisting of simple pictorial cartoons). Since subjects were encouraged by the experimental context to provide a coherent description, rather than simply thinking aloud about what they saw, the resulting descriptions were heavily structured as more or less logically sequential 'tours' of the scenes or routes.

This leaves open the question of what we can learn from verbal descriptions about people's interpretations of a digital map display. As quoted earlier (from Pick and Lockman, 1983), a map lies somewhere between the entirely symbolic and the truly pictorial, and so one might expect people's descriptions to include a mixture of literal visual description, and geographic interpretation, with an emphasis on intrinsic as much as deictic relations.

Eliciting these descriptions can cause biases, depending on the task context: it may be that peoples' tendencies towards specific explanatory strategies, given the artificial context of the experimental situation, would prevent their verbal descriptions from accurately reflecting their developing representation of the map.

An alternative way of studying the spatial language issue is by providing the language and seeing how well people respond to it. Laura Carlson-Radvansky (Carlson-Radvansky and Irwin, 1994; Carlson-Radvansky and Radvansky, 1996) has looked at how people respond to different reference frames implied by different spatial terms, when viewing pictorial stimuli. By asking students to agree or disagree with statements about pictures which relied on different assumptions about which way was 'up', she showed that multiple reference frames can be applied with equal ease to such a display (as we surmised earlier on when considering the *Abbey Road* example). Taking this further, she showed that the choice of whether to prefer an intrinsic or deictic reference frame depended partly on whether there was a functional relationship between the two objects whose spatial relation was being described. For example, when a postman was depicted as approaching a US-style mailbox, hand stretched out apparently ready to place letters within it, a functional relationship was deemed to be depicted and the phrase 'in front of' was deemed preferable to 'left of'. Where the postman was reversed so that his back was to the mailbox and his outstretched hand had no apparent functional meaning, this was not the case. Thus the ease of applying one reference frame or another appears to depend on the semantic meaning being construed - in other words, on the perceived context.

Arguably, most of Carlson-Radvansky's spatial terms were ambiguous ones: 'left' could have referred to the postman's left or the viewer's left, and 'above' could have meant the top of a picture or whatever was directly in line with the top of a depicted animal's head. In other words, as discussed by Garnham (1989), spatial relational terms generally have more than one meaning depending on the reference frame selected.

When viewing a digital map, however, many of the terms one could use for spatial relations apply *only* with one reference frame or another (i.e. they imply either a deictic or intrinsic reference frame *and make no sense* if used in regard to the alternative). Consider words such as 'above', which has no meaning when considering the intrinsic reference frame of the 2D map features - we don't know what is 'above' the church when such a third dimension is not represented (and one can generally assume only sky or birds exist above it anyway in real life); but 'above' in the context of the screen display (deictic) can mean 'nearer the top of the monitor'. Similarly, 'behind' has no meaning if a house is interpreted only as a red square on the screen, i.e. a deictic reference frame applied to the simple visual geometry: there are only electronics 'behind' the display. But if the red square is interpreted as a house, then the square 'above' it on the screen might be understood to be 'behind' it relative to its probable intrinsic front (facing a road, normally). There is no ambiguity in these terms - they imply one reference frame and have no meaning in the other. This is also true in other spatial situations, such as tabletop spaces (Pederson, 1995). Therefore, in the digital map context, instead of just asking people to say which reference frame is more appropriate (given that there's no point in producing a non-geographically-functional nonsense map which would be the equivalent of Carlson-Radvansky's stimuli) we may be able to use one set of spatial relation terms or the other to unambiguously imply or 'induce' one reference frame: a possibility explored further in the next chapter.

It should be noted that our use of language suggests that we *switch* reference frame with remarkable ease: imagine discussing a paper map of a university campus with a colleague and saying something like "The student union is that odd-shaped building by your thumb, just on the shore of the lake, and the square thing below it is the car park next door...". Here our switching between deictic/extrinsic and intrinsic reference frames is so effortless that it occurs with little awareness and in mid-sentence: from 'by your thumb' (extrinsic/deictic, i.e. viewer-centred) to 'on the shore of' (intrinsic to the depicted geometric objects and the 'real' ones they represent) to 'below' (deictic - i.e. treating the map as a vertical 2D space although it may be lying flat in front of the viewer) to 'next door' (intrinsic, but only to the 'real' objects), etc.

Whatever we mean by a reference frame, we clearly do not mean something cast in inflexible mental stone.

4.4.2 Reference frames and mental models etc.

How do we link this notion of a 'reference frame' back to the previously discussed concepts of 'mental models' and 'schemata'? Are they the same? Is the reference frame somehow a 'lower level' representation focusing only on object relations and not on the nature of objects themselves? There are two ways of considering the answer to this puzzle, which has not yet been addressed clearly in the cognitive literature. First, both experimental (e.g. Landau and Jackendoff, 1993) and neuropsychological (e.g. Moscovitch et al, 1995) evidence have shown in the past few years that our encoding of the nature of visible objects (i.e. 'what is it?') may occur separately from our encoding of the spatial relations between them (i.e. 'where is it?'). From this perspective, it would thus make sense to keep reference frames separate from mental models or schemata which can supposedly encode the visual characteristics and semantic meanings of the geographical entities displayed in a digital map.

On the other hand, clearly at least some of the reference frames we employ are dependent on our having constructed an underlying mental model of the space we're viewing: e.g. some form of convention (encoded in a schema?) makes us treat the edge of a piece of paper that's furthest from us as the 'top', and we're unlikely to use geographically-based terms such as cardinal directions or relations like 'opposite' and 'next door' unless we've interpreted the visual display in terms of its 'real world' semantics, although this may be less true in other cultures (Pederson, 1993). Thus some of our reference frames clearly depend on the semantics of our mental models, and have sometimes been assumed to be incorporated within them, e.g. Denis (1996). Perhaps the relationship between them will have to remain unclear for now, since both concepts are currently too fuzzy to be pinned down in a specific processing model.

4.5 Summary

We have seen that long-term 'deep' geographic knowledge, cartographic conventions (paper and digital), perceptual salience and task requirements are all likely to affect our understanding of a given map. In all cases, we may thus hypothesise a similar effect on the reference frame that we impose regarding object relations within the map, and that is reflected in our spatial language describing it:

- We would expect reference frames, and hence spatial language, to reflect awareness of geographic 'rules' and spatial understanding, e.g. in using relations such as 'next door', 'in front of' and 'opposite' correctly.

- We could expect those familiar with cartographic conventions, such as the assumption that north is 'upwards' and that the map takes an overhead orientation (i.e. like an aerial photograph), to use this knowledge in interpreting it and in applying appropriate reference frames.

- Perceptual salience, in terms of issues such as clutter and colour choices, has also influenced the ease of establishing reference frames, at least in some of the studies reviewed above. One might imagine such salience to have a lesser influence on a task or description using a reference frame based on the geographic interpretation of a map, than one focusing on its geometric characteristics.

- We may also expect that the task context will cause one reference frame to be more appropriate and hence more easily adopted (e.g. if the task involves correcting the visible geometry to match that shown on a paper map, 'above' is more relevant than 'next door to', but not if the user is trying to develop possible solutions to a neighbourhood dispute).

- If reference frames and underlying representations (mental models) are linked rather than separate, then we may expect that stronger awareness of the map as a symbolic representation might make us more likely to use an intrinsic reference frame based on the 'real' geography (e.g. saying houses are 'next door to' each other instead of just 'above' one another on the screen). Turning this around, we may also expect that if we instruct users with language implying such an intrinsic and geographic reference frame (i.e. emphasising the meanings of the visible objects), they may be more likely to adopt it and to utilise/encode geographic rather than surface geometrical characteristics of the map, resulting in a better mental model of the geographic space than if the task instructions focused on surface geometry.

To summarise, therefore, we have a situation in which a number of factors may influence the mental representation we build up of a map, even when it remains visible in front of us. We have some limited understanding of the structure of such a representation, both in terms of overall perceptually encoded/cognitively inferred knowledge (mental model, perhaps incorporating underlying generic schemata), and of the framework(s) we use to describe the relations between the visible objects (reference frame). The link between these concepts is still unclear in the psychology and cartography literature, but it seems reasonable to assume that they are mutually dependent to some extent even if they are based on separate neurocognitive pathways. In the next chapter we will examine potential influences on this spatial mental processing which come not from external factors within the map or task, but from differences among individual users.

Chapter 5: Individual differences: spatial ability, strategy and expertise

As stated in Chapter 2, substantial individual differences were apparent between users' responses in the USIS survey and observation study, which did not appear to be readily explainable by differences in systems, application areas, or simple measures of user training and experience. As suggested in Chapter 4, individual differences among users are among the factors which might be expected to affect spatial cognition with a digital map. This chapter will review what we know about them. The aim of the chapter is to gradually unravel the likely individual difference factors influencing digital map use, with a view to deciding how these should be measured in any cognitive study of this. It will be seen that the subject is complex and as yet still poorly understood.

Although, as in Chapter 4, the main emphasis in the review below will be on findings relevant to users' spatial cognition with the map, it should be remembered that the map is *digital* i.e. computerised and hence people's attitudes and performance with computers are influential as well. Hence where relevant, individual difference studies and views from the HCI domain will also be mentioned in the appropriate sections below.

Individual differences are generally considered not in terms of true individuality, which would make it difficult to draw any useful conclusions, but in terms of factors which tend to group or measure across the population such as gender, race, age, educational attainment, expertise and training, and apparent personality traits. Above all, however, psychologists and society as a whole have focused a great deal of attention on the notion of intelligence, and of possible components of it such as verbal or spatial ability. With regard to these, there have now been decades of factor analytical studies, attempting to tease out which types of ability appear to be closely enough correlated to indicate underlying 'real' factors.

Major reviews of spatial and other cognitive ability measures, however (e.g. Cooper and Regan, 1982; Lohman et al, 1987; Carroll, 1993) have urged a deeper consideration of the actual cognitive processes involved in a task, to move away from the assumption that all tasks and tests that are vaguely 'spatial' will somehow measure the same few factors of underlying ability. We might expect spatial ability, whatever it is, to be a key factor in predicting map use performance, but the review in this chapter will show that this is not a simple issue. Therefore, as well as describing studies which have tried to identify individual differences in computer and map use, this chapter will try to consider the actual nature of spatial ability.

In any case it is important to remember the broader context of individual differences beyond spatial ability, not only because our understanding of digital map users would otherwise be one-dimensional, but also because other sources of variance could be found to interact with spatial abilities in affecting user behaviour. For example, it is not inconceivable that a particular spatial ability measure could be found to predict an aspect of user performance *only* when users were novices, or *only* when their level of 'computer anxiety' was low enough to allow them to perform the task with reasonable confidence and concentration. Therefore expertise, age, gender and related factors are discussed below. Age and gender, unlike spatial ability, are clearly-defined entities and need no definition: expertise is, however, as multifaceted and vague as spatial ability, and its nature is therefore discussed further towards the end of the chapter.

5.1 Cognitive factors (abilities and strategies)

Individual differences in people's performance on tasks is often attributed to factors intrinsic to their mental processing of information: the most established of these is cognitive *abilities*, often assumed to be innate (although practice can often improve people's performance on many psychometric tests). *Strategies*, on the other hand, are assumed to be the product of training or experience, and thus suitable to be improved by appropriate methods. A more recent, politically-correct and woolly notion of cognitive or learning *'styles'* seems to lie somewhere between abilities and strategies in people's understanding of their changeability: something seen as relatively stable but not reflecting a value judgement about 'aptitude'.

Starting with spatial ability, or abilities, what abilities may be relevant in the map context, and how can we identify them? The first point to bear in mind, which will be discussed in more detail later, is that we as yet have no very clear definition of the components of spatial ability. Decades of factor analysis-based research in psychology have failed to yield a set of convincing and consistent subfactors (Lohman et al, 1987; Carroll, 1993). Therefore the best we can assume, when we consider abilities, is to tentatively guess that where one so-called ability 'test' score correlates highly with performance on an experimental task, the two tasks are both measuring at least one common cognitive factor. This is different from assuming that the experimental task therefore depends on exactly the combination of skill(s) involved in the so-called 'test'. This was pinpointed by Philip Johnson-Laird, a famous cognitive psychologist, during a discussion on questionnaire measures of visuospatial abilities:

> I always find it ironical that people doing individual differences research... somehow feel that it's the questionnaire that's predicting the experimental result, whereas it seems to me that the experimental result is telling you what the questionnaire is actually measuring, if you're lucky. [de Vega et al, 1996, p. 215]

While psychologists continue to debate the nature of 'fundamental' abilities, as far as geographers are concerned spatial ability focuses on the practical skills required to 'do' geography (Golledge, 1992):

> Spatial abilities include: the ability to think geometrically; the ability to image complex spatial relations at various scales, from national urban systems to interior room designs or tabletop layouts; the ability to recognise spatial patterns in distributions of functions, places and interactions at a variety of difference scales; the ability to interpret macrospatial relations such as star patterns; the ability to give and comprehend directional and distance estimates as required by navigation, or the path integration and short-cutting procedures used in wayfinding; the ability to understand network structures used in planning, design and engineering; and the ability to identify key characteristics of location and association of phenomena in space. This definition extends beyond that usually found in discussion of spatial aptitude tests... [Golledge, 1992, pp. 5-6]

Obviously, since Golledge was talking as a geographer not a cartographer, these abilities require clarification with respect to digital map use:

1. "the ability to think geometrically" presumably includes the ability to understand and navigate around a visual display such as a map, to calculate distances and use basic geometric 'rules' such as the notion that three points in space form either a straight line or a triangle, and that 4 successive right-angled turns in the same direction tend to lead back to the starting point (if paths are reasonably straight).

2. "the ability to image complex spatial relations at various scales": as we suggested in Chapter 4, we may not need to form mental images of a 'real' geographic space in order to perform many tasks with a map of it, but we may need to at least grasp

complex relations *within* that map (e.g. notions of neighbourhood and intrinsic front/back/opposite relations; symbols which indicate the third dimension such as contours, height markers and roof apexes; etc.).

3. "the ability to recognise spatial patterns" - obviously relevant to maps designed to illustrate distributions of phenomena, but also important in picking out roads, waterways, towns versus rural areas, and e.g. identifying housing estates from particular patterns of house shapes/sizes/relative locations.

4. "the ability to give and comprehend directional and distance estimates as required by navigation, or the path integration and short-cutting procedures used in wayfinding" - on one level, arguably only relevant to the use of maps explicitly for navigation/wayfinding. On another level, being able to trace a path through a map may be a task within, say, an urban planner's decision-making process. In the case of a digital map there is also another 'navigation' issue: finding your way around a map which extends way beyond the screen boundaries. For example, visible 'landmarks' within the map display (such as a particular configuration of lines) may become crucial to finding your way back to viewing an area you'd seen earlier.

5. "the ability to understand network structures used in planning, design and engineering" - crucial to GIS applications such as the utilities and highway maintenance, but possibly irrelevant to most other digital map users.

6. "the ability to identify key characteristics of location and association of phenomena in space" - this seems to the present author to be covered by numbers 1, 2 and 3 above.

In contrast to this list, most spatial ability tests in use in psychometrics have focused entirely on either:

- Mental rotation: e.g. the famous 'blocks' test (Shepard and Metzler, 1971); also simpler tests which only depend on 2D rotation such as the Cards, Flags and Figures tests (Thurstone, 1938).

- Visualising how things might look from another (non-rotated) perspective or after something is done to them, e.g. after moving in space, or after paper is unfolded; e.g. Form Board and Paper Folding tests from the ETS kit (Ekstrom et al, 1976).
- Spotting specific shapes or patterns among larger, more complex figures, e.g. the Embedded Figures Test (Witkin et al, 1971).

Later we will consider the supposed underlying spatial ability factors that such tests have suggested. For now it is worth noticing that none of them is particularly close to Golledge's broad notions of spatial ability, although all might be relevant to certain tasks within map use. Of the three, the first is important for identifying a map object such as a building from any orientation; the second would be useful in visualising planned changes to an object or area, and would also be helpful in interpreting contours and other symbols; the third is obviously useful in picking out desired features from a cluttered display and for identifying patterns.

However, most tests fail to cover the ability to notice novel patterns or objects, as opposed to picking out prespecified patterns, and they also have little to say about changing reference frame (interpreting the visual display in more than one different, structured, way) which was discussed in Chapter 4. There are no well-known standardised tests on describing or identifying routes or network structures, other than simple maze-solving tasks, and certainly the present author knows of no standardised test on visualising a geographic environment from a small-scale map. Psychologists could argue that such tasks *include*, as subtasks, the key abilities tested by the standard spatial ability tests, but since they also include other subtasks as well they may not be sufficient as explanations of behavioural differences.

5.1.1 Abilities and strategies in map use

This problem of the apparent failure of standardised psychometric tests to cover all the aptitudes relevant to map use has been emphasised by the findings of studies which attempted to correlate such test results with people's performance on map-reading or map-learning tasks. Such studies have also shown the importance of strategy in performing an appropriate sequence of actions, besides the aptitude with which the actions are performed. These studies will now be briefly reviewed.

The classic and most quoted study of map-reading or map-learning ability was performed by Thorndyke and Stasz (1980). These authors found that different people used different strategies when interpreting and memorising information from a map; all subjects used some verbal learning as well as spatial, but spatial recall was variable and seemed to depend on whether subjects encoded spatial location and enhanced this knowledge by noticing relationships or shapes. The authors found it was possible to improve people's performance by training them in appropriate techniques. However, there seemed to be some aspects of learning information from a map which were not improved by map-reading expertise or training; also, tests of general visual memory ability suggested that the highest-ability subjects benefited more from training than those with lower ability.

Thorndyke and Stasz's study, while focusing heavily on the relevance of visuospatial abilities, reminds us that the learning process is never *purely* visuospatial. In reality maps include a large amount of information which could be described as propositions using language, or which is textual in itself (such as place names). Furthermore, maps often accompany other information, chiefly in the form of text, and are intended to supplement it and integrate with it. It should thus be no surprise that a number of studies have shown an important role for verbal abilities and/or strategies in effective map learning. Sholl and Egeth (1982) found unexpectedly that mathematical and vocabulary tests, which they had included in a map-reading ability study only to contrast with the spatial ability variables, actually predicted map-reading performance better than the spatial tests they included (although there were a number of confounding factors in their experiment, and they had not necessarily chosen the most suitable spatial ability tests anyway; a recent study by Tkacz (1998), suggests that some spatial ability tests *do* correlate well with map interpretation tasks, though unfortunately Tkacz fails to cite exactly which tests she used).

More recently, Diana and Webb (1997) found that although providing a map to supplement a text generally improved children's learning of it, the effect was less significant for students with high verbal ability (who could presumably learn the text quite effectively without help from the map), although spatial ability made no difference[21]. Similarly, in some work with adult Army personnel (Schofield and Kirby, 1994), although spatial ability was a key factor in learning the features of a topographic map, teaching subjects to use a consistent verbal strategy to memorise the features dramatically improved performance. They commented:

> These results indicate that most subjects usually approach the task of map reading in a relatively passive way, without any coherent strategy, and then attempt to encode information using inefficient spatial codes. When a simple and effective verbal strategy was supplied, most subjects were able to adopt it, with uniformly beneficial results. This suggests that performance on many spatial tasks could be improved by verbal strategy training.

Studies such as these show that at least as important as ability measures is the role of people's learning strategies. Audet and Abegg (1996) found that relatively novice students can take a haphazard, visually-based, trial-and-error approach to problem-solving with digital maps, which improves with experience; meanwhile Schofield and Kirby's (1994) study showed that verbal strategies can facilitate map recall, and Thorndyke and Stasz (1980) showed that training in visual strategies could also help (at least for people with good visual memory). Griffin (1983) showed that students performing a task which involved matching a cartogram (with the area of each subregion proportional in this case to its population) with a corresponding geographic map (where the relative areas of the regions were geographically realistic) adopted one of several strategies for each component of the task, and tended to be consistent in their use.

This was supported by some studies reported by MacEachren (1995), who argued that various problem-solving strategies can be used both in comparing and learning maps, and that individuals differ in the degree to which they "organise map-derived knowledge in an analogue versus a propositional (and/or maybe a procedural) form (or in the tendency to retrieve knowledge in that form)" [p.173]. Here MacEachren's deliberate parenthesised alternatives remind us of the debate in Chapter 4 concerning the nature of the underlying representation - he seems to be suggesting that one

---

[21] The authors also noted that verbal ability and general overall intelligence were highly correlated: this is a common finding in individual differences research and is related to the way we define such abilities: most measures of general intelligence draw strongly on verbal or at least verbally described tests. Even where the test involves spatial abilities, the subject has often had to read or listen to quite complex verbal instructions to know how to complete it.

source of individual differences, whether deliberate or without awareness, is the extent of our reliance on retrieving one form of information rather than another (remembering the suggestion that the underlying representation may not be in itself either propositions or imagery, but something which can be used to construct either).

This is reminiscent of some of the claims made for the *hemisphericity* concept (see Springer and Deutsch, 1998), i.e. that brain asymmetries make people more or less 'verbal' or 'spatial' in their problem-solving abilities. The debate surrounding hemisphericity and its implications is too complex and uncertain to discuss here: we can only raise the possibility that if hemisphericity affects the way in which spatial knowledge is encoded or retrieved, then certain tasks may be affected by individual differences in it. However, since it is likely to interact with other factors such as gender, and to be ameliorated by expertise (either through training in appropriate strategies or by experience of having to perform spatial tasks that demand particular processes/knowledge), its significance in predicting people's performance may not be clearly demonstrable.

The possibility that strategies can differ, and can be more or less appropriate to a given display, was explored Goh and Coury (1994) with regard to various (non-map) digital displays such as shapes, graphs and numbers. They claimed that the verbal protocol analyses of subjects trained to respond to 'normal' and 'failed' system states (as if in a process control scenario) illustrated two common strategies: an 'analytical strategy' with separate processing of individual display elements, and an 'emergent feature strategy' in which the display seemed to be seen as a whole, i.e. "a specific cue from the display is processed as a single perceptual unit without decomposition into individual parts" [p.734]. Subjects adopting the latter strategy tended to show faster response times, and the authors suggested that the salience of the 'emergent feature' (the overall aspect of the display which indicated change) would be the key aspect in building an appropriate cognitive model to explain the effect.

Goh and Coury's 'emergent feature strategy', which was adopted consistently by some subjects and sporadically by others, suggests a 'top-down' approach to interpreting a display, i.e. the involvement of a learned cognitive schema rather than a set of perceptual analyses and deductions. However, Lloyd (1993) argued that the distinction between 'top-down' and 'bottom-up' processing is too simplistic, and instead drew on Cave and Wolfe's alterations to Treisman's 'feature integration' theory of attention (Wolfe et al, 1989; Cave and Wolfe, 1990). Cave and Wolfe argued that attention to a display involved an initial stage in which both top-down and bottom-up processing were happening in parallel[22], so that expectations are retrieved at the same time that initial sensory information is being provided. In addition, they argued, these initial parallel processes guide the next stage of processing, which proceeds to serially examine salient features of the display.

In the case of Goh and Coury's experiment, the experimental task was probably too artificial and context-specific to involve much top-down processing at any stage (at least, not from generic experience-based schemata about geographic space), and this is true in fact for the sort of artificial stimuli used in Cave and Wolfe's experiments, and similar work on the visual search paradigm in cognitive psychology. Lloyd's discussion of map interpretation of course invokes a far greater role for users' general experience and expectations. His interpretation of Cave and Wolfe's theory includes the suggestion that if subjects know nothing about what to expect (i.e. if the map display is completely new to them and they don't know what colour/symbol conventions may apply), bottom-up processing tends to predominate; if subjects are already familiar with the type of display and know what they are looking for within it, top-down processing will apply, with irrelevant stimuli being automatically ignored. This possible effect of expertise on strategy, and alternative ways of considering the issue, will be discussed further in the section on expertise later in this chapter.

This discussion of strategies demonstrates that even if there is such a thing as underlying 'spatial ability' factors, they may interact with other individual difference factors in ways which are unclear. The underlying ability may make it more likely that a particular strategy will be adopted in a given task context, or it may be independent of strategy and be a 'purer' measure of some aspect of processing capability (which is

---

[22] The use of the word 'parallel' in theories of visual search like Cave and Wolfe's is intended to imply that the scanning of the information itself takes place as a parallel process (as opposed to serially, viewing one element or area at a time). Thus the top-down and the bottom-up processes in the initial stage are *each* parallel, according to Cave and Wolfe, as well as running simultaneously.

what is generally assumed in psychometrics). We turn now to a more detailed look at what spatial ability has appeared to be when examined in psychometric studies.

5.1.2 Abilities and styles: the factor-analytic approach

For most of this century, individual differences research in psychology has been split into two main areas of interest: personality and intelligence. In both these areas, psychologists have spent decades attempting to identify valid dimensions along which humans can be placed through measurement, so that every individual could be represented as occupying a specific point in a multidimensional space. In both personality and intelligence research, fierce debates have raged between those who postulated the existence of only one, two or three dimensions, and those who chose instead to identify many more. The most well-known of these dimensions in personality theories is probably extroversion-introversion, e.g. Eysenck (1990); in intelligence research it is probably Spearman's 'general factor', *g* (Spearman, 1927).[23]

The differences between theorists have often been due to differences in their use of the modelling tool known as *factor analysis*, which attempts to identify underlying factors from clusters of strong correlations between a number of variables. For example, if a thousand people complete a large battery of different mental ability tests, and their scores on three of those tests are highly correlated with each other but not with any other tests, we may hypothesise that the three tests are measuring the same underlying factor. When *all* tests show some positive correlation, however, the number of underlying factors that we may identify depends on the mathematical parameters we choose for isolating each factor, and also on the degree to which we will allow those factors to correlate *with each other*. 'Orthogonal' approaches to factor analysis assume that only factors which are uncorrelated with each other can be identified as unique; 'oblique' approaches allow factors to be identified which themselves have some correlation with each other (because the cluster of variables

---

[23] It may seem ironic that the only area of psychology which focuses on differences between individuals has concentrated most of its efforts on grouping them together using as few dimensions as possible. This is certainly true of the research using factor analysis, in both personality and intelligence. While this approach has always dominated, in both areas alternative approaches have been suggested: e.g. Kelly's Personal Construct Theory of personality (Kelly, 1955), and the more recent cognitive and psychophysiological approaches to abilities that are discussed further in this chapter. Perhaps there is always a tension between our desire to protect our own individuality, and our attempts as behavioural scientists to produce universal theories.

which each represents are themselves quite strongly correlated with each other). It follows that oblique factor models claim to identify more dimensions of behaviour than orthogonal ones.

The differences are further complicated in oblique factor analysis by the possibility of producing 'higher-order' factors, by considering clusters of correlations between the factors identified in the initial analysis (the 'first-order' factors). These further analyses tend to support a hierarchical view of behaviour, in which one, two or three 'general' factors are underlying all the behaviour that has been measured by the tests, but at a lower level of analysis there are several 'domains' of behaviour. In the abilities debate, these 'domains' tend to be given labels such as 'visualisation', 'language', perception' or 'creativity'; while the higher-order factors are given labels such as 'crystallised' versus 'fluid' intelligence, or a single higher-order factor may be postulated ('general' intelligence).

The debate is now several decades old, but one recent re-evaluation of it is perhaps the most important contribution yet. Carroll (1993) published a painstaking and impressively encyclopaedic review of more than 460 factor analytic studies, in which he used researchers' original data where possible to recalculate factors and draw an integrated picture of their findings. He concluded by suggesting a 'three-stratum theory' of cognitive abilities, which was similar to previous suggestions of a 'hierarchical' view: one single general intelligence factor, underpinning all measures of mental ability, contributing to at least eight broad types of ability. These in turn predict behaviour on various standardised mental tests which have been designed to measure apparently unitary factors.

Carroll's theory also provided a potential way forward in the other major debate in this field, that of 'nature' versus 'nurture' and whether ability is inherited or a factor of the environment. As in the 'one factor or many?' debate, this debate has also been subject to compromise in recent years, with the assumption that both genetic and environmental influences have a bearing on intellectual development. Carroll cited Cattell (1982) in suggesting that some of his specific factors may be genetic, while others may be environmentally-induced. One might expect the latter to be more amenable to improvement via training or experience, while the former could influence the degree to which such training would be effective for a given individual (i.e. their ability to learn new skills in a specific domain). This would imply that some factors in cognitive ability were actually more 'fundamental' than others, which fits well with the 'hierarchy' view.

However, while we remain uncertain about the classification of spatial abilities, as will be seen below, it is impossible to make judgements about what causes a given one, because we can't be absolutely sure that it exists: in other words, we are prone to circularity in our consideration of what is 'fundamental' to behaviour, because all we ever have are observations of patterns rather than physical objects. However, this will hopefully change as cognitive neuroscience progresses our understanding of the link between spatial task performance and the brain's anatomy and physiology (e.g. Gothard et al, 1996; Goel et al, 1998).

In Carroll's discussion of the implication of his proposed model of abilities (Carroll, 1993, chapter 16) he accepted that his theory is rooted in previous similarly hierarchical theories of intelligence, dating back as far as Spearman (1927). However, Carroll's version has the advantage of sophisticated modern factor analysis techniques, due partly to advanced computer technology, and of having a vast collection of data to work on, collected painstakingly by researchers over more than sixty years. Such a hierarchical model may seem intuitive to us: we talk loosely and generally about 'intelligence' and 'stupidity' in everyday life, as well as more specific abilities (e.g. "I'm no good at languages", "I'm a hopeless navigator"). Yet Carroll's mammoth contribution to the classification of abilities illustrates that some of our everyday constructs (e.g. 'mathematical ability') are too vague and are not clearly supported by evidence in the factor analytic research: instead, a combination of higher-order factors seems to contribute in varying degrees to any given mathematical activity.

The problem with Carroll's analysis, as with all factor-analytic studies of cognitive ability tests, is that it can only tell us about the factors which appear to group together the behaviours exhibited in performing those tests. In other words, since, as we stated earlier, most test batteries contain limited tests of spatial behaviour, focusing largely on mental rotation, pattern visualisation and otherwise identifying abstract geometric figures (or sometimes rotated/reflected letters of the alphabet). Whatever we can say about what such tests have in common, we have no reason to assume this is all that there is to spatial ability in more realistic contexts. However, it does provide a starting point.

Carroll suggested (p. 643) that the specific pattern of abilities possessed by an individual, i.e. that individual's specific factor scores, may affect not only their final score on a given mental test but also the approach or strategy they use to solve the test problems. In other words, as mentioned earlier, he appeared to agree with the suggestion that cognitive 'styles' and 'strategies', where they can be identified, are strongly related to certain underlying abilities rather than being fundamental variables in their own right in any way. Nevertheless, as Lohman et al (1987) pointed out, the fact that people do use different strategies in solving the same problems, even with the artificially simplified tasks in aptitude tests, makes it difficult to interpret them as pure measures of any single ability no matter how carefully they are designed. Factor analysis may be able to only partly help in teasing out the different strategies, e.g. if a single test seems to load on more than one factor. But if, say, test instructions sometimes misled people into adopting a strategy which was severely ineffective and only ever produced low scores, regardless of their 'innate' ability to apply the right strategy, then all the factor analysis in the world (which depends on correlations between scores) would distinguish the strategy from simple low aptitude.

Style is a different matter, however. Like the extroversion-introversion dimension in personality research, field independence (FI) is the most common theoretical dimension explored in investigations of cognitive strategy or 'style'. This may be due to its intuitive appeal, its relative ease of measurement, or its tendency to yield significant results. Typically, this construct is measured using a paper and pencil test such as the Embedded Figures Test (Witkin et al, 1971), in which subjects have to identify a simple shape or pattern hidden within a more complex one. Field independent people are those who perform well on such tasks: their perception of the test shape is relatively independent of the field (background) within which it is displayed.

Carroll provided some evidence for one ability factor which he called 'Closure Flexibility', which he identified with the popular so-called 'cognitive style' of *field independence* (FI). He pointed out that such a factor is closer to an ability than to some kind of 'style' since "it does not represent a dimension along which individuals can choose to operate at any point, as might be true of a cognitive style" (p. 556). In other words, one is good or bad at FI tests, due to an underlying ability which may be either specifically spatial or perhaps broader (e.g. a general speed-of-processing factor: certainly there is much evidence showing a strong correlation between general intelligence scores and FI). One does not *choose* to fail to see the hidden pattern in the test figure. This point was also made (Hockey, 1990) in a useful and highly critical review of individual difference studies in HCI. Besides pointing out the apparent lack of test-retest reliability found in FI measures, he added (p.114): "If style is truly independent of ability, it should be possible to find task situations in which the style identified as 'field dependent' (FD) should be more rather than less effective than that used by FI individuals" - but none have been identified. (As an aside, however, Carroll conceded that the dimension may still be of interest beyond the domain of ability assessment, because it appears to correlate with certain aspects of personality.)

On other so-called 'cognitive styles', Carroll found only patchy evidence for their existence and finally concluded (pp.559-60):

> The overall impression presented by these results ... is that cognitive styles have not yet been well established and differentiated, and that most putative measures of cognitive style depend too much on speed and accuracy ability parameters. ... If we assume that people use whatever abilities they possess to make decisions and solve problems, it can be expected that the manner in which decisions are made or problems are solved will vary depending on what profile of abilities is present in an individual. Referring to different modes of behaviour as resulting from different 'cognitive styles' is merely a manner of speaking; there is no necessary implication that cognitive styles exist independently of profiles of ability.

Yet theories such as Carroll's still leave us with the tantalising question of what his ability definitions, based on factor analysis, do *mean*: if not aspects of personal choice or expertise, then what are they? Is there any biological substrate for each one in the brain? So far this is not sufficiently clear. Could they merely be mathematical artefacts? Carroll argued that many of his factors cannot be dismissed as such, because they reflect real differences in people's behaviour: "The fact that it is difficult to specify the precise physiological sources of such differences does not make the corresponding factors any less real" (Carroll, 1993, p. 642).

As for linking identified ability factors to general cognitive theory, Carroll suggested that the use of factor analysis to determine which factors are involved in specific tasks, and how different tasks differ in their involvement of ability factors, may give clues as to *how* the factors operate within the tasks. The present author finds it difficult to see this as a very productive approach to studying cognitive processes: the discovery that one task loads onto certain postulated ability factors, while another does not, tells us little about how the task is achieved in the brain until we understand the nature of those factors themselves (if it did, then we would presumably by now have a very detailed understanding of exactly how people process the questions in mental ability tests, since these have been factor-analysed ad nauseam). We still don't know what is going on, second by second, in the subject's mind.

A more promising approach is hinted at by Carroll's suggestion that different factors represent different *types* of measure: some measure speed of a process while others measure some kind of 'level' of performance; and some of the mental tests that appear to measure certain factors depend on declarative knowledge (e.g. about the individual's native language) while others depend on procedural knowledge (e.g. how to perform mental arithmetic). Again, some tests may depend more heavily than others on the capacity of the individual's working memory.

However, Carroll deliberately avoided becoming entangled in specific cognitive theories, apparently wishing to leave the actualisation of his hierarchy in a theoretical context to somebody else. He did briefly discuss evidence for psychoneurological bases for abilities, and suggests that his theory could be used to guide attempts to reveal such neurological evidence. This would mean attempting to find neurological substrates corresponding to specific higher-order factors, rather than specific ability tests that tend to measure more than one higher-order factor to some extent. In fact, Carroll stated that in the light of his re-evaluation of ability factors, many tests in common use in factor-analytic research should now be redesigned to differentiate more clearly between different factors, and particularly between those measuring 'speed' and 'level' of performance. He also emphasised the importance of the validity of a test in a given situation: both the test and the task (or job) for which individuals are being tested should depend (have high loadings on) the same factor(s).

5.1.3 Implications for understanding spatial cognition

Carroll's hierarchical model is not intended to isolate 'spatial' ability from other factors, as the author himself made clear. Although the factor labelled 2V, or 'Broad Visual Perception', appears to relate to most of the best-known spatial ability tests, we cannot assume that other factors such as 'Fluid Intelligence' (e.g. its 'Speed of Reasoning' component) or 'General Memory and Learning' (e.g. its component 'Visual Memory') are irrelevant to performance in any real-world spatial task.

In his specific chapter on spatial abilities (chapter 8), Carroll also pointed out that it is very difficult to isolate 'pure' factors of spatial ability from factor analysing the results of a spatial ability test, because such tests tend to involve a *sequence* of cognitive activities. Hence people may have greater or lesser ability to perform any given step of the task; also, as stated earlier, subjects can choose different strategies to perform them (e.g. skipping certain questions, or performing half of the required transformation and then guessing from non-eliminated answers, or performing steps in more than one possible order). Subjects who report using different strategies can show different degrees of correlation between scores on one spatial test and another, suggesting the use of different abilities (French, 1965). It will be noted that even in the factors that Carroll identified with some degree of confidence, as listed below, the description of each factor suggests a sequence of activities rather than a single one.

However, Carroll persevered in his factor analysis of a large number of datasets gathered using different spatial ability tests. Presumably he felt that with such a large number of tests in existence, all slightly different in the tasks they set, then a large enough sample of data could still reveal commonalities that reflected 'deeper' factors and hence abilities. Certainly, Carroll showed a great deal of caution in his assignment of labels to the factors he produced from re-analysing these datasets.

He was particularly tentative about the degree to which the analysis differentiated between a single 'Visualisation' factor and a 'Spatial Relations' factor: sometimes the former did not emerge as a clear single factor, and the evidence to separate it from the latter was not unambiguous. Carroll suggested that the 'Spatial Relations' factor reflected an element of speed, and the problems of definition could only be resolved through further work separating speed from difficulty of task performance on tests which apparently measured these factors. This in fact appears to be a frequently-occurring issue in psychometric tests, and one which has been the subject of much debate (e.g. the possible presence of a general 'speed of processing' factor which does not reflect ability to eventually solve a cognitive task).

Nevertheless, Carroll proposed that there was sufficient evidence to identify these major first-order factors underlying spatial ability tests (Carroll pp. 362-3):

> *VZ: Visualisation:* Ability in manipulating visual patterns, as indicated by level of difficulty and complexity in visual stimulus material that can be handled successfully, without regard to the speed of task solution.
>
> *SR: Spatial Relations:* Speed in manipulating relatively simple visual patterns, by whatever means (mental rotation, transformation, or otherwise).
>
> *CS: Closure Speed:* Speed in apprehending and identifying a visual pattern, without knowing in advance what the pattern is, when the pattern is disguised or obscured in some way.
>
> *CF: Flexibility of Closure:* Speed in finding, apprehending, and identifying a visual pattern, knowing in advance what is to be apprehended, when the pattern is disguised or obscured in some way.
>
> *P: Perceptual Speed:* Speed in finding a known visual pattern, or in accurately comparing one or more patterns, in a visual field such that the patterns are not disguised or obscured.

Carroll also considered that there was some evidence for a few extra factors, but that is was difficult to pinpoint their exact meaning and further research was needed before they could be properly identified (Carroll p. 363):

> *PI: Serial Perceptual Integration:* The ability to apprehend and identify a visual pattern when parts of the pattern are presented serially or successively at a high rate. (It would be desirable to determine whether this factor is distinct from factor CS.)
>
> *SS: Spatial Scanning:* Speed in accurately following an indicated route or path through a visual pattern.
>
> *IM: Imagery:* Ability in forming internal mental representations of visual patterns, and in using such representations in solving spatial problems. (It would be desirable to show that this factor is distinct from factor VZ.)
>
> *LE: Length Estimation:* Ability to make accurate estimates or comparisons of visual lengths or distances (without using measuring instruments).

Finally, a couple more factors were suggested by Carroll which were based not on traditional paper-and-pencil tests but on people's responses to visual 'tricks' such as the Müller-Lyer illusion or the duck/rabbit reversible figure (shown in Figure 5-1).

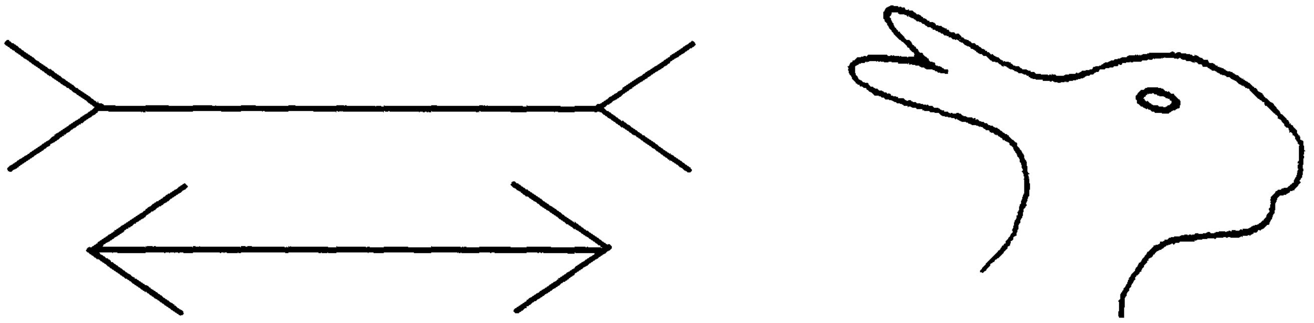


Figure 5-1. Left: the Müller-Lyer illusion, in which the lower line appears shorter than the upper although they are the same length. Right: the 'duck/rabbit' reversible figure: the two protuberances to the left can be seen as the duck's bill, or as the rabbit's ears.

Both of these types of task have been omitted from most standard aptitude test kits, possibly because of the difficulty of administration and of interpretation. Although the number of actual factors involved seems unclear, and the data incomplete, Carroll split them into two types (pp. 357-360). The following are summaries of his comments rather than direct quotations:

> *IL: Factors in the Perception of Illusions:* situations like the Müller-Lyer where subjects seem to make misjudgements based on illusory aspects of stimuli. Carroll was unconvinced about whether this counted as an 'ability' at all, but more as a 'response tendency' (the present author cannot see a meaningful distinction here). Subjects appear to differ in the extent to which different types of illusions create the misjudgements, so more than one factor tends to be found to reflect this.
>
> *PN: Perceptual Alternations:* situations where two different interpretations of a visual stimulus are possible, such as the duck/rabbit. Carroll cited some evidence that these may be related to some psychophysical measures, suggesting (for once) a neurophysiological basis to at least part of the tendency to switch views.

Carroll's final conclusion was that overall, the range of factors could not be properly understood without further research: he felt that 'procedures of measurement are in need of much refinement'. Thus his overall theory of abilities dissolves into some confusion at this lower level.

What can we then conclude about the application of all these abilities to 'real-life' tasks in the use of a digital map? The descriptions given by Carroll illustrate all too clearly that these factors were derived from somewhat artificial tasks. At first sight it takes a small leap of the imagination to see when one might find oneself (say) 'apprehending and identifying a visual pattern, without knowing in advance what the pattern is, when the pattern is disguised or obscured in some way'. However, on reflection it is normally possible to relate this back to real life: e.g. one could be assumed to perform such a task when one is looking for typing mistakes in a document one has written, particularly where the document does not consist of normal sentences. (This is of course an imperfect example, since it also involves linguistic skills, but it illustrates the point.)

Looking back at Golledge's suggestions regarding spatial ability in the geographic/cartographic context, we can try to tentatively map Carroll's factors onto these as follows:

1. "the ability to think geometrically" - this is presumably represented in Carroll's list by VZ, but only in so far as it involves mentally rotating/manipulating objects. Ability tests loading on VZ tend not to require much understanding of how individual components relate to one another within a pattern (in terms of the meaning of these relations, beyond *where* they are in relation to one another), but only whether two patterns are the same when rotated, reconstructed, reflected or transformed. Where speed might make a difference, SR, CS, CF, P and PI could all be relevant (depending on the precise task). Note that in Carroll's list, as he himself pointed out, we seem to have a lot more factors relating to speed than to eventual success, reflecting the time-limited way in which most ability tests have been administered (but not reflecting situations in real life where speed is unimportant, such as most office-based digital map use).

2. "the ability to image complex spatial relations at various scales" - IM, which Carroll expresses concern about since it may be indistinguishable from VZ, may be the nearest factor here. However, it does not involve different scales, and is mainly concerned with *vividness* of mental images, which may not be their key aspect in solving geographic/ cartographic problems.

3. "the ability to recognise spatial patterns" - partly as with (1). Note however that where Golledge uses the word 'recognise' he means 'identify as being of a certain type', not visually 'recognise' as is meant in psychometric tests where the exact

geometric shape of a known pattern is being sought. Hence Golledge's definition is more concerned with cognitive categorisation than with perceptual recognition per se: as such, Carroll would probably not count the ability as 'spatial' (in his terms 'Visual Perception') at all, or would see it as a combination of this and what he calls 'Reasoning' factors. Here is another example, then, of how cognition in map tasks inevitably goes beyond traditional spatial ability.

4. "the ability to give and comprehend directional and distance estimates as required by navigation, or the path integration and short-cutting procedures used in wayfinding" - where speed is important, Carroll's SS would presumably be critical here. LE, and possibly IL, would be important where distance estimates were being made, regardless of speed issues. One would also expect a strong verbal ability component in giving and comprehending directions (once again highlighting non-spatial aspects of such tasks).

5. "the ability to understand network structures used in planning, design and engineering" - this is again about interpreting patterns and routes, and also classifying entities. As such the factors mentioned in (1) and (3) above might be relevant: in fact it seems difficult to separate 'structure' from 'pattern' in the sorts of tasks used in most aptitude tests.

The only one of Carroll's suggested factors not mentioned above is PN, which requires more careful consideration. Although the reversible figures that produce the PN data may be seen as even more artificial than standard aptitude tests, in that we do not normally have to worry about cleverly ambiguous line drawings, they in fact have a strong link to the discussion of reference frames in Chapter 4, and hence to our consideration of interpreting visual stimuli like digital maps.

This link was pointed out in a study (Peterson et al, 1992), which was actually concerned with mental imagery and its properties. The authors were trying to establish whether a mental image of an ambiguous figure (such as the duck/rabbit) could 'be reversed' in the same way as a drawing of it (in other words, whether subjects who were asked to memorise the drawing could then, from memory only, spot the alternative interpretation of it). In considering this, they pointed out that some ambiguous figures require more of a transformation than others: e.g. with the duck/rabbit figure, the 'front' of the animal changes (in other words, its intrinsic reference frame) *and* one has to reinterpret the function of individual components of the drawing (the rabbit's ears become the duck's bill, etc.). With others, however, such as the famous Necker cube (an outline of a cube which can be seen as having one of two opposite faces foremost), the components do not change but the intrinsic (front/back) reference frame still does.

So, many of the reversible figures involve an ability or tendency to *change reference frame*. This can be linked back to the distinction made in Chapter 4 between a 'geometric' and a 'geographic' interpretation of objects on a digital map: e.g. if an oblong object that represents a house is seen as an oblong, then its uppermost line segment on the screen may be considered its 'top'. When considering it as a house, however, the 3D reality of houses means that this line is now a wall, i.e. either the 'front', the 'back' or a 'side', depending partly on its new interpretation and partly on its relation to other objects. The reference frame, and the function of the components of the object, have changed. Perhaps the ability to spot reversible figures would help a digital map user in a situation where, as in typical office situations, they were editing or examining the geometry of the display at one moment but interpreting its geography in the next? This speculation will be pursued further in the next chapter.

It is worth noting that images like the duck/rabbit figure, although slightly unrealistically drawn, do represent real objects and thus require spatial/semantic interpretation at a 'higher' cognitive level, unlike most spatial aptitude tests. Lohman et al (1987) pointed out that the historical shift in ability testing from using concrete objects to presenting abstract figures on paper was not theoretically justified, and made a big difference to the factors that tended to emerge from spatial ability tests. As Lohman et al say on page 261, "Not all that is figural is spatial... The fact that a test uses figural stimuli is no guarantee that it will require spatial skills."

However, the measures considered by Carroll were measures of how rapidly people's perceptions could switch between say the duck and the rabbit, once both had been seen. This is a different issue from being able to spot both interpretations in the first place. It is perhaps not surprising that certain neurophysiological substrates could be linked to the rapid alternations between the two interpretations: this is essentially a psychophysical task rather than a problem-solving one. Thus the actual PN factor itself is perhaps not the best indicator of performance, where maximum speed is not required.

To summarise: spatial abilities remain a complex and poorly understood topic, providing many potential pitfalls for the applied researcher. Besides the 'level versus speed' problem highlighted by Carroll (1993), other researchers have applied an information-processing approach to demonstrate that people's ability to handle spatial information is poorly represented by many psychometric tests, and can change substantially across time and context (e.g. Lohman et al, 1987). The best that an applied researcher can do, in the present climate of uncertainty over the validity of existing psychometric tests for spatial ability, is to focus on aptitudes which appear to be relevant to the real-life tasks under consideration, e.g. the interpretation of images of real-life objects.

5.2 Other individual difference factors

5.2.1 Gender

The area of gender differences in spatial ability and geographical knowledge is fraught with sensitivity. The main problem in interpreting studies in this area is, of course, the fact that every researcher has an bias (usually quite a strong one) about whether or not the differences exist, and about what causes them to appear to exist in some studies. Widely and worryingly accepted folklore states that men have greater spatial abilities than women; however, as we saw above, we are not sure how to define 'spatial abilities' at the best of times, and verbal abilities may be just as relevant even in some supposedly spatial tasks. Furthermore, a *tendency* for more men than women to excel at certain types of spatial task (and for women to excel at certain types of verbal task) does not mean that any given person of either gender will always surpass anyone of the opposite gender, or that studies will yield consistent results for different population samples. In fact, any differences found between the genders are generally less than the variance of test scores *within* each gender (Hyde, 1981), suggesting a relatively unimportant effect. All this would be true even if the findings on spatial ability differences were consistent between studies, but even this is not the case (Stumpf and Eliot, 1995).

Nevertheless, as Stumpf and Eliot discussed, gender differences are fairly consistent in some types of test claiming to measure some component or other of spatial aptitude. These tests tend to be those measuring mental rotation of abstract shapes, and performed under a time restriction so that speed is crucial. Another set of tests which tend to be fairly consistent are 'spatial visualisation' tasks, where something (again usually an abstract geometric form) is seen from one perspective but needs to be visualised from another, or where a set of items needs to be mentally fitted together. In other words, returning to Carroll's (1993) analysis, VZ and SR are the two factors showing most consistent gender differences. Such findings have, as with so much else in psychology, led to speculation about a potential evolutionary role for poorer spatial aptitude in women. Tentative findings concerning effects of some sex hormones on performance in such tasks have even led to the suggestion that it would have been evolutionarily useful to reduce women's mobility at certain stages of reproduction (Sherry and Hampson, 1997). It is hard to swallow the idea that being less able to rotate the famous Shepard and Metzler cube constructions would have stopped prehistoric woman from going out, however.

When we move away from abstract psychometric tests and consider skills and stimuli relevant to environmental space, some gender differences still emerge. They tend to focus on particular tasks which involve, as with the psychometric studies, visualisation of movement and rotation, but gender differences are not found for more 'passive' spatial tasks such as memory for object locations (Montello et al, 1999). This in fact replicates findings by Vecchi and Girelli (1998) that even for abstract lab-based tasks, dynamic path-following showed gender differences favouring men but more 'passive' spatial tasks did not.

If tasks involving static spatial interpretation and memory are not particularly favourable to men, then we should have no especial reason to see poorer performance by women on many map use tasks, particularly those favoured by many lab studies. Sure enough, the evidence is ambiguous or non-existent for worse performance by women (e.g. Gilmartin and Patton, 1984), and it appears to depend on the instructions they are given and the socio-cultural expectations placed upon them (Sharps et al, 1993; Kitchin, 1996).

The effects of even subtle changes in such instructions and expectations should not be underestimated. In one educational study, just calling a display a 'map' instead of a 'diagram' seemed to increase the amount that students learned from it (Kealy and Webb, 1995). But the reverse is often true for older women in particular: Caplan and Lipman (1995) found that older women who learned a route using a sketchmap were under certain conditions significantly worse than younger women at recalling it, although the same difference was not true for men. To explain this we could argue about the potential lack of navigational expertise of older women who may have travelled less in unfamiliar places, etc., as in fact we will do below, *but* the same study showed that the older women's performance was actually *just as good* as younger women's, *if* the sketchmap was *not* labelled as a 'map' (but as a 'diagram'). It is possible that older women may have become intimidated by the *thought* of having to interpret a map, even though they could do it reasonably well. They also performed more badly than younger women when the drawing contained more landmarks and thus looked more like a normal map, although for older men the presence of landmarks improved performance. In other words, the gender difference in this case (and almost certainly in some other studies) could be explained as a demand characteristic of the experiment, rather than an effect of either ability or expertise.

This finding is reminiscent of a number of studies cited by Brosnan (1998), where the same aptitude test was administered and described either as a 'perceptual ability' test or as a test of 'empathy'. The test was the Embedded Figures Test, measuring so-called 'field independence', often seen as a measure of 'cognitive style' but defined by Carroll (1993) and others as a spatial ability measure as mentioned earlier. Gender differences were nullified or even reversed in the 'empathy' condition, and this was largely due to changes in women's performance: women seemed to be performing according to stereotype rather than to stable cognitive traits. Despite the fact that Brosnan, following other writers (e.g. Turkle, 1984), discusses the field independence attribute mostly as a 'cognitive style' measure and uses it to partly explain gender differences in computer-based tasks, his own results suggest that the findings on this reflect social stereotyping and not some underlying preferred 'style' of thinking.

Such evidence has to be borne in mind when we interpret other data on gender differences. Kitchin (1996) studied gender differences in everyday geographical knowledge, and found no important differences. He argued that socio-cultural subtleties in some earlier studies had served to reinforce social stereotypes (which could be especially true if people were aware that they were being studied as women or as men per se).

There is also an emotive aspect to maps, which may be stronger for women. Maps may be subtly associated with moments of pressure and stress, and with problems more than with solutions, for many people (Muehrcke, 1978, p. 11):

> Maps make many of us nervous because we confront them only in emergency situations. Typically, we turn to maps when we're lost, late, frayed of nerve, and close to panic. This, obviously, is not the ideal time to nurture a love for maps... Unfortunately, however, not everyone has taken the time to learn map skills. As a result, many people are unable to move with confidence through their environment or make decisions... Realtors, tourist trap owners, and con men routinely make substantial profits by preying on people's environmental ignorance.

Certainly, women can be intimidated by maps and can underestimate their abilities with them. This was demonstrated anecdotally in a recent survey by a British campaigning group, the Women's Environmental Network, who asked members to gather information on local cases of breast cancer and on local environmental hazards, and to draw a sketchmap to show their approximate locations. Respondents were prompt and helpful, but largely reluctant to draw the sketchmaps, prompting the organisation to appeal in its newsletter (Lynn, 1998):

> Do women think they cannot draw maps? Surely this can't be true!! If bees can convey directions by waggling their bottoms at each other than we are convinced women can draw maps. We would like to encourage you to participate by asking you to take a couple of minutes to do a little exercise...Take a few minutes to draw me a map to illustrate the route [you would recommend to a visitor touring the town]. Remember, distances don't matter and I don't care what your drawing is like...

Self et al (1992) also commented on this in a major review on studies in this field. They used evidence from various studies and commentators to point out that the types of space of which women and men tend to have most experience tend to differ, and hence studies of gender differences at any given scale of space will call on differences of experience. Stretching the point slightly, though it is more likely to be true of older generations, they argued that women's environmental experience is often more concentrated at the scale of a small and simple single building, or of walkable local spaces (e.g. shopping centres), due to their spending greater time in these environments. Men may spend more time on average studying spatial configurations such as maps and diagrams, and may gain more knowledge of a wider local area, particularly through driving more frequently. But Self et al also suggested that women's tendency to be far greater readers of fiction than men may give them greater experience of imaginary environments (though perhaps less so in the age of male-dominated computer adventure games), and also may spend more time contemplating distant and exotic places such as holiday destinations. It is possible that studies of gender differences often focus on the type of task in which one gender or the other has greater experience.

Translating this to map use environments, if older women's confidence problems can be overcome, it would seem that they may even have an advantage over men in aspects of geography and visualising spaces from verbal descriptions, especially at certain scales. Therefore we do not need to delve into the 'nature versus nurture' debate, nor make unlikely conjectures about evolutionary roles, to appreciate that men's and women's *experiences* of space will partly determine their strategy in learning or using any spatial information.

We might thus expect that we will see gender gaps most prominently in tasks where women either have less experience of that scale or environment, or have less confidence in their ability to perform the task than men. Women might be just as competent, if not more so than men, at picking out (for instance) the details of a house or residential street, but may appear much less competent at route-finding. When performing a task that they do as much as men in the context of their job (e.g. in a GIS-using office environment), we should not expect any gender effects if such effects are entirely due to fear and experience. However, it is difficult to make any strong hypotheses when interpretation of performance is so difficult. Nevertheless, the tendency for women to underperform in many circumstances, for whatever reason, is common and undeniable and may be predicted in any study which does not attempt to adjust its instructions to reduce intimidation. In circumstances where subjects are faced with an unmistakable map, displayed on an unmistakable computer screen, perhaps little can be done (beyond education to dispel the myth that they 'must' be 'useless with maps').

5.2.2 Age

While there seems to have been no work explicitly on adult age as a factor in map tasks, developmental work (e.g. Blades and Spencer, 1987; Spencer et al, 1989) has shown that children develop an ability to understand maps as symbolic environmental representations by the age of four to six, and at six and above they can relate a map even to ambiguous landmarks. The concept of a map representation is therefore understood easily long before adolescence, at least in our culture. The main two indicators suggesting any relevance of age to digital map tasks in adults are the interaction between age and gender causing performance deficits in older women under certain circumstances as described earlier, the effect of age on general cognitive performance (e.g. the slowing of response times and task learning times), and the relationship between age and attitudes/performance with computers in HCI studies.

This last issue is perhaps the most pertinent to digital maps in particular. Unlike maps, computers have changed a great deal in this century, and digital computers in fact did not exist in the first half of it. The information technology boom of the 1980s, and the 1990s rise of the internet and of multimedia, have left most older people bereft of the skills and acceptance of younger generations, who are now using computers from the age of five or less. Although Egan (1988) pointed out that age and expertise can be (either positively or negatively) correlated, depending on the population under study, and so age and expertise are hard to separate, there does seem to be some evidence of problems with computer-based tasks among older groups.

Czaja and Sharit (1993) performed a detailed and thorough laboratory-based study, in which women of various ages performed data entry, file modification and inventory management tasks. The tasks were intended to be representative of typical computerised office work, and subjects were monitored for various physiological measures (mainly based on heart rate and respiration) as well as previously-validated subjective rating scales and objective performance measures. Younger subjects performed the tasks more quickly and with fewer errors, reported less difficulty and fatigue, and showed more stable respiration rates and faster heart rate recovery than older subjects. The authors suggested that the older subjects could not cope as well as the younger women with the increased mental workload in the more complex inventory management task, although the older subjects' subjective ratings indicated no greater perception of workload.

Czaja and Sharit do not appear to have considered that their experiment required subjects to perform tasks, to use a system and to handle data which were all new to them, and which were perhaps still quite novel even after initial practice sessions. Previous computing experience was measured and found to significantly affect performance measures, but not subjective or physiological ones. However, the subjects were effectively novices when it came to *these* specific tasks and *these* specific databases. Since it is generally known that older people have greater difficulty in learning novel tasks, the practice sessions may not have lasted long enough to progress from a 'novice' to an 'expert' situation. The role of task expertise will be discussed further below: the point to note here is that the Czaja and Sharit study indicates stress and performance problems for the older women *only* when performing relatively novel tasks in an experimental scenario.[24]

In a later review paper (Sharit and Czaja, 1994) the same authors explicitly addressed the issue of experience, and whether expertise in a task could offset the effects of age on performance. While stating that most available evidence appeared ambiguous, they argued that the accumulation of domain knowledge could offset older people's slowing down of cognitive functions, so that overall their *job* performance did not deteriorate. In other words, they argued that the effect of age would be smaller where

---

[24] It should also be noted that Czaja and Sharit's interest in older workers partly rested on their statement that the introduction of IT entails *increased* information processing demands on the individual. In practice, as more 'intelligent' systems replace older methods and machines, the 'deskilling' process may often *decrease* the need for operators to process information. For example, the present author once worked in a life insurance company, which had recently installed new decision support software. Its staff, who had previously been required to analyse information and make decisions about individual insurance cases, were now reduced to the status of data entry clerks. Their resultant demoralisation contrasted sharply with the satisfaction felt by the in-house system developers. It is not known whether the new system ultimately improved their effectiveness.

workers could use their knowledge to anticipate situations and necessary actions, than where information processing speed and working memory continued to be vital after the initial learning period. They also argued that as well as interventions in job design, physical ergonomics and training methods, user interface design could play a part in ameliorating the effects of age, e.g. by minimising the amount of information the user had to store in working memory while performing a task. (Of course, HCI texts (e.g. Thomson, 1985) have long specified that user interfaces should *always* be designed to minimise working memory load.)

Overall, therefore, age may affect people's ability to adapt to a novel digital map task, and may slow down their response times in problem-solving. However, there is little reason to expect experienced digital map users to show any significant performance deficit with age. As for age effects in technophobia or 'computer anxiety', which might be expected, Brosnan's (1998) review of studies of this showed mixed results, and if anything a *more positive* attitude from older people. As Brosnan tentatively suggests, younger people may feel pressurised or anxious because they feel they are expected to be computer literate, but may not be (especially those who start using computers only in mid-career). Older people may be less afraid of knowing nothing and needing time to learn.

5.2.3 Expertise

As suggested above, gender and age may be related to people's levels of prior experience of various scales and types of space, and hence the types of expertise that they bring to a map use task. Other factors that studies have suggested to be relevant include the place and culture of one's upbringing, and the subjects taken in previous education (Wiegand and Stiell, 1997). The latter factor, indeed, is too often assumed to be a measure of expertise in itself: at least one published study (McGuinness, 1994) has compared geography students with others (most commonly psychology students) and labelled the geographers 'experts' and the others 'novices' on that basis alone. The tasks included asking students (at sixth-form level) to make inferences about relationships between spatial phenomena, which involved choosing to display certain data layers in a GIS. The experimenters argued that the 'novices' (psychology students) tended to display too much data on the screen, and to use inappropriate reasoning when explaining their approach.

This approach is a rather trivialising one in terms of understanding expertise: nobody doubts that geography students have developed better strategies for handling and interpreting maps, which was amply demonstrated by McGuinness and colleagues, but otherwise we can't be clear about what exactly their 'expertise' entails. For a start, if geography students are made to perform a mapping task under experimental conditions, the experiment's demand characteristics are obvious: they may know or suspect that they have been chosen for the experiment because they are geography students with map knowledge, and therefore are aware of being expected to know and behave in a geographically literate way. Psychology students, on the other hand, will invariably wonder about the purpose and design of the experiment, may make incorrect assumptions about the purpose of the task, and may even guess that they are expected to behave more poorly (if they know geography students are also involved). Such effects of an experiment's demand characteristics have been well known since the early 1960's (Orne, 1962). It is not clear whether they were avoided in McGuinness et al's studies.

Furthermore, in studies such as these, geography students may be more skilled at *presenting* their knowledge of a map or space than psychology students, and it may be this skill (rather than the knowledge itself) which is captured in the results. Thus such studies have to be interpreted with extreme care; in other situations, sometimes geographical training seems to have no difference on people's ability to perform basic map-based tasks (Golledge, 1992).

Elsewhere, rather than geographic knowledge the key aspects of GIS 'expertise' have been defined on a more mundane level as "understanding of the database structure, graphics expertise, knowledge of the domain of the data and knowledge of the database query language" (Hearnshaw and Medyckyj-Scott, 1990, pp. 6-7). However, this assumes that users are at least having to enter and query attribute data, and perhaps manipulate the graphics as well as viewing them; in other words, this definition focuses on the HCI aspects of digital maps, as much as the users' *map* understanding. We saw in chapter 3 that these are quite separate aspects of digital map users' tasks.

However, two key aspects of expertise studies are the apparent tendency for more appropriate knowledge *schemata* to be used by students versed in geography (and related subjects such as meteorology: Lowe, 1994), and the use of more appropriate *strategies* in encoding and examining complex spatial information. These twin aspects are now fairly well-established in cartographic research (Eastman and Castner, 1983; Gilhooly et al, 1988). Recent studies such as McGuinness's (discussed above) and others (Audet and Abegg, 1996), both looking at strategies of digital map use by students with a view to the educational context, seem to confirm that the twin issues of schemata selection and problem-solving strategy are still as relevant with digital maps as with traditional paper-based cartography.

In the case of Audet and Abegg's study, rather than attempting to evaluate performance as in McGuinness's, the focus was on think-aloud sessions in which students with varying levels of GIS experience produced verbal protocols to show their apparent problem-solving strategies. As with McGuinness's study above, the use of appropriate, logically constrained, strategies was seen among 'experts', with 'novices' more likely to use trial and error or to rely on running spatial analyses rather than logical deduction to solve problems with the displayed data. In both these studies the extent of the importance of a good underlying representation of the system and data, versus strategies for solving the problems logically (which could perhaps be induced by other aspects of knowledge or experience), is not clear. However, it is certainly clear that for schoolchildren using GIS, and thus presumably for adult learners of it, expertise is about more than simple procedural learning.

As stated in Chapter 4, it has been suggested that experts are more likely to perform relatively 'top-down' processing, while novices will probably perform more 'bottom-up'-oriented processing (taking in the surface characteristics of the geometry etc.). However, another way of viewing this is that the map holds a 'deep structure' in the same way that a story or problem can be seen to have (e.g. Wagenaar et al, 1988; Dufresne et al, 1992), and that this 'deep structure' is what experts understand (perhaps via their more appropriate schemata for it)[25]. The term 'deep' here refers to the semantic content implicit in the display or map, not some imaginary 'depth' in the brain. This gets us away from the problems inherent in the 'levels of processing' notion implicit in the 'top-down' versus 'bottom-up' distinction (i.e. the tendency towards parallel processing as mentioned in the last chapter). We don't need to assume any explicit processing model in order to allow for some form of experience-driven cognitive influence on the task.

The 'deep structure' concept could however be applied both to familiarity with the geometric *patterns* of a digital map, and to understanding of its geographical representation: in other words, it might be possible to be an expert in what digital maps *ought to look like* separately from expertise in what they *mean*. We could expect GIS technicians who spend their working lives digitising, correcting and updating the spatial data to be experts at spotting, say, incomplete polygons and unusually shaped road junctions, while we might expect people with a stronger geographic background (e.g. who use maps in tasks such as wayfinding but don't have to consider their visible appearance) to be more aware of unusual building configurations and other notable landmarks. Experts who have to interpret data patterns such as contours or distributions of phenomena would see patterns in these which others would fail to pick up. Thus the content of the supposed 'deep structure' of a map or other visual representation, and hence the type of expertise required for the task, may be highly specific to that and similar tasks, rather than some general 'map skills'. However, experience in examining and spotting the deep structure within *any* type of visual image or array may be relevant to some extent, so perhaps we should check whether this type of expertise also has an effect on learning and performing map-based tasks.

This analysis seems more likely to relate to relevant expertise than a more simplistic assertion by Nyerges (1995) that declarative, procedural and configurational levels of spatial knowledge would be progressed through by novices with 'conventional' understanding of space, and that professionals in spatial information use would have a

---

[25] Thanks to Thom Baguley for raising this concept with me.

level of understanding beyond the conventional configurational level. This application of a set of knowledge levels only previously applied to environmental cognition, i.e. how we learn about a space when we're immersed within it, seems only tenuously appropriate in the digital map context. Given that children grasp spatial representations in maps from pre-school age onwards, and given that we seem to automatically organise any visual array into 'chunks' and apply semantic and perceptual categorisation to it, what kind of novice could make only declarative or procedural statements about a visual array of which they have a clear configurational overview, i.e. a map? However, Nyerges did usefully distinguish between computing and cartographic experience relevant to using GIS and digital maps, between conventional and professional spatial knowledge, and between problem-solving competence and tool competence (at handling the computer) - all these may be seen as aspects of 'expertise' which need to be teased out in any study.

Gilhooly et al (1988) examined map use 'expertise' in terms of subjects' ability to remember maps. They discussed the well-known study by Thorndyke and Stasz (1980), which was discussed earlier, which among other findings suggested no difference in recall between 'expert' (geographers, again) and 'novice' (non-geographers) map users. Gilhooly et al argued that this was due to those researchers' use of a planimetric map (in other words a map at the scale of a 'plan' of a building or campus), whereas geographers' specialised schemata would be relevant only for the types of map (contour maps, thematic maps) that they regularly encountered in their professional lives. Gilhooly et al's own experiments suggested that where a map was of a familiar type, experts tended to encode the map into memory in larger chunks, and focused more on patterns across the map than on individual features or place names.

This again suggests that a conceptual inference process is occurring that influences the way the map is understood, and that an expert schema/model encourages the user to focus on semantically task-relevant information rather than on the most perceptually salient features. Since the most common uses of GIS are probably for planimetric maps like the city streetmap used by Thorndyke and Stasz, it is possible that academically qualified geographers would not have a specific advantage of 'expertise' if made to perform typical tasks at that scale - however, urban planners might be at a greater advantage. However, we do not really know enough about whether or how experience with *any* maps, or with any visual information sources in general, could influence people's responses to a digital map task.

5.3 Summary

To summarise our understanding of expertise, therefore, we have seen that the types of knowledge relevant to a digital map task may include:

- knowledge of what is expected of the user within the task and experimental context ('demand characteristics')
- understanding the HCI aspects of the computer software and hardware, and having what Nyerges (1995) called 'tool competence' to handle it
- using appropriate schemata to selectively and efficiently process information relevant to the task; this may include 'professional' spatial knowledge besides conventional experience of space
- using a problem-solving or learning strategy appropriate to the task
- using an ability to spot relevant 'deep structure' in the map, possibly gained not only from specific map experience but from studying other types of visual image or array

Meanwhile, we have already seen that other relevant individual difference factors may include:

- spatial ability - but since the definition of this is so unclear, we may need to focus on specific cognitive components that seem especially reflective of the task at hand
- gender - but we can try to prevent 'spurious' gender effects by trying to word the task instructions so as not to intimidate female subjects or invoke social stereotypes

- age - but only in terms of slower cognitive functioning, which may be compensated for by greater experience and more positive attitudes in some situations

The next chapter will describe the design and planning of an experimental study to examine some of the cognitive processes suggested in Chapter 4, with an eye on using realistic tasks which can be related back to the analyses in Chapter 3. It should be clear by now that all of the above individual difference effects either are non-straightforward to measure (requiring measurement of a whole series of factors, in the cases of expertise and spatial ability), or else are apparently weak factors in digital map use age and gender. Hence, and since the practicalities of an extensive multifactorial individual differences study (requiring a very large number of subjects for any predictive power) would be almost impossible in the circumstances of this doctoral research, the above individual difference factors will be carefully measured and monitored for the sake of interest and for providing pointers to potential future research, but will not form the backbone of the study.

Chapter 6: Experimental study of factors influencing digital map interpretation: design and method

At the end of Chapter 4, it was hypothesised that the reference frame or model implicit in the language and requirements of a digital map task, i.e. the problems solved by the user regarding the map, would affect the way it was represented cognitively by them. This could be seen as the major hypothesis prompting the experimental work which will be described in the next two chapters. From Chapter 5 we can add a subsidiary general hypothesis that some variables measuring expertise would make a difference to the nature of this representation. Other individual difference variables, such as age, gender and some form of spatial ability measure, *might* be expected to affect overall performance with a digital map, but there is to date no evidence that they would affect the nature of the reference frame/model developed in a given spatial task.

It was decided to examine the first of these hypotheses experimentally, with an eye to the other issues if time and resources allowed. Broadly speaking, a dependent variable of 'performance' (e.g. reaction time, accuracy, correct recall scores, correct problem-solving) could be examined under different conditions which manipulated the implicit reference frame as the main independent variable (probably through use of appropriate linguistic spatial terms). Additional independent variables could include individual difference measures, and the effect of having to switch between two alternative reference frames rather than simply considering one.

This chapter describes the process of designing and conducting the experiment. The next chapter will describe and discuss the results of the experiment.

6.1 Developing the paradigm

No previous study existed which had explicitly tried to make subjects develop and use a cognitive representation of a digital map. The nearest studies, as described in Chapter 4, had tested map learning or memorisation, or asked subjects to describe a paper map. In addition, most studies had concentrated on the geographical representation of the map, rather than considering its alternative existence as a geometric display, and no studies outside the education domain had considered digital maps. There was therefore no existing experimental paradigm to draw on, so the author embarked on a process of developing one. In other words, a decision was taken to base the study on careful and thorough design, considering all likely behavioural and practical issues, and this in the event took many more months than the conduct and analysis of the experiment itself. The process is described in this section, since it formed relevant and illuminating research in its own right.

6.1.1 Selecting the main tasks

In planning an experiment to test these hypotheses, a key issue to be tackled was the choice of task. Did the hypotheses, or the desirability of achieving ecological validity, necessitate making subjects perform realistic GIS tasks, and if so, which?

The author spent a considerable amount of time examining the practical and methodological drawbacks of making subjects physically perform realistic GIS tasks. They would, almost inevitably, be unfamiliar in their precise procedures even to relative GIS experts, since all systems operate slightly differently and tend to be customised in individual ways. In addition, the performance of such tasks would introduce artefacts such as the requirement to learn new sequences of actions, the visual distraction of screen areas devoted to the user interface, the involvement of complex motor control, or an increased influence of technophobia and/or low self-confidence among non-computer literate subjects. In any case, as stated elsewhere in this thesis, user interface issues for GIS are changing rapidly to follow those for other popular software, by reflecting operating system developments such as Microsoft Windows 95, and so it was inappropriate to make these the main focus of this experiment. (As it turned out, the only GIS readily available to the author was a freeware one with which she was unfamiliar, which proved unable to work on the available computers, and which would have made it quite difficult to present simple tasks to the subjects.)

An alternative method which was considered would focus on the map itself, by making users perform simpler responses such as clicking a mouse on a named object, which could be either named according to its geographic ("the cinema on Broad

Street") or geometric ("the red square near the top right") representation and location. Here again, however, subjects' motor co-ordination and degree of familiarity with a mouse would influence results, as would distance from start to target in each trial; these factors were not of interest.

Therefore an experimental method was needed which relied on making subjects study a typical digital map and solve some sort of semi-realistic problem(s) about its visual appearance, without having to perform any physical interaction with it. Rather than do anything to alter, search or manipulate the map, the user would only have to answer questions about it. This also permitted a direct focus on the relationship between 'space' and the language used to describe it, as discussed earlier in Chapter 4. By asking questions about either the surface, *geometric*, characteristics of the map, or about its *geographic* representation, subjects would be induced to consider the visual array via different reference frames and schemata. One might expect subjects to differ in the degree to which they found the geographic or geometric questions harder or easier, depending on prior expertise and preference.

The question then was how we 'measure' subjects' representations of the map to see which aspects are most salient to them. If we accept Hayward and Tarr's (1995) and others' assertions that people's use of language will reflect their underlying representation of a space, then we could ask subjects to describe (aloud or in writing) what they understood of what was presented. Studies of people's descriptions of visual displays and maps were reviewed in Chapter 4, and at first it seemed that these would be bound to indicate the extent to which subjects adopted one reference frame or another.

Initially, it was thought that at the start of the experiment, before trying to induce any particular reference frame in the subject's mind, an unstructured verbal report in answer to an open-ended "Describe what you see" (or a more structured set of such instructions) would elicit the extent to which subjects took notice of different visual interpretations of a map display. This, it was hypothesised, would allow quantification of the relative degrees to which the 'surface' geometric characteristics such as colours and shapes, and the geographic representation of houses and roads, were noted.

However, advice from more experienced researchers made the author reject this option. Open-ended description at the start of the experiment, where subjects were wondering what was going on and doing their best to interpret the map intelligently, would inevitably lead to them treating it as a map: it was not so 'unmaplike' as to not be recognisable as such. The descriptions would probably be highly unstructured and difficult to analyse, as well as varying greatly in length, and in any case it would be difficult to draw any conclusions just from the relative frequency of geometric or geographic terms (e.g. what would we know about subjects' representation from a mixed statement such as "That purple triangle is probably the village green"?). In any case, at this stage subjects would be acting on initial impressions of the display, rather than a gradually-learned internal representation of it, so it might only indicate some form of pre-existing bias (but would be a very time-consuming and uncertain way of finding it).

Related to this problem, Carlson-Radvansky and Radvansky (1996), in their study of spatial language and reference frames discussed in Chapter 4, tried leaving open blanks for subjects to complete with any word. They found that this often failed to elicit any spatial terms at all, even in the narrow context of their specific questions, and therefore rejected such an open-ended response method. With a less constrained context like that planned for the present study, one could assume that obtaining an analysable number of unprompted spatial terms would be even harder.

Instead, the author decided to measure response-related variables such as errors and timings, which are more common and probably more reliable indicators. They also have the advantage of more *directly* (if crudely) reflecting subjects' task responses, rather than adding an extra 'task' to it (in having to verbalise).

The way in which subjects would be required to respond to the questions was carefully considered. Subjects could be asked open-ended or multiple-choice questions that required written or circled/ticked answers, but this would make it difficult to time responses; initial pre-design piloting suggested that people's error rates would generally be very low except for quite complex questions, so response time would be a more sensitive measure of performance than errors and would need to be measured accurately. Subjects could be asked to say their response aloud, but again this could cause inaccuracies and recording/timing difficulties. Subjects could be asked to hit a button once they decided on a response, and then perhaps be asked to say/write it, but this would introduce a slight degree of artificial complexity into the task (and risk them forgetting to hit the button before responding).

Overall therefore, the most simple and accurate way to measure subjects' responses, especially given the limited resources of the current study, was considered to be a two-choice response paradigm. In this method, subjects are presented with a stimulus or question, and two responses on the left and right sides of the screen. Subjects have to hit one of two buttons or keys, using their left or right hand, corresponding to the response that they think is correct. The restriction to two responses minimises complexity and hence motor errors ('slips'), although it is important to balance the correct responses between the left and right hand. Handedness (being either left- or right-handed, or neither) should also be recorded, to double-check that this does not significantly bias responses. These precautions were taken.

6.1.2 Selecting the 'post-test'

It would then be possible to test the extent to which these question types induced either a geographic or geometric reference frame/model, by administering a surprise post-test. Our major hypothesis, that the task influences the nature of the cognitive representation, predicts that subjects who had answered questions about the map's geometry, but not its geography, would perform significantly less well on a post-test requiring geographic interpretation of the display.

To perform this post-test, an obvious method is to show subjects the real environment that the map represents, from an immersive viewpoint, and test the extent to which they know where they are and where/what other objects are. This can be achieved either by taking the subjects to the actual place (impractical unless the experiment is entirely performed nearby), by immersing the subjects in a VR (virtual reality) 3D representation of the place (expensive in terms of equipment and time preparing such an environment), or by showing the subjects still photographs or videotape footage of the place. Obviously, still photographs would be the cheapest and simplest option, provided they contained sufficient unambiguous landmarks to answer the questions posed to subjects. They also had the advantage of avoiding any effects of disorientation, and minimised the likelihood of subjects being overwhelmed with irrelevant contextual information.

The subjects were to perform a task, or answer a further question, which would force them to interpret the map geographically by matching it to the objects in the photograph. As mentioned in earlier chapters, most studies which have looked at transfer of knowledge from maps to the real environment have been concerned with issues of *learning* (i.e. *memory* for) the map information. The present study is not about memorising a map, however, but about interpreting it in the context of changing task demands. This is partly for reasons of real-world task validity, since GIS users probably rarely have to memorise their digital maps, and if they ever had needed to in the past then that requirement is bound to have decreased still further with the reduced cost and increased speed of plotting or printing, and with the ever-decreasing size and cost of portable computer equipment. Furthermore, the post-test was to be a surprise, to avoid subjects deliberately trying to guess its nature and memorise potentially useful information in preparation for it. There was therefore a danger that subjects would not remember enough from the map, not having expected to recall it, to raise their performance on the post-test above chance regardless of their previous tasks. Therefore, it was desirable to leave the map visible during the post-test, so that subjects could refer to it rather than having to recall it.

To further avoid introducing artifice or complications in the post-test phase, it was decided to continue the two-choice response paradigm so that no new task learning was required.

6.1.3 Selecting the map stimuli

Digital maps are expensive, at least in the UK, largely due to the monopoly of the Ordnance Survey at most scales. The USIS studies had shown that Ordnance Survey data was almost the only background digital map data used by most UK organisations, at least at that time, and so it was desirable to obtain it. However, the cost of buying even a tiny amount of data 'off the shelf' would have been prohibitive. Under a deal with CHEST, the Combined Higher Education Software Team, UK universities could at this time obtain small amounts of sample Ordnance Survey data for seven specific areas of the UK, by paying an annual licence fee.

The data in question was listed as covering Port Talbot, Folkestone, Bristol, Glasgow, the Peak District, the Lake District and the Gower peninsula. Since the present study required a fairly feature-rich display which could generate plenty of questions about visible map objects, and since urban environments are more relevant to most people's daily life than the rural terrain favoured by many military map studies, town centre street-map data was sought. Also, if the area had to be visited and photographed, it needed to be somewhere within reasonable travelling distance for a day trip, and so the 'Bristol', 'Folkestone' and 'Port Talbot' data were examined.

These proved to be problematic, and many hours were consumed finding appropriate map sections. With help from a street atlas of the Bristol area, the numbered map 'tiles' supplied proved to cover not Bristol itself but only part of a housing estate in Nailsea, a village some ten miles away. The data was too sparse in features to be of any use. Similarly, the 'Port Talbot' data turned out to only cover part of a village called Margam, and a nearby lake. This left the 'Folkestone' data, which was loaded painstakingly into a GIS for close scrutiny since it appeared to include at least part of a couple of towns. It again covered largely rural areas, much of them recently occupied by the rail terminus and sidings for the Channel Tunnel, as well as some villages and suburban estates. None of the map areas therefore included a town or city centre. This was presumably due to the Ordnance Survey's decision not to give anyone cheap access to commercially valuable data.

Fortunately, the 'Folkestone' data also included a few villages which were clustered about a central point, typically with a war memorial on a village green or road junction. Two of these villages, Saltwood and Newington, appeared to include a number of named houses and public landmarks in this central area, which in both cases was shaped like an isosceles triangle. The triangle for Newington was smaller, in a different position, and 'pointed' northwards (towards the top of the map); the Saltwood one 'pointed' towards the left (west). This meant that though the villages were essentially similar, the visual layout and the visible features were so different that they could not be confused (i.e. one could not look at the Saltwood map and momentarily expect to see the Newington visual layout). The data for these villages was therefore chosen for use as described below. (The reasons for needing two maps are discussed further on in this chapter.)

The role of familiarity and predictability of such villages was also considered. In these respects, a village centre was felt to be broadly similar to a town centre. British people may expect a town centre to contain a town hall, cinema, shops, car park, post office or library; they may expect a village to contain a church, vicarage, village hall, war memorial, pub, green, small school (often converted to housing), village shop, or cottages having suitably 'rural' names. Although these expectations were not tested, either generally or with the chosen map stimuli, no surprisingly atypical items were visible in either map, while most of the above features were present in either one or the other (but mostly not both).

However, as discussed in depth by Campari and Frank (1995), cross-cultural issues would be pertinent here: a typical English village, along with its associated features and names, would not be nearly as familiar to non-British subjects, who might be less able to utilise their familiarity with such villages in the map-oriented tasks. In addition, the Ordnance Survey style of mapping, and the choice of features included, could also be less familiar to non-British subjects. Furthermore, the wording of the questions in the experiment generally required an extensive and fluent vocabulary (see Appendix 3 for the actual questions and for statistics on their readability).

For these reasons, it was ensured that all the subjects in the present experiment were either native British, or had lived in Britain for at least 15 years and had English as their first language (there were only two such non-natives). No subjects, during their thorough debriefing, ever raised any queries about the villages, other than curiosity about their geographical locations.

6.2 Planning the experiment

6.2.1 Question construction

The starting point for devising the main-session questions, which would induce the subjects to think more or less 'geographically' about the display, was to check that asking such questions could be related back to real-world GIS tasks as discussed in earlier chapters; in other words, the extent to which one could imagine these questions being posed in real working environments. The following questions illustrate this:

- "What's next door to (or opposite) the village hall?" - a real-life query could involve this question because the hall's owners had applied to build an extension, because they wished to extend the bar's licensing hours and neighbours might object, because the hall was being sold for development and there was doubt over the property boundary, etc. All of these could arise where a GIS was used by a local authority, for example.

- "Do you have to cross the road to get from the electricity substation to the school?" - an important issue for a utility company trying to lay or repair a cable, for instance; also for highway safety analysts in a local authority.

- "Can you see the village green from the house next to the pub?" - e.g. for planning permission purposes where the pub landlord wishes to erect a new sign or extension which could spoil a view.

- "Where does the farm track join the main road?" - e.g. for highway safety considerations, revision of property boundaries, or analysing a crime scene.

- "Which farm does the stream run through?" - e.g. for analysing environmental impact of a factory upriver.

- "Is the telephone call box right next to the road?" - e.g. for cable-maintenance purposes.

It will be seen by the reader that all of Knapp (1995)'s visual task categories with maps, discussed in Chapter 3, are present here: Identify, Categorise, Locate and Associate. So also are most of the various non-interactive tasks identified by McCann (1982): e.g. symbol interpretation, determining relative locations, etc.

Thus it seems that most such questions we can ask about a map have valid meanings in real-world work with maps and GIS. This is not only true for geographically-meaningful questions, however; many GIS operators spend many hours digitising and editing the visible geometry of the map, and have to deal with red squares (houses, or more accurately roof edges), purple curves (roadsides) and light blue shapes (boundaries, paths etc.) just like those in the experiment maps. We should not assume, therefore, that the geometrically-oriented questions will be less familiar to 'expert' subjects than the geographically-oriented ones. Questions about the deictic spatial relations between the visible objects may be as realistic in terms of real-life handling of digital maps as questions about the geographical interpretation, despite not being considered explicitly in cartographers' analyses of digital map tasks as reviewed in Chapter 3.

Accordingly, two sets of 20 questions[26] were constructed for each of the Saltwood and Newington maps. The questions were compiled very carefully to avoid ambiguity and maximise clarity, and statistical tests were performed on readability scores to check that the 'map' and 'geometry' questions did not differ significantly (see Appendix 3). For each question there was a correct response, and a 'wrong' response was chosen which was as different and obviously wrong as possible in most cases (since in cognitive experiments, it is better to aim at very high accuracy to make the response-time data more meaningful). For example, where the answer to a question was the name of a building or other feature in one part of the map, the distractor response was the name of another feature in a completely different and part of the map, so that as soon as the relevant rough location was identified the distractor could be discarded. Similarly, in the Newington 'geometry' questions, red lines were contrasted with purple ones in terms of their 'bendiness', because the red lines were mainly very obviously straight and angular (since they represented the walls of buildings). However, of course such a strategy could not be applied to all questions, e.g. those with 'Yes/No' answers, so some questions still required fairly careful inspection of the display to ensure a correct response.

---

[26] A number chosen as a balance between producing enough data for a reliable mean RT measure (and to have given Ss a thorough session with the map prior to the post-test), and avoiding subject fatigue or boredom. Since the experiment was original and did not resemble any previous studies, and since there seemed to be no suitable power analyses to help establish a suitable number of questions, this number was based on intuition and piloting, plus a practical need to keep the experiment duration to under an hour.

After piloting, further wording changes were made to some questions to eliminate some slight ambiguities, and to alter some questions which had been too open-ended (e.g. "Is there a red circle anywhere?" or "Is the village church shown on this map?" caused Ss to perform an open-ended visual search which, if the answer was negative, would be affected by Ss' strategies since some would double-check and others wouldn't).

For the post-test, it was decided that matching the photograph with the map was the main task, and there was a need not to complicate the response times by making the questions too complex. It was also felt that to absolutely ensure that the questions could not be answered from the map alone, finding the precise location of the photograph would need to be an integral part of the question, and the question would not name anything visible in the photograph that would give the subjects a clue as to the location. Accordingly, one single question was used for all the post-test trials: "What is behind the photographer in photo x?" - which forced the S to identify the location of photo x purely from the visible objects within it. Piloting showed that subjects were generally able to answer these questions, with typical response times of 30 seconds or less.

6.2.2 Physical stimuli

Several versions of the maps of Saltwood and Newington were extracted from a GIS via a screen capture program, all at the same scale (1 inch to 20 feet, i.e. 1:240) but panned across to show slightly different areas and hence combinations of features. Versions were selected which contained a suitable assortment of unambiguously labelled features in the village centres, so that sufficient questions could be constructed about them. The colour scheme implemented by the Ordnance Survey was largely retained, and the background colour was set to black. This was partly to reflect the many systems which had been set up this way in the USIS observation study, and this aspect of many GIS displays is one of the characteristics which have tended to make digital map displays quite distinctly unlike traditional paper maps, so that subjects would be less influenced by any resemblance.

The two maps were displayed at the same scale; slight editing of some features was done to tidy lines and text which had been rendered disjointed or illegible during the screen capture and image conversion process. A few further edits (mainly to text labels) were made after the experiment was piloted, to ensure consistency between the objects and labels on the map and the signs etc. in the photographs, since one or two inconsistencies slightly confused the pilot subjects. The final versions of the two maps are shown in Figures 6-1 and 6-2. The versions shown here are identical to those shown on screen as far as is feasible on inkjet-printed paper: the screen display was a laptop screen approximately the same size as this thesis, and the colours and resolution appeared very similar to their reproduction here.

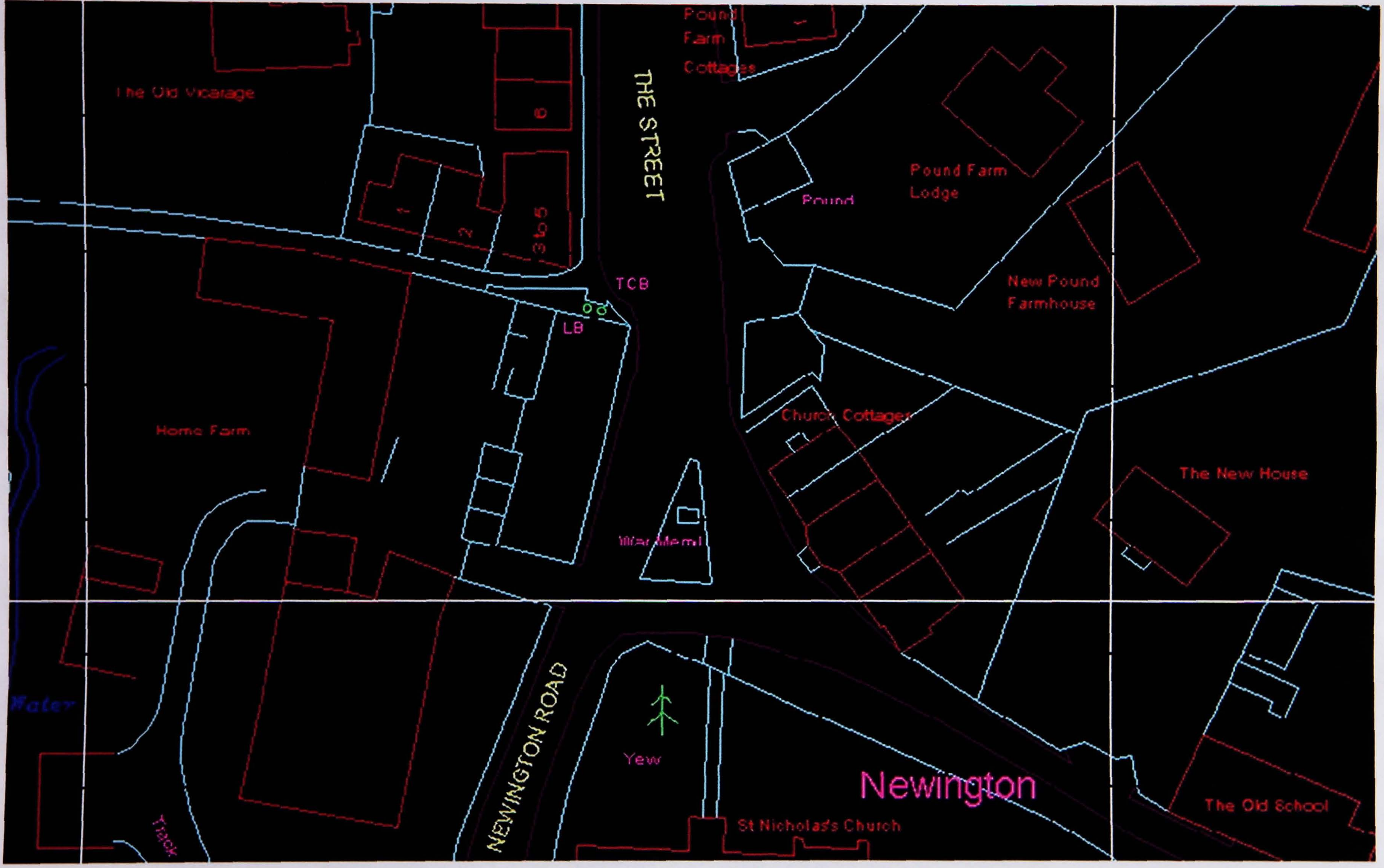


Figure 6-1. Newington map used in experiment: note the small triangular road junction with war memorial (an important orientation cue for the post-test photograph task, for both maps)

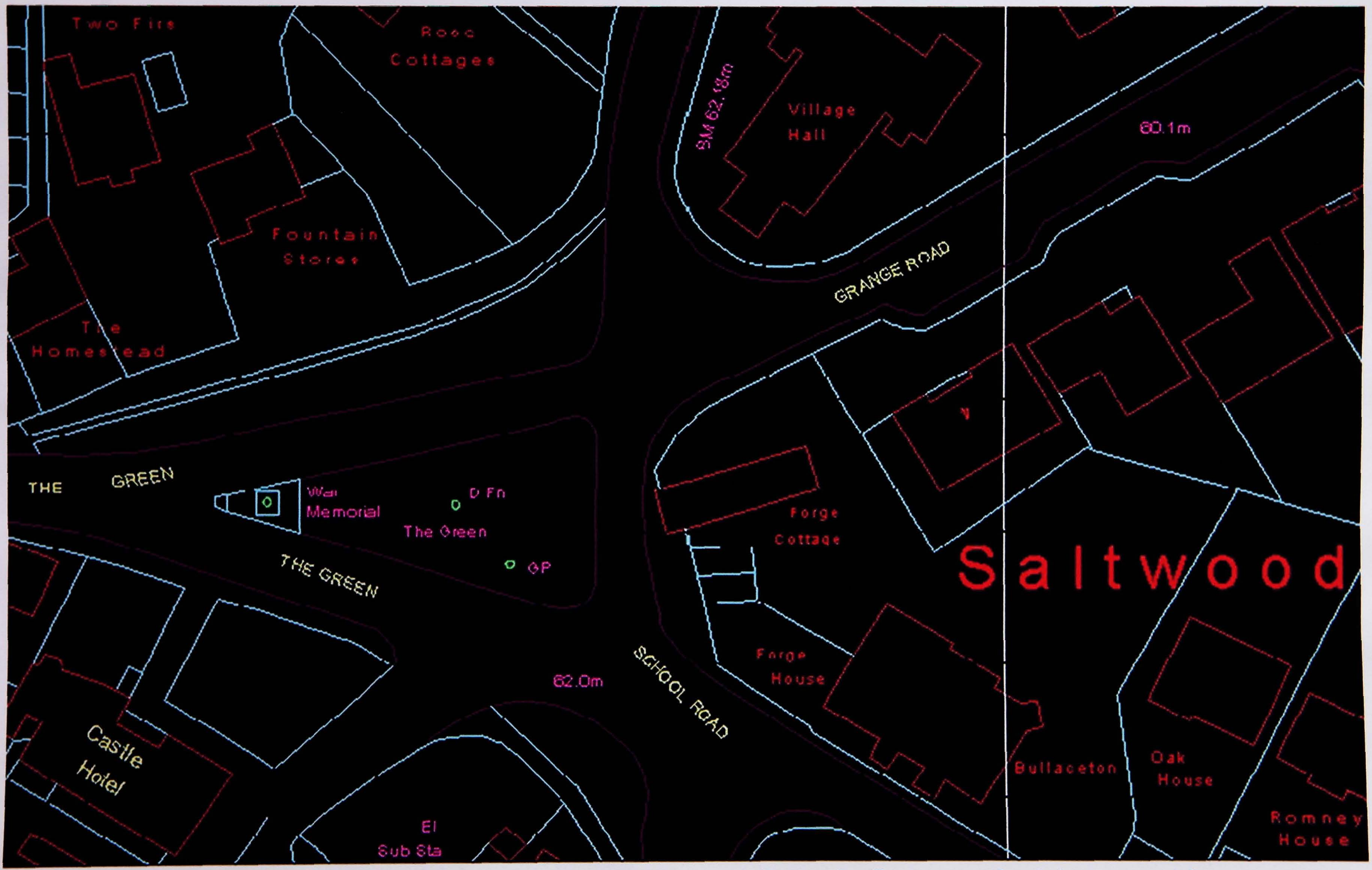


Figure 6-2. Saltwood map used in experiment. Note the purple triangle forming the village green, and again its war memorial

The experiment procedure was programmed on a Fujitsu laptop PC with a high-resolution (800x600 pixel) screen, using an experiment generator scripting language called EXPE[27] (Pallier et al, 1997). Difficulties in obtaining and retaining a laptop with a suitably large and high-resolution screen caused delays in the preparation of the experiment; eventually one was hired for several months from a commercial supplier. The programming involved thorough testing for bugs and crashes, which was complicated by attempting to use a beta release with incomplete documentation (although the generator's author, Christophe Pallier, was extremely helpful by email). It also had to run on a laptop whose physical hardware had somewhat unpredictable and undocumented display parameters, after having had to change laptop and recalibrate these in the middle of programming. For these reasons it took some two months of work to complete the program. The reliability of the generator, and the care put into the program, paid off: there were no equipment crashes while running the experiment, and not a single subject's data was lost.

For the post-test, around 20 photographs of each village were taken by the experimenter at various vantage points, on the same afternoon. Each photograph focused on a landmark which was either clearly labelled (e.g. Saltwood's 'Village Hall' and 'Castle Hotel', and Newington's 'Pound Farm Lodge' and St Nicholas's Church) or else was unambiguous (e.g. the war memorials in both villages). It was taken from such an angle that the photographer's position was fairly unambiguous, at least in terms of which landmarks would be found behind it. The final selection was based on optimising this unambiguity of the photographer's position, and also the clarity with which identifiable landmarks were visible and matchable against the map.

In piloting, subjects had little trouble correctly identifying all the photographs' locations; one photograph of Newington proved easier to identify if placed after one of the others in the order of questions, so it was rearranged accordingly. (It would have been impossible to avoid such learning effects across the four photographs, and

---

[27] The author is very grateful to Christophe Pallier, one of EXPE's authors, for allowing her access to a beta copy of version 6 and for his helpful responses to her queries about the scripting language.

so it was best to ensure that the task was as straightforward as possible to optimise all subjects' performance.)

Since the photographs were to be displayed simultaneously with the map, it was easiest to mount the photographs on card for subjects to view separately from the screen. This was not ideal, in that subjects had to move their heads to switch view from the screen to the card, but the amount of physical movement was minimised by placing all the photographs on the same piece of card. Subjects turned the card over before hitting a key to commence the post-test session, although the instructions exhorted them not to look at the photographs yet. They thus had them visible without the need for (much) further motion, although a few subjects, while struggling to identify objects in the photographs, picked up the card for a closer inspection.

To achieve this, while making the photographs large and clear enough for identification but not making the card unwieldy or too large to view without stretching, only four (5" x 7") photographs were presented. (It was deemed necessary to show more than one photograph, in case subjects had a specific problem with some scenes, and also to measure response times beyond the initial 'shock' of having to suddenly view this type of visual scene.) The card for each village (actually a page from a 'selfix' photograph album) showed the four photographs two above two, and labelled from A to D. So that subjects had no chance to view the next item between trials (though they were given only a few hundred milliseconds gap between trials anyway), the questions were pseudo-randomly mixed in terms of which photograph they asked the subject to view, so the order was not ABCD but DACB (and was the same for all participants, for the reason of optimising ease of performance by allowing visual learning effects across the four trials - these would have occurred in any case, so it was best to keep them consistent across all participant conditions). The photographs are shown in Figures 6-3 and 6-4 (as stated above, in real life they were larger, so that building names could be read easily).

The questions asked in the four photograph trials for both Saltwood and Newington are given at the end of Appendix 3. As with the main-task questions, the incorrect 'distractor' responses were chosen to be as implausible as possible, so that the viewer could eliminate them very quickly once the photograph's main landmark and approximate angle had been identified.


A

B

C

D

Figure 6-3. Newington photographs used in experiment

A

B

C

D

Figure 6-4. Saltwood photographs used in experiment

6.2.3 Individual difference variables

The study was designed to focus on the main hypothesis about the relationship between cognitive representation and task, in terms of the effect of task on performance on the photograph post-test. As such, power analysis of the experiment design (described further later in this chapter) suggested that 27-30 subjects would suffice to provide a strong test of the hypothesis. Obviously, for a reasonably powerful and definitive study of individual difference factors, many more subjects than this would be required, and this was unfeasible in the current study. However, it was felt that including age, gender, expertise and preference variables in an *exploratory* analysis might yield pointers for future research. It would give some indication of correlations or clusters among such variables, and tentative interactions with the experimental conditions. It would also allow a methodological exploration of how a future standard questionnaire on map-relevant skills could be developed.

6.2.3.1 Expertise measures

As discussed in Chapter 5, the meaning and measurement of 'expertise' relating to map-based tasks are very poorly defined. Most of the (few) studies considering expertise have started by recruiting subjects from two different populations, e.g. trained cartographers versus psychology students, without attempting to investigate the prior differences between them. Gilhooly et al (1988) tried to do a more thorough job by administering a test of contour map interpretation, a biographical questionnaire and a self-assessed measure of how much each subject used maps in their leisure activities etc. Taking the top and bottom 30% of scorers on the contour map test, they found that top-scorers were significantly better at memorising a contour map than bottom-scorers, but not at memorising a planimetric map (similar to the street maps used in the present study). In other words, their attempt at an expertise measure turned out to be too task-specific (in the sense of specific to a certain type of map).

With no more precedents than this to go on, it seems reasonable to hypothesise that expertise closely related to the task and map type at hand will be more relevant to the experimental task than other types of expertise. However, as we saw in the last chapter, we have at the very least to consider:

- expertise in map reading, particularly at street-map scales and with British OS data
- expertise in using a GIS, particularly viewing the particular type of geometric vector data display used in this experiment
- expertise in other related types of visual display, such as route navigation systems
- expertise in using computers in general, and sense of self-confidence when asked to perform any task with one
- expertise in generally inspecting visual data of any sort, especially where it has a symbolically representational purpose

Potential *priming* was also an issue: as discussed elsewhere, it was deemed desirable to avoid potentially biasing results by using language that implied an interest in spatial ability or in maps. However, the expertise questionnaire needed to be administered before the experiment, so that people who felt their experimental task performance had been poor did not then mark down their own self-rated expertise as a result. Therefore a questionnaire was designed and piloted which measured a wide range of experiences with many types of visual data and computer software. The questionnaire took only a few minutes to complete, and was designed to fit onto a single A4 sheet so as not to intimidate subjects. Subjects had to circle a number from 1 to 5 on a standard Likert-type scale. The list of items is shown in Appendix 5; the way in which analysis proceeded is described in the next chapter.

In addition to this, initial questions on the first side of the sheet asked for the subject's name, age, gender and occupation. There was then a tabular set of questions intended to identify subjects' level of education, training or self-teaching in a number of subjects (of which only geography and cartography were really of interest, although others were included both in case visual experience in general proved relevant, and as above to disguise the purpose of the experiment). Although the pilot subjects dealt well with this section of the questionnaire, a substantial minority of subjects in the experiment itself failed to complete it in a useful way (e.g. ticking 'Learned without formal qualification' for everything, or claiming not to have learned anything about any of the subjects). It was therefore not analysed further.

Finally, the questionnaire included a couple of simplistic attempts to draw a distinction between verbally- and visually- oriented subjects, by asking people to state whether they preferred to work with visual or verbal materials and whether they preferred to navigate with maps or directions. Again, although the pilot revealed no problems with these questions, and although a 'Neither' option was provided, many subjects in the experiment proper were unable to choose an option or else ticked more than one. Caution was therefore taken in using these variables, and they were subjected to only limited analyses as described in the next chapter.

6.2.3.2 'Spatial ability' measure: reference frame switching

Chapter 5 showed how the psychometrics literature has failed to produce an unambiguous factor for spatial ability which could be clearly related to the tasks in the current experiment. It also showed how, unless an aptitude test *is* theoretically relevant to a specific task, it may be worthless trying to collect ability measures. The most hopeful specific suggestion to emerge from reviews of existing tasks and tests was that relating to the way in which certain ambiguous or reversible figures force subjects to change reference frame, reinterpreting the spatial relations between different components of the displayed figure.

Although the factor examined by Carroll (1993) was concerned with rapid alternations between perceptions, a more pertinent point to the present experiment might be the ability of subjects to switch reference frame at all in the first place. If one has been viewing the duck/rabbit figure as a duck (or viewing the digital map as a geometric array of coloured lines and shapes with largely deictic spatial relations implied between them), and must then switch to seeing it as a rabbit (or as a map, with intrinsic relations between the objects), the ability to perform this switch at all and the speed with which it can occur are more relevant to the measures in the current study than the ability to then 'flick' back and forth between the two perceptions (although the latter would arguably be relevant to the 'random' condition, described later on).

The author therefore decided to include a pre-test in which subjects were asked which interpretation of some reversible figures they could see, and were then asked to respond when/if they could perceive the alternative objects. This task was designed to use the same two-choice response paradigm as the rest of the experiment, to avoid the need for practice and to avoid introducing response-specific variations into the performance measure. Five reversible figures were identified which all involved switching both a change in the functional meaning of components and switching reference frame so that spatial relations were different, as defined by Peterson et al (1992). These are shown in Figure 6-5:

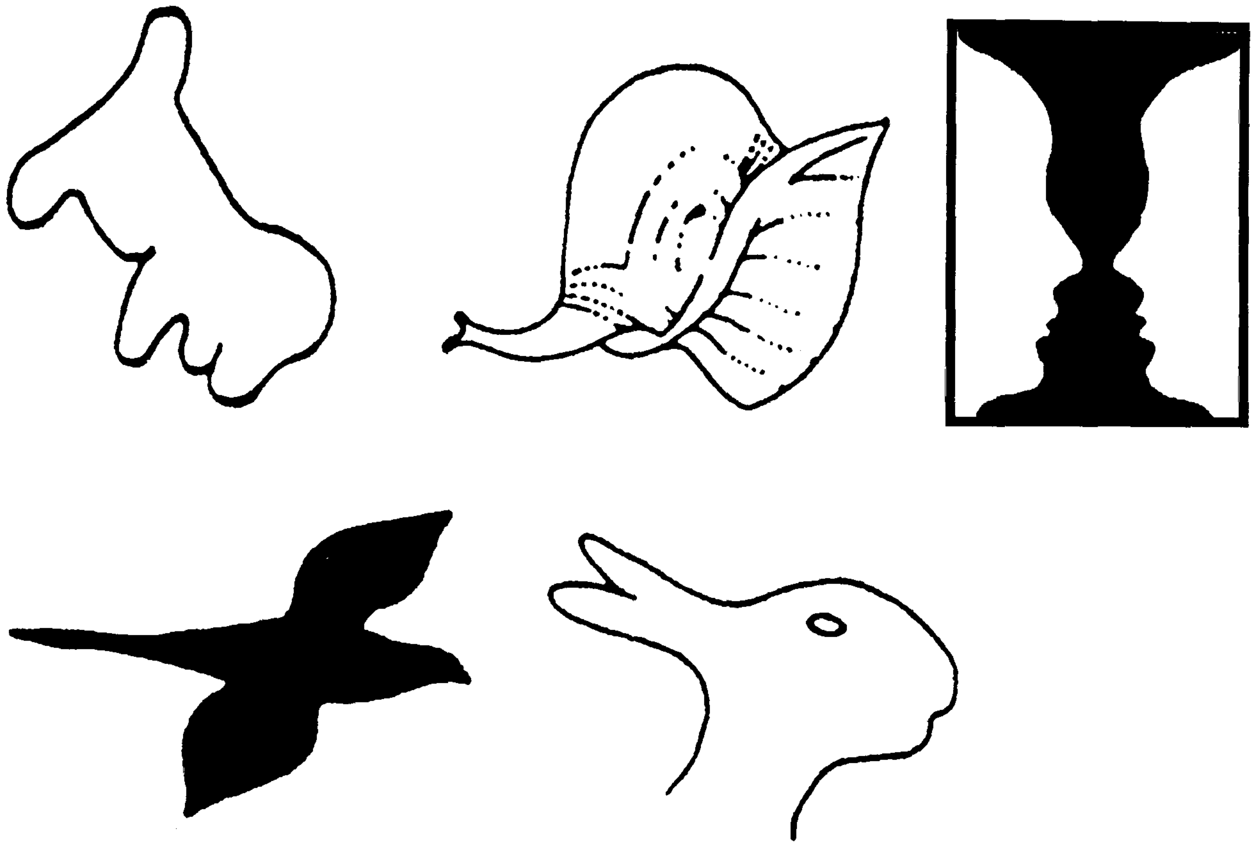


Figure 6-5. Reversible (ambiguous) figures used in the experiment: the chef/dog, snail/elephant, vase/faces, goose/hawk and duck/rabbit[28]

The chef/dog figure (Chambers and Reisberg, 1985), at top left, may be seen either as a cartoon of a chef with a tall hat and long nose, who is looking downwards and to the left, or as a cartoon puppy dog whose head is at the lower right and tail at the tip of the image. This is the only one of the figures requiring a rotation in terms of where the 'top' is. The snail/elephant figure (Fisher, 1976) may be interpreted such that the protuberance at the lower left is either the trunk of an elephant whose ear is formed from the right-hand third of the image, or the head and body of a snail emerging from

---

[28] Note: these versions of all these ambiguous figures were copied directly from Peterson et al (1992).

a shell (in which case what was the elephant's ear becomes the snail's foot, or a leaf). The well-known Rubin vase figure (Rubin, 1915), in which the black area is a vase or else the white areas are two faces, requires a figure/ground reversal as well as the realignment of components and spatial relations.

The goose/hawk figure (Tinbergen, 1948) shows either a leftward-flying goose whose head and long next stretch out towards the left, or a rightward-flying hawk with a short neck and long tail. Finally, at bottom right, the duck/rabbit figure (Jastrow, 1900) was explained in the last chapter: the two protuberances on the left are either a rightward-looking rabbit's two ears, or the upper and lower bill of a leftward-looking duck.

Subjects' responses and reaction times to both the initial and follow-up questions were recorded. Additionally, in the debriefing session at the end of the experiment, subjects were asked whether they had previously encountered any of the reversible figures, and whether they had had any problems perceiving them. The analysis of responses to the reversible figures is described in the next chapter.

6.2.4 Experiment design

One subsidiary purpose of the experiment was to examine whether subjects differed in their performance with tasks concerning the geometric visual properties of the map layout (henceforward referred to as 'geometry questions' or G), as opposed to tasks concerning the geographic representation within the map (henceforward referred to as 'map questions' or M), and whether this could be tentatively associated with prior expertise and/or preference. This implied that each subject would need to be tested on both map and geometry questions (a within-subjects variable).

This immediately raised issues of order and learning effects. Obviously subjects who had just answered 20 questions of any sort about an image would be more familiar with it than those who had not, so a second block of trials would always be easier than the first one. If the map and geometry questions were mixed, however, subjects would have to keep switching between reference frames, which one might expect to impair performance (this in fact became another hypothesis in the experiment as discussed below).

Therefore it was necessary to avoid order/familiarity effects by splitting subjects into different groups, one of which answered a block of map questions first and then a block of geometry questions, and the other reversed. Each block concerned a map of a different but essentially similar village (Saltwood and Newington, as above), so that neither question block was affected by familiarity with the visual layout. To further minimise possible order effects, half of each subject group saw the Saltwood map first, and the other the Newington map. This precaution also avoided any possible distortion if, say, the Saltwood map proved easier to interpret than the Newington one. This yielded a 2x2 design.

All subjects were then subjected to a post-test focusing on the second (most recent) map they had seen. This meant that for one group, the map in question had only been considered up until this point as a geometric display, while the other group had already begun to learn its geographic representation.

As just mentioned, it was also considered interesting to see what would happen if subjects were made to switch randomly between the two question types. Carlson-Radvansky's previous work with reference frames (Carlson-Radvansky and Irwin, 1994) had made subjects switch randomly between extrinsic and intrinsic frames throughout the experiment, apparently without considering whether this would itself affect results. If the two reference frames were held as separate cognitive representations, one would expect the effort of switching between them to worsen performance in the current experiment, compared to either all-map or all-geometry conditions. Also, by the time they reached the post-test these subjects would have seen the second (and first) map *partly* as a geographic representation, and partly as a meaningless set of coloured lines and shapes. Therefore one might expect this group's post-test performance to be intermediate between that of the 'map first' and 'geometry first' groups. The group was again split equally between subjects seeing Newington first and others seeing Saltwood first, for the same reasons as above, yielding six subject groups in all:

Location order	Question order: M = Map first; G = Geometry first; R = Randomised		
Saltwood (S) first	MS	GS	RS
Newington (N) first	MN	GN	RN

Table 6-1. Conditions in the experimental study, balancing for village and order effects

Although this yielded a balanced design which would eliminate order effects when comparing responses to different question types, further precautions were also taken. To minimise task learning effects, which could have always made the second block of questions easier, a practice session was constructed. The trials in this short session were designed to ask general, easy, questions, mixing various types and scales of simple visual image and map (but no maps of the type or scale used in the main experimental trials). Subjects had to answer ten two-choice questions in the same way as in the main trials. The experiment generator was programmed so that if subjects made more than 3 erroneous responses, the session was repeated (but in fact no subjects did so).

In this practice session there were two questions about each of five visual images. One of each pair of questions focused on its visual appearance, while the other required interpretation of the objects depicted (to give subjects some practice in having to think in more than one way about visual stimuli, as well as giving them practice in the two-choice response task). The stimuli and questions were pseudo-randomly mixed such that each question pair was separated by at least one other question, so that subjects did not know what to expect on any trial. They are listed and explained in Appendix 6.

6.2.5 Subjects

Tentative power analysis, based on piloting, established the minimum number of subjects likely to be needed to test the main hypothesis (see below). The pilots indicated that a mean difference between conditions of about 2 seconds would be reasonable to expect, and with an apparent standard deviation of about 6 seconds this led to an effect size of 0.67. To obtain a power of 0.8, therefore (i.e. have an 80% chance of avoiding a Type II error - finding no effect when one really existed), the G*POWER statistical power analysis program (Erdfelder et al, 1996) suggested a minimum total sample size of 27. Given that there were 6 experimental conditions and

ANOVA is most reliable with equal cell sizes, this implied running 30 subjects. Erring on the side of caution, however, since 0.67 is a large effect size by the standard of most behavioural studies (and since we wanted to have some leeway to test subsidiary hypotheses without compromising too much on Type I or Type II error probabilities), the sample size was increased to the maximum available within the time, which under the circumstances[29] was not much more. Therefore, 36 subjects took part in the experiment: 18 male and 18 female.

Since age and expertise were seen as relevant issues (see Chapter 5), it was not deemed appropriate to rely on undergraduate students: the age and experience range would be severely limited, and they would also be more 'experiment aware', i.e. alert to the demand characteristics of the experiment (Orne, 1962). Instead, effort was made to identify and gain co-operation from some (mostly) highly experienced digital map users (from the highways and planning departments of Surrey County Council in Kingston-upon-Thames). These 20 subjects were employees of Surrey County Council, most of whom worked with GIS on a daily basis. All but four were completely untrained in psychology or research.

The remaining 16 subjects were people known to the experimenter, who mostly had little or no experience with digital maps (with a few exceptions). Subjects' ages ranged from 20 to 58 (mean=38, s.d.=11.3). The experimenter checked beforehand (though only by verbal enquiry) that all users had adequate vision and were not acutely colour-blind. No subjects reported any problems identifying details or colours on the screen (although a few later quibbled with the name used for one pinkish-mauve colour).

Asked to rate their GIS experience on a scale from 0 to 6, the 18 regular GIS users' mean score was 4.1 (s.d.=0.76), as opposed to 1.1 (with s.d. 1.0) for the 18 non-users (Mann-Whitney U=8.00, $p<0.001$).

---

[29] The circumstances being: limited availability to real-life GIS users and to willing non-psychologist friends and family, given the time pressure induced by having to use a hired computer at £100/month!

Subjects were allocated randomly to conditions, apart from ensuring equal division of males and females in the two major conditions (map-first versus geometry-first). Later checks for age and expertise differences between conditions showed no imbalance (see Chapter 7).

6.3 Specific hypotheses

The following main hypothesis was made:

The perspective-taking task (the photograph questions) would be harder (longer RTs) where the geometric reference frame (G) had been the one most recently used rather than the map one (M). For the randomised condition where they had to keep swapping between G and M questions, it was hypothesised that the practice at swapping reference frames would enable the Ss to solve the problem more easily than the Ss who'd only applied geometry to this map, but *not* more easily than the Ss who'd become more familiar with its 'mapness'[30] (It was planned to test this hypothesis via an analysis of variance, with photo-task-RT as the dependent variable, and most-recent-condition as the independent variable.)

Additional tentative hypotheses included:

1. During the 'main' task sessions (answering the M and G questions), the geometric reference frame would be easier to apply overall, since a 'shallower' translation/ interpretation process is necessary between the basic perceptual stimuli and the descriptions given in the questions (and since the map arguably discourages 'deeper' interpretation, by not readily resembling typical paper-based maps in visual style). Having to switch between reference frames in the R condition would impair performance to below either all-M or all-G, because of the extra effort

---

[30] Alternative hypotheses were considered: (a) the need to have studied the map more closely while swapping reference frames could make the photograph task easier in the R condition than for *either* group of non-R Ss, who could have put less overall effort into familiarising themselves with the map's layout; OR (b) the Ss could have become confused about the task by being given conflicting indications about the nature of the display, and could end up performing worse on the photo question than any of the non-R Ss. However, it seemed more consistent with the main hypothesis to assume that half as much experience with a relevant reference frame was better than no such experience at all, but not as good as a whole session.

assumed to be necessary to switch between them. (Planned test: analysis of variance with main-task-RT as dependent variable and condition as independent.)

2. The relative ease of application of the geographic as opposed to geometric reference frame would be partly related to map use expertise (i.e. experience in translating between geographic entities and small-scale 2D maps). Map expertise would (tentatively) appear significant to a greater extent than other visual experience. However, this wouldn't prove that expertise necessarily *makes* people apply different reference frames more easily: the fact that they have put themselves in situations where they gained such expertise, e.g. cartography/GIS training or extensive route navigation, may indicate a predisposition towards applying them. Therefore it could be an underlying spatial ability/preference that really affects performance. (Planned tests: tentative cluster analysis to establish appropriate subject groupings across expertise measures, followed by inclusion of these groupings as an extra independent variable in the overall ANOVAs above.)

3. Age and gender effects, even if found to affect overall performance on the spatial tasks, would not interact with question type (M or G), or with condition M first, G first or R. Gender effects were not predicted to be significant, given the highly verbal aspect of interpreting the questions, and given the many null or ambiguous results of previous studies. Care was taken to avoid using the word 'map' or implying any interest in spatial ability, when recruiting and running subjects (until the post-test was reached), since previous work (Sharps et al, 1993) has suggested that this would impact on results, especially for older women; in any case, mentioning 'maps' could have biased all subjects' expectations towards the geographic reference frame. Age was expected to slow down response times overall, but not to interact with any other variables. (Planned test: inclusion as tentative extra factors in the overall ANOVAs: *however* since the null hypothesis was expected to be true, lack of statistical power would probably make any null results extremely tentative. However, in the practical constraints of the study, this was unavoidable.)

4. Performance on the reversible figures tasks would predict performance in the photograph task where it involved a switch of reference frame (i.e. interacting with the predicted M-versus-G effect), and would also predict performance in the R main-task condition (but not in the all-M or all-G condition). (Planned tests: inclusion of some measure of reversible-figures performance, e.g. mean RT with a correction for general speed of response across all tasks, as an extra factor in the overall ANOVAs.)

6.4 Procedure

Some of the instructions and timings in the computerised procedure were altered after initial piloting; e.g. the training session was shortened to 10 questions (from an original 20) because subjects found it tiring, and certain instructions were rephrased to reduce annoyance or to improve clarity. The procedure was then kept constant for all the 36 experimental subjects.

The full experimental procedure is detailed in Appendix 4. In summary, it consisted of the following for all subjects:

1. Completion of demographic/expertise questions (on paper).
2. Brief verbal instructions from experimenter.
3. Working through on-screen instructions training Ss to make responses.
4. Practice session.
5. Reversible figures session.
6. First main-task session (either M, G or R questions, and either Saltwood or Newington).
7. Pause and opportunity for short break if required.
8. Second main-task session (G if previously M and vice versa, still R if previously R; Saltwood if previously Newington and vice versa).

9. Photograph task (with reappearance of second map).

10. Debriefing and brief post-experiment interview.

The next chapter will describe the results of the experiment, including both qualitative observations and quantitative analyses.

Chapter 7: Experimental study: results and discussion

7.1 General observations

7.1.1 Qualitative observations

In general Ss went straight through the experiment procedure in around 20-30 minutes, more quickly than expected. Most commented afterwards that they learned fairly quickly to check one of the two answers, and if it seemed completely wrong to simply choose the other without wasting time evaluating both. At least one subject, however, never adopted this strategy and thus probably performed more slowly across all questions.

Certain questions caused problems for some subjects: these were especially those which asked how many of something appeared on the map (e.g. colours, white lines, yellow words, lines crossing a specified object). Subjects often differed from the experimenter in the way they counted entities - e.g. splitting lines into more segments than expected. This will be considered further below, since it occurred more for *g* questions than for *m* and hence could have affected the results.

Some subjects commented on the maps themselves, disliking colour choices or the way that 'solid' objects were displayed only as outlines so that space within or outside buildings wasn't clearly differentiated. Many subjects, when asked what they thought the point of the experiment was, suggested something to do with evaluating the suitability of these visual features, although they were actually the standard settings supplied by the Ordnance Survey.

Some noise was experienced during some subjects' sessions, but in general this seemed not to distract them significantly.

7.1.2 *Response times (time in secs from question presentation to keypress)*

Question type	Mean	Standard deviation	Minimum	Maximum	Normally distributed?
Map	10.8	2.6	6.9	16.8	Yes
Geometric	10.9	2.6	6.8	15.9	Yes
Random	10.9	2.5	6.7	16.6	Yes
Photos	19.7	7.4	7.8	43.3	No

Table 7-1. Basic response time statistics by condition and task

Table 7-1 shows basic statistics for the response time variables for the three main-task question types (whether presented first or second), and for the photograph task. It will be noted that the photograph task took significantly longer on average than the main tasks, as expected, and that its distribution was not normal due partly to very large outliers (note the maximum response time of 43 seconds). This will be considered further below.

7.1.3 *Accuracy (number of correctly-answered questions)*

Question type	Mean	Standard deviation	Minimum	Maximum	Normal?
Map (/20)	18.8	1.0	17	20	Yes
Geometric (/20)	18.7	1.0	17	20	No
Random (/20)	18.8	1.3	16	20	No
Photos (/4)	3.5	0.8	1	4	No

Table 7-2. Basic accuracy statistics by condition and task

Table 7-2 shows that most of the time, accuracy was high on the main tasks, with the lowest score being 16 correct questions out of the 20. Similarly, the mean of 3.5/4 questions correct on the photograph task demonstrates that most subjects were able to perform it correctly. However, as might be expected by such high-scoring results, the accuracy data was generally highly skewed and hence unsuitable for parametric analyses. The intention was to concentrate mainly on the response time variables anyway, and this will be the main focus below.

7.1.4 Subject characteristics: checks for intergroup differences and overall effects

Cross-tabulation checks were performed to establish whether, inadvertently, the three groups of Ss assigned randomly to the different main-task conditions (map first, geometry first, random) differed significantly on any other independent variables which could bias the results of further analyses.

Chi-squared analysis comparing the groups on gender, expertise, verbal/visual preferences and direction/map preferences showed no significant effects. A one-way ANOVA testing for age differences between the three conditions (after checking that age was normally distributed within the sample) also found no significant differences.

It was expected that age and/or gender might affect overall response times, e.g. if women were more nervous or if older people tended to have slower reactions. However, a t-test comparing male and female Ss on mean main-task RT (across all 40 trials: checks showed that this was normally distributed) showed no effect of gender; similarly, a Pearson correlation of age with response times showed no significant effect of age (Pearson r=0.16, p>0.05).

7.1.5 Outliers

All subjects were able to complete all the tasks without allowing the computer to time-out on any questions, although a few Ss claimed to have had to guess or partially guess once or twice because they could not agree with either possible answer. The main-task questions which gave problems, and hence created a few large outliers in the response time results, are listed in Table 7-3.

Map/ condition	Question	Cause of difficulty, and effect on RT
Saltwood/ geom	Which of these two words is shown? (Church; *Hotel*)	S32 couldn't find either word for >1 min.
Newington/ geom	Where is an upside-down red L shape shown? (Top right; *Middle left*)	S23 couldn't find this for >40 secs.
Newington/ geom	How many colours are there in total? (Four; *Eight*)	S23+S31 couldn't agree with either number for >29/41 secs, so 'guessed' (took highest value). (One or two other Ss also did this.)
Newington/ geom or random	How many lines cross the red object near the central blue triangle? (Ten; *Five*)	Several Ss later reported failing to agree with either number and taking the closest. S12 took >110 secs to do this.

Table 7-3. Questions causing severe problems (response time outliers)

The 5 extreme outliers mentioned in the table (out of a total of 40x36=1440 main-task response times) reduced reliability for items within each condition, and could have heavily distorted the subsequent data analyses. They were therefore replaced via mean-substitution (i.e. the mean of other subjects' responses).

For the photograph task, where response times generally varied more than in the main tasks, one subject produced one response time so extreme (more than 43 secs, on the 3rd photograph question for Newington) that it created significant problems by itself. The data in general suffered from mild skew and kurtosis, i.e. deviated slightly from a normal distribution, but this improved if the subject who produced the extreme outlier (who also caused two other lesser ones) was excluded from analysis. However, the main-hypothesis test below produced identical results in terms of where the significant differences lay between conditions, whether tested with a one-way ANOVA (excluding this subject) or with the non-normality-requiring Kruskal-Wallis ANOVA (including the subject).

7.2 Test of the main hypothesis: effect of condition on photograph task RTs

Village:	Saltwood			Newington		
2nd q type	Mean	sd	N	Mean	sd	N
Map	16.4	4.1	6	14.9	6.6	6
Random	17.7	5.6	6	19.2	4.0	6
Geometry	21.2	7.2	5[31]	25.4	5.9	6

Table 7-4. Response times in the photograph task, by condition and village

The main hypothesis, plus a check to ensure no effect of the different villages, were tested with a two-way analysis of variance: the independent variable was condition ('*m* second', '*r*', '*g* second'), and the dependent variable was mean response time across the four photograph questions. Recall that if the first set of main-task trials that Ss received was with map-related questions, the second set was with geometry-related questions and vice versa; the photograph task related to the second map, so those who had *m* first and *g* second were expected to perform significantly worse than those who received *g* first and *m* second, with the *r* condition performing intermediately.

The ANOVA supported this hypothesis: F=5.44 (with 2 d.f.), p=0.01. A Bonferroni comparison between the conditions showed that *m*-first did indeed lead to longer response times on the photograph task than *g*-first (mean difference=7.9 secs, p=0.007), but only weakly approached a significant difference from *r* (mean difference=5.1 secs, p=0.12). The mean difference between *g*-first and *r* (2.8 secs) was nowhere near significant (p=0.70). This could be taken to suggest that having answered *some* questions employing a geographic reference frame/model is enough to slightly facilitate performance in a subsequent task which requires such a model, relative to having only been permitted to consider the display's surface geometry. Figure 7-1 shows the marginal (i.e. estimated, when error variance removed) means for the three conditions. It also demonstrates that relative to the effect of condition, there was no real effect of which village was being viewed at the time (F=0.58, 1df, p=0.45), nor of any interaction between condition and village (F=0.73, 2 df, p=0.49).

---

[31] Note that this table, and the ANOVA that follows it, excludes the subject who produced a strong outlier on the photograph task, as discussed above.

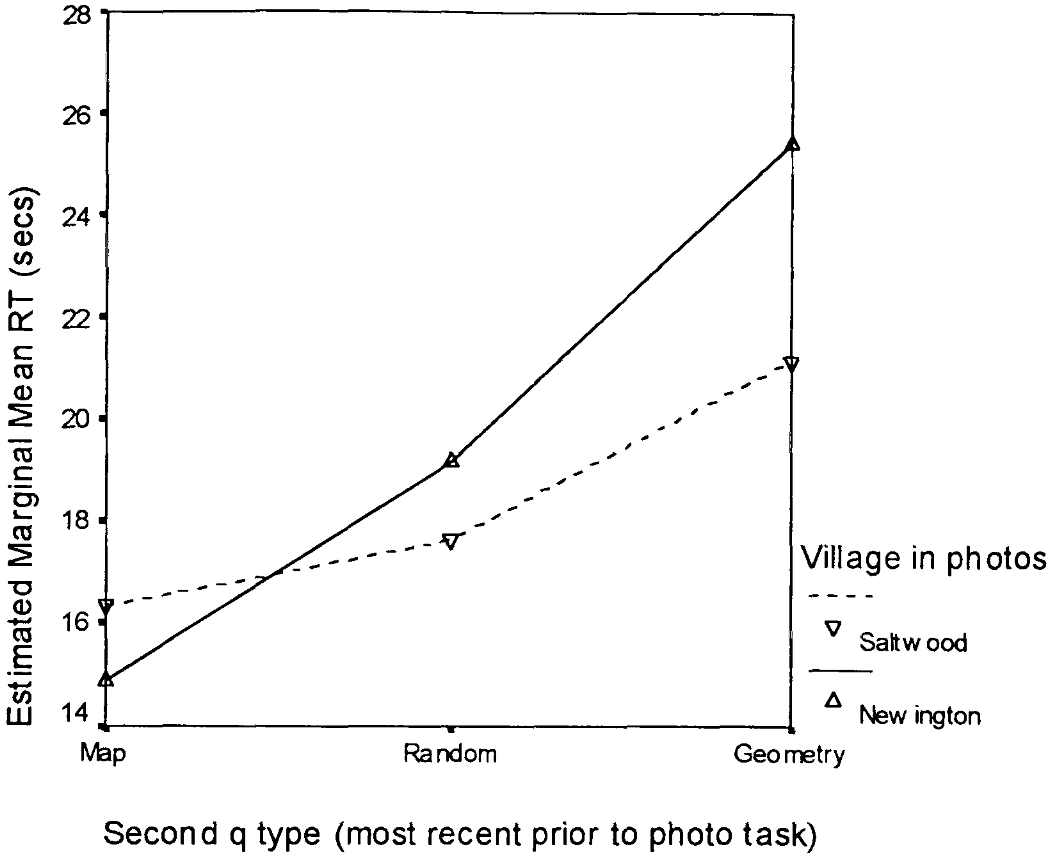


Figure 7-1. Response times in post-test photograph task, by main-task condition

The strong difference in mean photo task RT between the *m*-first and *g*-first condition suggests that the subjects had undertaken some level of geographic interpretation in order to answer the *m* questions, but had not had the chance to do so with the *g* questions. Ss in the latter situation still managed to perform each photograph task in an average of 23.5 secs, so performance was not entirely impaired: this would probably not have been the case had the map been removed so that the photograph task was performed from memory, but Chapter 6 explained that this option was rejected since it would have constituted an unrealistic (and possibly unperformable) task paradigm.

7.3 Testing subsidiary hypotheses

7.3.1 Effect of condition on main-task RTs

A subsidiary hypothesis raised in Chapter 6 was whether response times in the main task sessions would vary between question types. It was suggested that geometry-related questions, all other things such as comprehension being equal, could be responded to more quickly owing to the requirement of a 'shallower' interpretation (not having to translate the visible map objects into geographic entities), and that in the *r* condition where questions were mixed, and where Ss had to keep switching between the two reference frames/models, response times would for that reason be slowest of all.

However, a one-way ANOVA with main-session response times (including both first- and second-block trials) as the dependent variable, and with question type as the independent variable, produced no significant results (F=0.006; p=0.994). The mean response times to the *m*, *g* and *r* questions were 10.8, 10.9 and 10.9 seconds respectively: effectively identical. While this obliges us to accept the null hypothesis regarding relative difficulty of question types and of continuous reference-frame switching, it arguably makes the finding of strong effects on the subsequent photograph task even more impressive: any artefactual explanations of the main finding in terms of extra fatigue or confusion (due to having just performed a tougher task) can probably be rejected.

7.3.2 Expertise: defining groups and testing effects/interactions

The crudest measure of relevant expertise available was to split subjects into those who were regular GIS users (all but 2 of the 20 Surrey County Council employees) and those who were considerably less experienced with GIS (the rest). The classification was based less on the expertise questionnaire than on their own informal descriptions of their knowledge and experience, and the written descriptions they gave of their jobs. These fell neatly (and coincidentally) into equal groups of 18. When the GIS user/non-user distinction was added to the main-effects ANOVA (i.e. 2x3, GIS use x condition, with photo RT as the dependent variable and omitting the one large outlier as explained above), no significant effect or interaction was obtained.

As explained in Chapter 6, the purpose of including many potentially relevant types of visual expertise in the pre-experiment questionnaire was to see if expertise fell into identifiable 'types' which could be related in any way to subjects' performance. Accordingly, a cluster analysis was planned on the expertise data, the results of which could form a new categorical variable for tentative statistical analysis. Obviously the point of this exercise was largely exploratory and could not be treated as conclusive or useful without further replication, not least because of the small sample size (although cluster analysis is not subject to the same stringent sample-size requirements as more parametrically-based methods such as factor analysis). Since expertise was not the major focus of the study, such replication was not intended to occur within the present doctoral research.

Cluster analysis methods were carefully investigated to choose an appropriate one for the data: the Pearson distance measure was chosen because it tends to group people according to the overall profile of their responses rather than their absolute scores (Aldenderfer and Blashfield, 1984): this was deemed important since inspection of the data showed that some Ss were far more reluctant than others to circle '5' despite having had extensive exposure to the same types of display/information. Average linkage within groups was chosen as the clustering method, which tries to make the average distance between cases within a group as small as possible (in other words, each group is as homogeneous as possible within itself). Examining the range of solutions, one with five groups appeared to be the easiest to interpret:

1. 16 subjects. Largest group (an artefact of most clustering methods), consisting of subjects whose strongest collective feature was fairly extensive general experience of many types of printed map (streetmaps, railway network maps, Ordnance Survey/road atlas maps, plus travelling (hiking and visiting unfamiliar towns/countryside). Most had little or no technology or GIS expertise.

2. 3 subjects. Small, technology-literate group, highly literate in virtual reality and/or computer graphics, but not in GIS or maps. All three were highly advanced software engineers (working in different organisations).

3. 9 subjects. Generally technology-literate group, with specific expertise in various maps and in GIS.

4. 7 subjects. Experience of various maps and GIS and also for some route planning/ navigation systems, although not very strongly technology-literate.

5. 1 subject. Little expertise in most map types or in GIS, but strongly artistic with experience of scenic photographs and of creating imagery.

Unsurprisingly, the earlier GIS user/not distinction was found to be significantly related to this 5-way grouping, in that significantly more regular GIS users were found in groups 3 and 4; however the cell sizes were too small to calculate a reliable chi-squared measure for the cross-tabulation.

It was difficult to know how to use this 5-way grouping in an analysis, owing to the different group sizes. However, it was considered interesting to compare group 1, with much map and travel experience but little GIS expertise, with a combination of groups 3 and 4 (GIS users but not advanced technologists). This gave a two-way comparison with coincidentally equal groups of 16 (although the latter group dropped to 15 after the outlier on the photograph task was omitted). It was thus similar to the earlier analysis crudely comparing professed GIS users with non-users, but eliminating those few subjects' data that appeared to indicate a less typical pattern, and which could thus have previously confounded the analysis.

The two-way grouping was run in a two-way ANOVA with the main condition effect as the other independent variable and mean photograph task RT as the dependent. However, even this analysis showed no significant effect of expertise type (F=0.36, 1 df, p=0.55), and no interaction with the main effect of condition (F=0.14, 2 df, p=0.87), as shown in the graph:

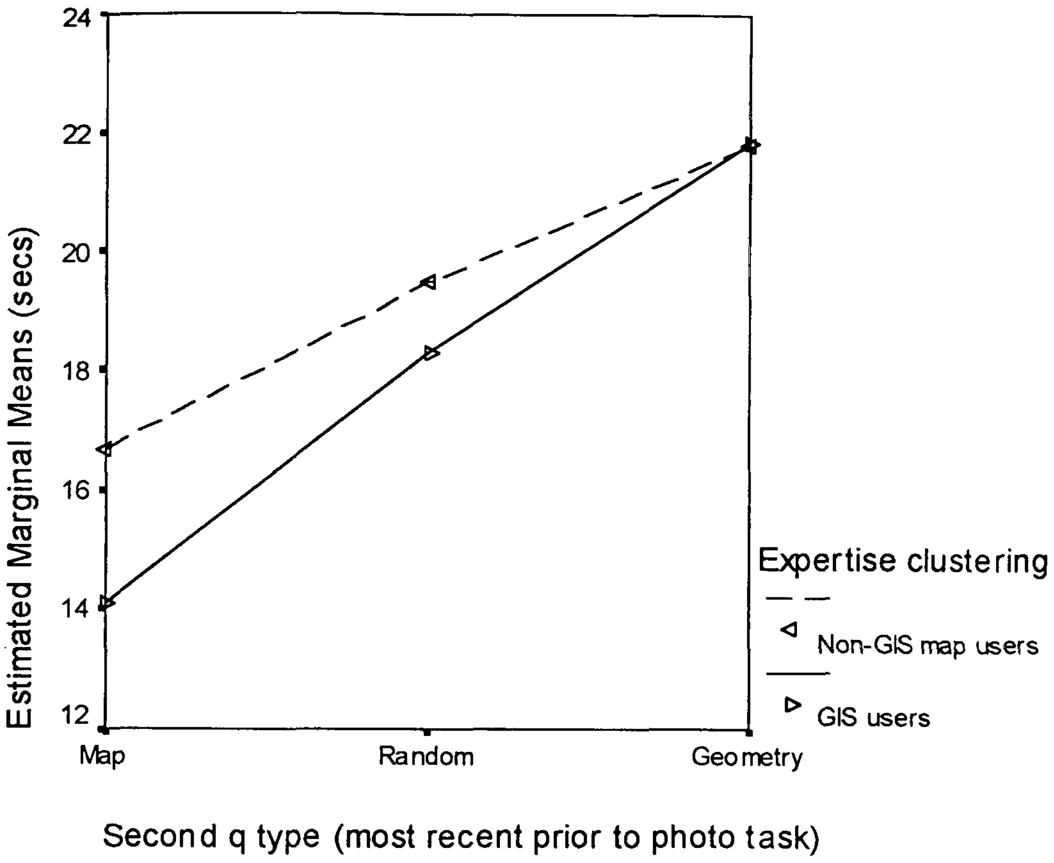


Figure 7-2. Post-clustering two-way expertise grouping (non-significant effect and interaction) shown with the main condition effect (significant as reported earlier)

7.3.3 Reversible figures performance and effects/interactions

Most subjects had no problems responding to the reversible figures, although two claimed to have a problem deciding which object they had spotted first, saying they had noticed both simultaneously. The objects first spotted in each figure, and the number of Ss responding 'No' when asked if they could see the alternative object (i.e. failing to reverse the figure successfully), is shown in Table 7-5, along with the mean response times. Altogether, 12 subjects managed to spot both interpretations of all 5 figures; 23 subjects failed to reverse just one figure; only 1 subject failed to reverse two.

	duck/rabbit	chef/dog	vase/faces	goose/hawk	elephant/snail
initial perception	duck=15 rabbit=21	chef=19 dog=17	vase=31 faces=5	goose=19 hawk=17	elephant=19 snail=17
n of failures	2	12	4	1	6
Overall mean reversal RT	2.7	3.5	2.8	2.9	3.9
Mean reversal RT (failure)	2.1	3.6	6.7	6.9	4.2
Mean reversal RT (success)	2.7	3.4	2.4	2.8	3.9

Table 7-5. Successful reversal and response time data for the five ambiguous figures

The table shows that the chef/dog figure, which as discussed in Chapter 6 involves a rotation of reference frame (and two fairly unrealistic cartoon figures), caused subjects the most trouble in terms of failures, although the likelihood of initially spotting one or the other object first was approximately equal. The only figure for which this was not the case was the Rubin vase/faces figure. The figure in question is often referred to as 'the Rubin vase'; since in the debriefing 21 of the 36 subjects explicitly mentioned that they'd seen it before[32]; it is reasonable to surmise that many would have heard it referred to as such and this would have biased their initial perception of it. The fact that the number of reversal failures was still by no means lower than the others, and that the mean response time to it was also not faster than for any figure, suggests that familiarity with a figure did not make a difference to the difficulty of the task or its validity as a measure of reference-frame reversal.

The response time data shows that the slowest reversals were for the elephant/snail figure. Note that the RTs were greater for reversal success than for failure, although not always enormously so. The overall mean RT for reversal success (answering 'Yes' to the second question) was 3.0 secs, while for failures (answering 'No') it was 4.3 secs.

In the previous chapter it was explained that the issue of switching reference frames might be expected to be relevant in two ways in this experiment:

---

[32] Not all Ss were explicitly asked this, as it was only added to the debriefing after the first few Ss. 5 had seen the duck/rabbit, 2 the goose/hawk, 1 the chef/dog, but 0 the elephant/snail. Two other Ss had not seen any, but had seen the famous "wife/mother-in-law". Certainly, the reversible figures idea was familiar to almost all Ss.

1. in the expected performance penalty for the *r* condition where subjects had to keep alternating between *m* and *g* questions
2. in the effect of condition (*m*-first or *g*-first) on photograph task performance

Since the first of these did not prove to be a significant effect anyway, as shown above, there was no point in testing to see if reversible figures performance (suggesting ability/speed in reference frame switching) made any difference to it. Regarding the photograph task performance, however, a Spearman's rank correlation showed a significant relationship between RT for (successful) reversible figure responses, and mean photograph task RT (omitting the photo RT outlier mentioned elsewhere, Spearman's r=0.395, n=35, p=0.009). However, this doesn't necessarily imply that reversible figures ability was relevant to the photograph task: both RT measures could just reflect a general 'speed of response' factor that varied between individuals across tasks. One way to test this would be to perform a partial correlation, controlling for general response speed by including a non-reversing response time measure (such as the mean response time in the main task sessions, while Ss were responding to *m* or *g* questions), but since neither photo RT nor reversible figures RT were normally distributed, partial correlation analyses could not be performed.

However, reference frame reversal ability, if measured by performance with the reversible figures, would probably be reflected both in the RT of successful reversals and in the number of reversals which Ss managed to spot. An attempt was made to combine these into a 'score' by dividing the mean RT for successful reversals by the number of these reversals (so that the score was smaller for faster RTs and/or for greater success, and larger for slower RTs and fewer successful reversals). This also correlated very highly with photo task RT (Spearman's r=0.440, n=35, p=0.004), but again was not normally distributed (suffering significantly both from skew and kurtosis), so further analyses were not possible and the effect could still be explained away in terms of general speed.

Attempts to rerun the main-hypothesis ANOVA described earlier (effect of condition on main photo task RT), using various second factors based on reversible figures performance (e.g. top and bottom 25%, total success versus at least one reversal failure, etc.), all led to insignificant results and no interactions with the main factor of experimental condition. Further work is obviously required if the reversible figures task is to eventually provide an predictive psychometric measure of reference frame switching, and even then it may not be sufficient to predict the current experimental results (because switching between geometric- and map-oriented views of a display may involve more than simple reference frame switching: see below for further discussion).

7.3.4 Other individual difference variables: age, gender, visual/verbal preferences

To avoid generating a mass of data on interactions which were of no theoretical interest, the analysis of these variables was split into a set of successive ANOVAs rather than a single multiway number-crunch. The separate effects of age and gender on response times in the photograph task were of interest (for the same reason stated earlier, i.e. the general assumption of older women especially having problems with spatial tasks and maps), as were their interaction with the impact of the main task condition (because, for instance, if it is the language of maps which affects older women's confidence more than other ages and males, then having dealt only with the surface geometry of the map could be less of a detriment to their performance than for other Ss).

A 3-way (4x2x3) ANOVA was thus performed using age group (split by decade, i.e. 20-29, 30-39, 40-49 and 50-59), gender and main-task-condition, with mean photograph task RT as the dependent variable (still omitting the extreme outlier case as above). This produced no significant effects or interactions (apart from the previously observed effect of main-task-condition); however, the small sample size and apparent effect size meant that the observed power values were extremely low, varying between 0.085 and 0.564; clearly not enough to reject the likelihood of Type II errors. We can only say therefore that the data collected was insufficient for us to identify any effects of gender or age on the photograph task or on the impact of induced reference frames.

Separate 3x2 ANOVAs including either verbal/visual preference, or map/direction preference, alongside main-task-condition, still with mean photograph task RT as the dependent variable, also produced no significant results, and again observed power turned out to be too low to conclude much from these. Part of the reason for this observed power - very much a post-hoc calculation - was that the size of mean differences between the various individual difference groups were minuscule - at the level of tenths or hundredths of a second when the overall means were around 14 or more seconds - and were not always even in a meaningful direction. So to say that the analyses were too weak is not to imply that they were 'nearly' reaching significance: there was no evidence that they were there at all. But much larger samples would be needed to comfortably accept the null hypotheses that expertise, age, gender and 'verbal-or-visual' distinctions made no difference to the tasks under investigation.

However, one more analysis, looking at whether gender and/or age interacted at all with verbal/visual or map/direction preferences in predicting photograph task RTs, did yield one surprising significant F value: the interaction between gender and map/direction preference gave a much higher F than any others in these analyses (F=7.75, p=0.015). Given the number of analyses being performed on the same data, significance at the 5% level is hardly conclusive (due to the increased probability of happening upon a random effect if enough numbers are crunched - akin to the possibility of obtaining tails 3 times in succession if you toss the coin many more times). We can tentatively suggest, however, that this is mild evidence that for male Ss, preferring directions to maps when wayfinding indicated something which caused longer RTs in the photograph task, whereas for female Ss this direction preference was linked to *shorter* RTs. Having no preference (or preferring to have both available), however, was linked to *longer* RTs for females but much *shorter* RTs for males. For both males and females, the RTs of Ss who preferred maps were intermediate.

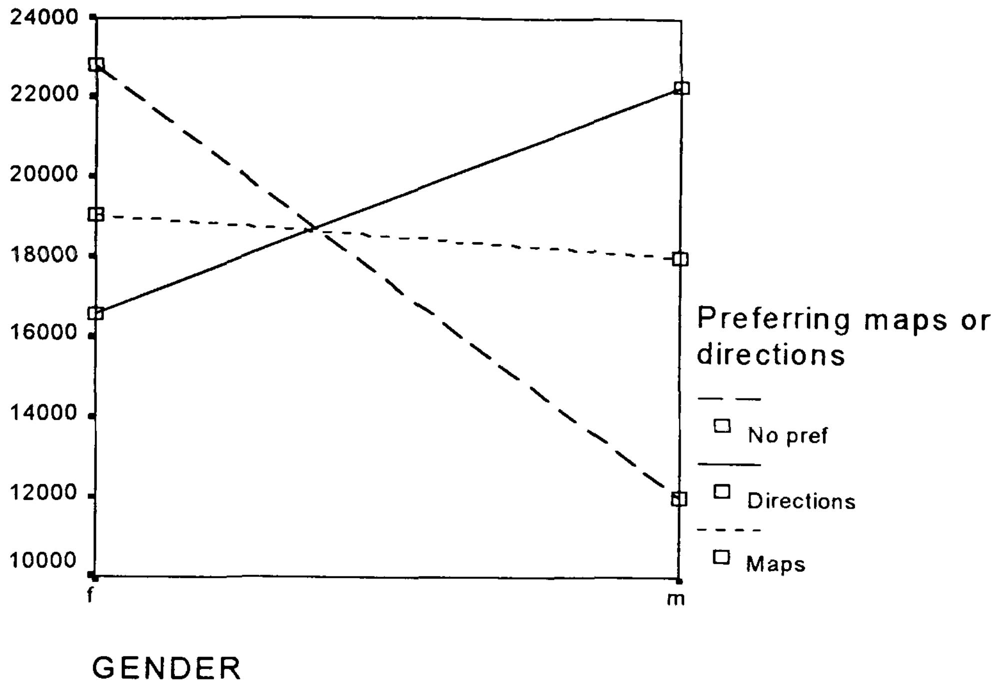


Figure 7-2. Graph showing relationship between mean photograph-task RT (y axis), gender (on x axis), and navigational preference (separate lines)

This finding, if it indicates a real effect and not a Type I error, is difficult to explain. The directions/maps question was intended as a crude measure of verbal/visual preferences in a relevant applied scenario. Here it may be suggesting that for male Ss, having a 'verbal' preference inhibited the 'visual' photograph task, as might be expected, but Ss who had no preference or preferred to work with both forms of information performed best; this makes sense as it could imply that Ss who could cope with the linguistic requirements of the tasks and could interpret such information as well as the spatial cues would be at an advantage.

For women, however, for whom the effect of the map/directions distinction on RT was visibly less strong, we could suggest that the preference conditions did not really differ at all and the graph shows random 'error' variance only. Or we could suggest that perhaps those women for whom wayfinding was more dependent on interpreting linguistic cues may have had such highly-developed language skills that they were better able to pick up knowledge about the map from the earlier tasks than the men; those who had no preference or preferred both information sources could have been expressing a lack of confidence in their ability to perform any spatial tasks (as previous studies suggest, and some anecdotal evidence in this one such as "ooh I'm no good with maps" etc.).

However, all of this is clutching somewhat at straws: the result requires further replication.

7.4 General discussion

The main hypothesis of the experiment, that the linguistic/analytical demands of the task affects the nature of the mental model built up of a digital map, and hence also the spatial reference frame used to understand relations between the depicted objects, appears to have been supported. Note however that the methodology and tasks used, in maintaining some contact with ecological validity, were by no means 'pure' tests of reference frames and of the types of spatial relation encoded; subjects will have encoded not only different forms of the relations between objects, but also different information about their actual nature. Subjects in the *m*-first, *g*-second condition will probably have developed less awareness about the existence of the various portrayed features of the village they saw second, besides being less aware of their relative locations in 'real' space.

The presence of the Saltwood and Newington war memorials, for instance, will have been irrelevant to the *g* questions but assumed massive importance when matching the photographs to the maps in the final task. There was a strong focus on colour in some of the *g* questions, although the colours in the map were almost all arbitrarily applied by the Ordnance Survey (at least, arbitrary as far as can be deduced: the OS themselves may have applied some logic to it), other than the Newington churchyard yew tree being shown in green. Therefore the changes in colour of all the objects when viewing the photographs could have had greater impact for the *m*-first, *g*-second subjects, potentially increasing their hesitation. The lesson to be drawn is that we have demonstrated an effect of task on the mental model which is developed by the user, but the effects of task went beyond the issue of the reference frame used to encode spatial relations (although this can be assumed to be a key aspect, the clearest one linguistically distinguishing the two main-task question types, and the most relevant one to the final photograph task).

Nevertheless, a key point to bear in mind is the finding that forcing subjects to focus on the surface geometry of the map did impair their performance when asked to relate the map to the 'real' space it represented. One could argue that many GIS currently cause users to spend too much time dealing with the visual geometry of the map, e.g. by having to undergo lengthy processes to digitise, edit and polygonise it. These results suggest that this may inhibit their geographical knowledge of the space which is represented, and hence decrease their effectiveness in tasks requiring such understanding. It thus lends support to the efforts of the NCGIA and Varenius initiatives, and other researchers elsewhere, to develop better and more geographically-intuitive ways of interacting with GIS (e.g. by not having to specify things in terms of 'polygons' or 'line segments').

The fact that the main-task questions themselves were not found to be significantly harder to respond to in the *g* or *r* question sets than in the *m* sets suggests that subjects had no more difficulty responding to questions demanding a geographical interpretation than to those asking about simple visual characteristics that required them to ignore that interpretation. Nor did the *r* condition subjects have any apparent trouble switching rapidly between the two question types; the mental model they built up of the map could clearly incorporate both aspects without any kind of contradiction or inhibition. This was shown by their performance in the photograph task being indistinguishable from the subjects who had only had to consider the second map's geographic representation. It has previously been assumed (Laurini, 1995) that users have difficulty dealing with aspects of a digital map which have no reality in the 'real' space it represents, such as error/projection data, status and currency, etc. However, the variables Laurini considered were relatively abstract properties of the map, not its basic visual geometry, so the current findings do not necessarily contradict this perceived problem (after all, the USIS re-analyses suggested that the equally abstract and non-geographically-relevant computer commands and syntax appeared to create the most common types of error for even highly experienced users).

The failure to find any really significant individual difference effects was predictable given the restrictions on sample size in this study, and the highly exploratory nature of some of the measures used. A much larger study would enable not only a better test of the relevance and reliability of these measures, but also an opportunity to build a predictive general linear model (using multiple regression/ANOVA) to see how much the different variables affected performance.

Chapter 8: Conclusions and future directions

8.1 Development of ideas during and since this doctoral work

Since the author registered as a Ph.D. student in October 1993, several external publications and events have added weight to the appropriateness of the work and have strongly influenced its development. It is probably useful to summarise these, and also more recent external work which points towards new directions for further research.

Chief among the post-1993 developments were MacEachren's 1995 book *How Maps Work*, which is such a thorough and probing look at the cognitive and semiotic issues of map design that it is inevitably quoted throughout this thesis. Among the ways in which MacEachren's work confirmed the author's understanding of the field were:

- the importance of incorporating up-to-date and thorough psychological knowledge and theory into our understanding of cartographic design and use, rather than paying scant reference to it as many authors have done;

- the relative dearth of experimental work by psychologists (as opposed to cartographers) on map interpretation, despite its potential as an interesting area of applied cognition (a point also made by Blades (1997), and reinforced at the 1999 Varenius workshop described below);

- the severe lack of progress in defining or understanding 'expertise' or 'ability' in this context;

- the increasing relevance of work in this area at a time when public access to the creation, editing and viewing of map-based information is massively increasing. MacEachren ended his book with the view that this increased access means that "understanding how maps work will become an increasingly important, while increasingly interdisciplinary, endeavour" [p. 461].

This idea of trying to apply spatial cognitive concepts in the context of digital maps, considering such issues alongside the usual human-computer interaction issues, was also partly inspired by a project at Loughborough looking at pictorial databases (Lansdale et al, 1996). Although the focus with that project was on spatial *memory* (since it concerned database retrieval, and pictorial recognition by experienced users), it led to some highly specific experiments on object location which demonstrated the potential link between the theoretical and the applied in such a context (Lansdale, 1998).

On the more theoretical side, de Vega et al's 1996 book *Models of visuospatial cognition* (de Vega et al, 1996) was a useful update on the attempt to bring together different viewpoints about spatial cognition, although focusing heavily on the mental models concept to the exclusion of issues such as reference frames. It also introduced this author to Michel Denis's group's work in France on verbal descriptions of maps and routes. However, work elsewhere on reference frames which also emerged during the course of the research strongly influenced the direction of the final experimental study, particularly papers by Logan (1995) and by Carlson-Radvansky (Carlson-Radvansky and Irwin, 1994; Carlson-Radvansky and Radvansky, 1996).

More recently still, and unknown to the author until she was writing up this thesis, work on the concept of 'implicit learning' has relevance to the tasks performed in the final experimental study. Wright and Whittlesea (1997; Whittlesea and Wright, 1998) performed experiments to examine the assumption that such learning is unselective and uncontrolled by the subject. This would imply that in the study described in Chapters 6 and 7, the subjects answering geometrically-oriented questions about the map would still be somehow 'absorbing' the geographically-determined structure of it. If this was the case, and if the implicit learning issue had been known to the author, it could have been tested to some extent by comparing the 'm first, g second' subjects' photograph task performance with the performance of an extra group of subjects who hadn't seen the map at all prior to performing the photograph task. However, since Wright and Whittlesea's results, obtained in far more tightly-controlled conditions, suggested that in fact the task subjects have to do, and the cognitive strategies they choose to use to perform it, *does* influence what they learn about underlying structure, so we have no reason to assume (although the possibility is interesting) that subjects in the 'g' condition still managed to learn the geographical content of the map, with no apparent intention or control.

Indeed, the results obtained in Chapter 7 lend some partial support to Wright and Whittlesea's conclusions, since the learning subjects did was *all* to some extent implicit (they weren't expected to memorise the map at all, and had no idea of the experiment's purpose, and most were unaware even that two different types of question were being asked let alone that they were thereby learning two different types of information). Despite the lack of explicitness or awareness of the two main experimental conditions, there was a significant effect on the photograph task performance, suggesting that what was learned was influenced by appropriateness to the task in the same way as in Whittlesea and Wright's work.

Meanwhile, in February 1999, the present author was invited to attend a research workshop run by the US National Science Foundation-funded Varenius initiative (the successor to the NCGIA), on the topic of 'Multiple Modalities and Multiple Frames of Reference for Spatial Knowledge'. The attendees were largely a mixture of psychologists and geographers, since the continuing aim of Varenius is to further 'geographic information science' (as discussed in Chapter 1). However, the contributions of many of the psychologists present did not concern geographic space or maps of it, but experiments on table-top or room-sized space, whose results are difficult to generalise. Nevertheless, although at the time of writing this the official report on the meeting is not yet available, a few interesting pointers did emerge which connect with the topics explored in this thesis:

- Various presenters commented that the nature of the task people are asked to do in experiments, mainly on wayfinding or orientation, determines the reference frame they use to encode spatial information, just as it did in the present experimental study.

- The initial perspective or orientation from which a spatial array or scene is viewed was assumed to be a key variable (see also below), although few studies have looked at direct 'overhead' views, let alone at overhead symbolic views like maps.

- The concept of reference frames was discussed, with such questions being posed (and unanswered) as: Do frames have geometry, in terms of having axes, a heading, etc.? Are different frames engaged for different cognitive processes? Can multiple reference frames be used simultaneously? Is this concept a useful one to discuss, like grammar in language, while perhaps not being a psychological reality? (Most attendees felt that the psychological reality of reference frames was already proven from recent neuropsychological evidence, and from our coping with phenomena such as relative movement.)

- In attempting to define when multiple reference frames could be assumed, one discussion group proposed that two reference frames are distinct if, and only if, they hold different sets of minimally sufficient information for describing spatial relations. This means that the geographic/geometric distinction drawn in the present experimental study did indeed reflect at least two separate reference frames, since the relative positions of all the objects in the map display could be described in two different ways (intrinsic, assuming real-world properties and relations, or deictic in terms of position on the screen).

- Geographers present noted that while most psychology experiments on reference frames have considered contrasts such as egocentric versus allocentric, and focus on relative location (objects' positions either relative to each other or relative to the person), geographers tend to try to define absolute frames of reference such as co-ordinate systems and grids. Thus to them, a 'geographic' reference frame would imply these attempts at absolute referencing (rather than the relations between objects in geographic-scale space, implied by this author's use of the word 'geographic'). However, it was pointed out that even then there is at least one reference object to which all positions are relative, i.e. the earth itself.

8.2 Limitations of the present work

8.2.1 Timeliness

Much of the work in Chapters 2 and 3 of this thesis was performed on systems which will by now have been replaced or extensively upgraded - that being the nature of modern technologies. Nevertheless, the lessons learned from the reanalyses of the USIS work, concerning the complexities of users' tasks and the difficulty of making prescriptive HCI judgments for such sophisticated technologies, remain current. The literature review in chapter 4 integrates ideas from various disciplines in a way not previously attempted, and the review in chapter 5 attempts to move us on a little further towards more practically useful definitions of spatial 'ability' and 'expertise'. This latter review is very timely, since the recognition that spatial ability in particular has been an oversimplified concept in the past is frequently occurring now in the literature (e.g. Lohman et al, 1987; Carroll, 1993; Stumpf and Eliot, 1995; Vecchi and Girelli, 1998). The final study, while moving more towards cognitive experimentation and away from pragmatic HCI (compared with the earlier analysis chapters), nevertheless serves as a reminder that task, design and cognitive representation are probably interdependent, important factors in work performance with spatial data interfaces such as GIS.

Since this thesis contains various observations and assumptions about GIS user interfaces, i.e. the means by which users interact with digital maps, and since many of those were based on early 1990s technology, the author visited a commercial GIS exhibition in the autumn of 1998 to observe recent advances in the technology and its usability, and hence to assess how up-to-date these assumptions were, and whether any of the USIS-derived recommendations would still be timely.

The most obvious feature of the current GIS products being exhibited was the increased uniformity of their user interfaces. Thanks to the increased speed and market dominance of the Wintel PC, and the subsequent standardisation enforced by Windows 95, most GIS now have similar menus, toolbars of buttons, dialog boxes and hierarchical online help files. (NB the USIS users of systems like Intergraph™ and Smallworld™, which pioneered this interface style in the early 1990s, often did not know what half the buttons were - hopefully that is not true of the newer ones). Import/export seems to have been addressed better with new standards and software modules, as well as the ability for some GISs to directly read data from ordinary databases and add 'spatial' fields to them so they can be linked directly, without conversion, to locations on a digital map (Oracle®, 1997). Thus geographic data becomes just one aspect of an overall information system, a trend recognised as inevitable by some in the industry (Spooner, 1999). Mobility of geographic data is now a major interest, with most vendors now selling add-on tools to enable WWW display of GIS data, and the production of ever-smaller portable computers and GPS terminals.

Undoubtedly, for all its faults, the Win95 user interface has made using most GIS much more predictable and logical than formerly. Help files, using the Microsoft Help application and modelled on its hierarchies of topics, are much more comprehensive and accessible than in 1993, when some users had to type words like 'apropos' or trawl the menus to find a 'help' command. (However, I saw little evidence of using examples to illustrate online help topics, as we recommended in USIS.)

A recent article in *Mapping Awareness* (Toon, 1998) exhorted the GIS industry to develop 'intelligent agents' that would pick up on users' tasks and learn to automate repetitive procedures without the need to program macro scripts. Where possible, however, GIS vendors already seem to be offering selections of commonly used scripts for users to run or adapt. Given their evident interest in following Microsoft's lead in other user interface components, and given Microsoft's latest incorporation of at least semi-'intelligent' features in products like the 'Office' suite (albeit still not learning agents to the extent suggested in the article), GIS are bound to incorporate some aspects of this concept in the near future.

What may, ironically, slow this development down is the GIS vendors' keen awareness of the variations in their users' tasks and requirements. Despite the apparent uniformity of the user interface and functionality of most current GIS, the companies still pointedly stick to favoured market sectors and promote themselves to those sectors as the one who 'understands' them. While this helps to ensure survival in a crowded marketplace, it also demonstrates the real differences in usage between, say, an environmental agency and a telephone company, as shown by the USIS results. GIS customers, individually and through their user groups, have always been forthright in suggesting problems with and improvements to their expensive investment, and this is one area of the software industry where the developers seem to genuinely pay attention to some extent.

Meanwhile, the continued improvements in display screen technology, along with more intuitive tools for altering visual map features, have similarly transformed the maps towards more recognisable and 'realistic' (i.e. paper-like or aerial-photo-like) layouts. Displaying aerial photographs or colour relief maps as raster backgrounds to vector data is now considerably less costly (in terms of memory, speed, visual clarity and monitor price) than in older systems.

Nevertheless it should always be remembered that as computing technology progresses, those users who have access to the latest technology frequently suffer problems with 'bugs' and with unnecessary functionality burdens; at the same time, there is always a large body of users (e.g. in small businesses and the public sector) who continue for years to use the previous generation of hardware and software with all its attendant problems. Thus both the usability and map interpretability problems discussed in this thesis, based on GIS as they appeared in the early and mid 1990s, continue to be relevant to many users (and anyway may not all be solved in the later systems: e.g. a GIS vendor at the show admitted to this author that the majority of changes to their systems between versions had served to further increase the complexity of functionality, rather than to simplify use).

8.2.2 Theoretical gaps

The theoretical gaps in the earlier chapters of the thesis have already been spelled out. We do not have a consistent and comprehensive way to define just what a GIS task is, in terms of the varied knowledge and cognitive models one might apply to it as an analyst, especially given the HCI community's past tendencies to focus either on simplistic technologies or on dynamic process control. In a situation where most of the literature on GIS has concerned conjecture more than empirical study, it is hard to progress beyond a certain stage of thinking. Having shown that spatial problems are apparently *not* the issue, but a dearth of suitable task representations and usability guidelines are, the author felt there was no further for her to go along the HCI path at the end of Chapter 3. However, more recent ideas about task-related representations (Barnard and May, 1999) suggest ways in which we can integrate consideration of users' cognitive processes with HCI analysis for complex systems, so perhaps that theoretical gulf may be bridgeable.

Chapters 4 and 5 of this thesis have been largely concerned with bridging other theoretical gaps, and with making hypotheses about people's behaviour with digital maps (and probably with other kinds of map).

As mentioned earlier, much of the work in psychology concerning reference frames has treated orientation (or viewing perspective) as a major variable. The present experimental study, in Chapters 6 and 7, ignored it. Clearly the overhead, north-up perspective of the map display was different from the various immersed perspectives of the photographs in the post-test - but the key problem is matching the visible objects against features on the map, so as to identify the approximate position of the photographer (unlike many psychology experiments, where the subject is asked to *imagine* a different perspective and must then build up a *mental* image of the changed relative positions of objects). Similarly, main-task questions such as "Is the War Memorial visible when standing outside the Castle Hotel?" *could* be solved by imagining oneself standing on Saltwood village green and looking in the right direction - but they don't *have* to be solved this way (checking the map to see if anything stands between the two will suffice). Thus perspective-taking (i.e. rotating one's mental orientation) was not an absolute requirement at any point in the present experiment, although it may have played a part in it. If it was relevant, then obviously this could be used to help explain the longer response times for the photograph task than for the main task.

Another issue relevant to the present work is linguistic priming: i.e. the wording of the 'm' questions effectively primed Ss on the names and natures of the real-world objects (i.e. made them more likely to respond quickly to them than to other words), as well as on the spatial relations between them. However, as discussed earlier, it is more sensible to assume that reference frames include informational content about 'what' things are, than to imagine them as referring solely to spatial relations. After all, a house only has an intrinsic 'front' because it is understood as a house (or at least as a building with a main entry door). The point is still that the language used in the task influenced subjects' understanding/ model/ reference frame of the represented space, to the extent that performance on an inherently spatial, geographical task was affected: the degree to which the repetition of any individual names made a difference to this is unlikely to be significant given the complexity of the tasks and displays.

This implicit assumption that the reference frame incorporates knowledge of 'what' things are, as well as 'where' they are, is still relatively uninvestigated, and is also not as simple as it may appear. As Tversky (1999) points out, 'what' versus 'where' issues are complicated by the fact that parts of any given object also have a 'where' as well as a 'what' associated with them (e.g. the front of a house), among other complications. It's not possible, therefore, to separate 'what' from 'where' in any simple way and suggest the relative importance of the two in a complex task such as map interpretation.

8.2.3 Methodological weaknesses

The two empirical phases of the current work, namely the reanalyses of data from USIS and the experimental study reported in Chapters 6 and 7, each suffered from limitations in terms of their methodologies and generalisability.

The USIS work was intended, as stated in Chapter 2, to examine HCI aspects of GIS, and to do so using established methods which were largely developed to enable system developers to make generic improvements to design. No direct capture of users' cognitive models was attempted, and so the conclusions drawn about cognitive issues were made on the basis of indirect and non-verifiable evidence (although arguably, this is always the case since we have no way of 'directly' measuring cognition without influencing it in some way).

The experimental study, on the other hand, was limited largely by practical and resource constraints. Its inability to properly investigate individual difference issues, which would ideally have been investigated in conjunction with the other experimentally manipulated factors in order to see which factors appeared to predict final task performance, was due largely to the limits on sample size. However, it was very carefully controlled, and the main effect of task/reference frame on performance is unambiguous.

8.3 Agenda for future research

Further research, to investigate the factors identified in this thesis as relevant to performance with digital maps, could include:

- An in-depth analysis of a GIS use situation, attempting to apply the more recent and more theoretically/cognitively based HCI analysis techniques to the system, paying special attention to its handling of the spatial and cartographic element.

- More research grounded in real workplaces, rather than in usability laboratories, since USIS showed that a great deal of data could be collected with very little intrusion on organisations or their workers. However, in future such studies would benefit from the use of the far superior video equipment now available (the USIS video camera was of home video standard, borrowed from an archeology Ph.D. student).

- Such studies could also utilise verbal protocols, especially if the videotaped interaction was played back to the user for them to 'talk through', since the verbal data collected in USIS was rather haphazard. More would be gained from this if care was taken to avoid biasing the users' expectations of what the researchers were interested in hearing - use of the very HCI-focused checklists in the USIS observation study meant that 'natural', potentially cognitively revealing, verbal data was unlikely.

- Longer-term observation of digital map users would allow the collection of extensive error data (frequencies, types and effects): taxonomies of errors such as Reason's (1990), based on theory in cognitive ergonomics, could again highlight the complex interaction between the various representations and requirements of complex GIS.

- At the other end of the scale, a focus on the use of public digital map systems (such as those sometimes now found in shopping centres and/or travel agents), and of internet mapping services, would help us develop an understanding of the needs of novice users of such systems - as the second part of Chapter 2 showed, many recommendations in the cartographic and ergonomic literature for digital maps seem inapplicable in many situations beyond the military battlefield or psychophysics laboratory.

- A proper individual differences study of realistic but controlled (paper or digital) map use task(s), taking a suitably large sample for factor analysis (well over 200) and examining age, expertise and some measure of spatial ability relevant to the task itself (such as reference-frame switching, as the present study attempted).

- A series of small laboratory experiments examining specific aspects of digital map design on task performance. Since many subjects commented on the unhelpful colour choices used by the Ordnance Survey (e.g. purple for road edges, red for walls of buildings, bright blue for paths but deeper blue for waterways, etc.), colour would be a prime candidate for manipulation since no previous studies have considered this at streetmap scale. Also, the degree of zoom, the amount of data (i.e. number of layers) displayed, and the familiarity of the type of layout depicted, could all make a difference to the tendency to adopt one reference frame or another.

- The starting point for the Chapter 6/7 experimental study was consideration of the extent to which typical digital map tasks require geographical interpretation of the map; having made the fairly crude distinction between 'geometric' and 'geographic' tasks, further breakdown of typical tasks (perhaps using the task taxonomies reviewed in Chapter 3) would lead to more specific testable hypotheses about the type of information which needs to be encoded, which in turn could lead to specific recommendations about which types of feature and label to emphasise in working digital maps in typical office situations (as opposed to some of the other map use scenarios which produced some of the less appropriate recommendations considered in Chapter 2).

- In terms of furthering our understanding of the psychology of map use, the issue of expertise must be better understood, not only from the psychometric point of view suggested above, but also in terms of cognitive skills and strategies which may be encouraged by training (and by cartographic education in schools, where appropriate). As stated earlier, a question about whether one feature is visible from another can be solved in more than one way, and the cognitive processes are quite different in each. Above all, these processes are largely common to other types of visual representation: the only unique aspect of maps is their geographical representation in a (normally) 2D format. This specific aspect is of great interest in environmental psychology, in terms of our understanding of space, but most other aspects of map use tasks are more general skills which may well be transferable from other types of visual task. Teasing this out would make much greater progress towards modelling 'spatial' ability (both innate and learned) from a cognitive processing perspective.

8.4 Agenda for improving digital maps and GIS

As stated earlier, it seems that GIS are becoming more uniform and predictable in terms of user interface design. The continuing attempt by GIS vendors to target specific market sectors may also help their products to be reasonably appropriate to users' tasks - hopefully a lot more so than previously, as discussed in Chapter 3. Such improvements could be helped a lot by considering organisation- and user- centred design practices (Eason, 1993), particularly when the GIS is customised for use in a specific organisation (a process which may be decreasing, except in very large organisations - the positive side of this is the transferability of users' skills, which at least two USIS users had found to be very low, from their M.Sc. in GIS to their job using an unfamiliar system).

With computer developments over the past decade, it is natural to imagine the future being less about 2D maps and more about immersive virtual environments. However, virtual environments also need to be navigated around, modelled, summarised and interpreted, even if one can choose to 'fly over' them. For example, even in the technically advanced area of computer games, maps are still part of many players' armoury. One of the most recent and (currently) most advanced computer games, *Legend of Zelda - Ocarina of Time* played on a Nintendo 64™, involves exploration of a number of fictional landscapes spread over a wide geographic area, but there is still a small (and disappointingly non-interactive) map in the corner of the screen telling you where you are in the 'world' as a whole. Some games will undoubtedly choose to make this map interactive in some way, just as with real-life maps in route navigation systems and similar applications.

Recent (as yet unpublished) work at De Montfort University by Howell Istance has been examining how people navigate round an immersive VR environment using a map, and their preferences for its orientation and symbolism. In time, the distinction between a 'virtual' environment and a 'map' of it will probably blur, but the role of a zoomable 2D overview of both real and 'virtual' space will continue to help people to access/edit information within and about it.

8.5 Conclusion

This thesis has travelled a long way in terms of disciplines, methodologies, levels of description and types of concern. Such an exercise can hardly be recommended to any sane Ph.D. student, with the amount of extra effort required to gain up-to-date knowledge of several different disciplines (e.g. the amount of database searching in 'science' (including computer science), 'social science' (including psychology) and 'arts and humanities' (including geography and cartography), besides more specialist sources). The result may be deemed unsatisfactory from a traditional single-disciplinary viewpoint, in terms of the amount of ground that would be covered if that was the only viewpoint in the thesis. However, if interdisciplinary links are not increased in applying research to real-world problems, integrated understanding of necessarily interdisciplinary problems will not be achieved. The author hopes that this thesis moves us a little further towards that integrated understanding, with regard to digital maps.

References

Albrecht, J. (1994). Universal elementary GIS tasks - beyond low-level commands. T.C. Waugh and R.G. Healey (eds.) SDH 94: Sixth International Symposium on Spatial Data Handling 5-9 Sept 1994, Edinburgh, UK. Advances in GIS Research Proceedings: Vol. 1. (pp. 209-222). London: Taylor and Francis.

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster Analysis . Quantitative Applications in the Social Sciences, vol. 07-044. Beverly Hills and London: Sage Publications.

Anderson, R. H., & Shapiro, N. Z. (1979). Design Considerations for Computer-Based Interactive Map Display Systems. (Report no. RAND/R-2382-ARPA). Santa Monica, CA, USA: Rand Corporation.

Arnheim, R. (1969). Visual Thinking. Los Angeles, USA: University of California Press.

Audet, R. H., & Abegg, G. L. (1996). Geographic information systems - implications for problem-solving. Journal of Research in Science Teaching, 33(1), 21-45.

Barnard, P. J., & May, J. (1999). Representing cognitive activity in complex tasks. Human-Computer Interaction, 14, 93-158.

Bartlett, F. C. (1932). Remembering: a study in experimental and social psychology. Cambridge, England: Cambridge University Press.

Bernhardsen, T. (1992). Geographic Information Systems. Norway: Viak IT.

Blades, M. (1997). Review Essay: Making Maps. Journal of Environmental Psychology, 17(2), 171-173.

Blades, M., & Spencer, C. (1987). Young children's strategies when using maps with landmarks. Journal of Environmental Psychology, 7, 201-217.

Boersema, T., Zwaga, H. J. G., & Jorens, K. (1993). Visual search in simple simulations of realistic scenes. Perceptual and Motor Skills, 77(3, pt 1), 867-881.

Boyce, S., & Pollatsek, A. (1992). An exploration of the effects of scene context on object identification. K. Rayner (ed.), Eye Movements and Visual Cognition: scene perception and reading (pp. 227-242). Springer Series in Neuropsychology, vol. New York: Spring-Verlag.

Brosnan, M. (1998). Technophobia: the psychological impact of information technology. London and New York: Routledge.

Brown, A. (1993). Map design for screen displays. The Cartographic Journal, 30(December), 129-135.

Burnett, G., & Joyner, S. (1995). Vehicle navigation systems: getting it right from the driver's perspective. NAV 95: Mapping and the Display of Navigation Information. The 1995 Conference of the Royal Institute of Navigation 7-9 November 1995, London. (Paper no. 25) London: Royal Institute of Navigation.

Calkins, H., & Obermeyer, N. (1991). Taxonomy for surveying the use and value of geographical information. International Journal of Geographical Information Systems, 5(3), 341-351.

Campari, I., & Frank, A. (1995). Cultural differences and cultural aspects in GIS. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems (pp. 249-266). NATO ASI Series D: Behavioural and Social Sciences, vol. Vol. 83. Dordrecht: Kluwer.

Campbell, H., & Masser, I. (1992). GIS in local government: some findings from Great Britain. International Journal of Geographical Information Systems, 6(6), 529-546.

Caplan, L. J., & Lipman, P. D. (1995). Age and gender differences in the effectiveness of map-like learning aids in memory for routes. Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 50B(3), 126-133.

Card, S., Moran, T., & Newell, A. (1983). The Psychology of Human-Computer Interaction. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Carlson-Radvansky, L., & Irwin, D. (1994). Reference frame activation during spatial term assignment. Journal of Memory and Language, 33, 646-671.

Carlson-Radvansky, L., & Radvansky, G. (1996). The influence of functional relations on spatial term selection. Psychological Science, 7(1), 56-60.

Carroll, J. (1993). Human Cognitive Abilities: a survey of factor-analytic studies. Cambridge, UK: Cambridge University Press.

Cattell, R. B. (1982). The inheritance of personality and ability: research methods and findings. New York: Academic Press.

Cave, K., & Wolfe, J. (1990). Modeling the role of parallel processing in visual search. Cognitive Psychology, 22(2), 225-271.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? Journal of Experimental Psychology - Human Perception and Performance, 11(3), 317-328.

Cocks, F. (1991). A pilot study to investigate the effectiveness of a GIS map as a visual communication device. M.Sc. thesis (for M.Sc. in Geographical Information Systems), September 1991 (Dept. of Geography). Unpublished dissertation, University of Leicester, Leicester, UK.

Coleman, D. J., Morriss, P., & Zwart, P. R. (1992). Quantifying GIS usage in an organisation: developing management tools. Proceedings of the 1992 Australasian Urban and Regional Information Systems Association (AURISA) Conference November 1992, Gold Coast, Queensland, Australia.

Cooper, L., & Regan, D. (1982). Attention, perception, and intelligence. R. J. Sternberg (ed.), Handbook of Human Intelligence (pp. 123-169). Cambridge, England: Cambridge University Press.

Cornelius, S., & Medyckyj-Scott, D. (1991). "If only someone had said!". Mapping Awareness, 5(7, September), 42-45.

Cuff, D. J., & Mattson, M. T. (1982). Thematic Maps: their design and production. New York: Methuen.

Curiel, J. M., & Radvansky, G. A. (1998). Mental organization of maps. Journal of Experimental Psychology: Learning, Memory and Cognition, 24(1), 202-214.

Czaja, S. J., & Sharit, J. (1993). Stress reactions to computer-interactive tasks as a function of task structure and individual differences. International Journal of Human-Computer Interaction, 5(1), 1-22.

Dark, V. J., Vochatzer, K. G., & Van Voorhis, B. A. (1996). Semantic and spatial components of selective attention. Journal of Experimental Psychology - Human Perception and Performance, 22(1), 63-81.

Davies, C. (1994). GIS: the view from the front line. Mapping Awareness, 8(3, April).

Davies, C. (1995). Tasks and task descriptions for GIS. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive aspects of human-computer interaction for geographic information systems (pp. 327-341). NATO Advanced Science Institutes Series D, Behavioural and Social Sciences, vol. Vol. 83. Dordrecht, Netherlands: Kluwer.

Davies, C. (1998). Analysing 'work' in complex system tasks: an exploratory study with GIS. Behaviour and Information Systems, 17(4), 218-230.

Davies, C., & Medyckyj-Scott, D. (1993). Usability as a component of product quality: the current state of Geographical Information Systems. Unpublished presentation to ESRC/AGI Workshop on GIS in local government July 8-9, 1993, Sheffield, UK.

Davies, C., & Medyckyj-Scott, D. (1993). USIS Project: Report on Questionnaire Survey of GIS Usability. Leicester, UK: Economic and Social Research Council Midlands Regional Research Laboratory, University of Leicester/Loughborough University of Technology.

Davies, C., & Medyckyj-Scott, D. (1993). The USIS Project: surveying user opinion on GIS user interfaces. EGIS '93 Conference Proceedings March 29 - April 1, 1993, Genoa, Italy.(pp. 474-483). Utrecht/Amsterdam, The Netherlands:

EGIS Foundation.

Davies, C., & Medyckyj-Scott, D. (1994). GIS usability: recommendations based on the user's view. International Journal of Geographical Information Systems, 8(2), 175-189.

Davies, C., & Medyckyj-Scott, D. (1994). Is your GIS up to scratch? GIS Europe, 3(1, February), 15.

Davies, C., & Medyckyj-Scott, D. (1995). Feet on the ground: studying user-GIS interaction in the workplace. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive aspects of human-computer interaction for geographic information systems (pp. 123-141). NATO ASI Series D; Behavioural and Social Sciences, vol. Vol. 83. Dordrecht: Kluwer.

Davies, C., & Medyckyj-Scott, D. (1996). GIS users observed. International Journal of Geographical Information Systems, 10(4), 363-384.

Davis, E. G., & Swezey, R. W. (1983). Human factors guidelines in computer graphics: a case study. International Journal of Man-Machine Studies, 18, 113-133.

De Graef, P. (1992). Scene-context effects and models of real-world perception. K. Rayner (ed.), Eye Movements and Visual Cognition: scene perception and reading (pp. 243-259). Springer Series in Neuropsychology. New York: Springer-Verlag.

de Vega, M., Intons-Peterson, M., Johnson-Laird, P., Denis, M., & Marschark, M. (1996). Models of visuospatial cognition. Oxford, UK: Oxford University Press.

de Vega, M., & Marschark, M. (1996). Visuospatial cognition: an historical and theoretical introduction. In: M. De Vega, M. J. Intons-Peterson, P. N. Johnson-Laird, M. Denis, & M. Marschark Models of Visuospatial Cognition (pp. 3-19). Counterpoints: Cognition, Memory, and Language. New York and Oxford: Oxford University Press.

Denis, M. (1996). Imagery and the description of spatial configurations. M. de Vega, M. J. Intons-Peterson, P. N. Johnson-Laird, M. Denis, & M. Marschark Models of visuospatial cognition (pp. 128-197). New York and Oxford: Oxford University Press.

Diana, E. M., & Webb, J. M. (1997). Using geographic maps in classrooms: the conjoint influence of individual differences and dual coding on learning facts. Learning and Individual Differences, 9(3), 195-214.

Diaper, D. (1989). Task Analysis for Human-Computer Interaction. Chichester: Ellis Horwood.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. Psychological Science, 8(4), 302-307.

Downs, R. M., & Stea, D. (1977). Maps in minds: reflections on cognitive mapping. New York: Harper and Row.

Dufresne, R. J., Gerace, W. J., Hardiman, P. T., & Mestre, J. P. (1992). Constraining novices to perform expertlike problem analyses: Effects on schema acquisition. Journal of the Learning Sciences, 2(3), 307-331.

Eason, K. D. (1993). Gaining user and organisational acceptance for advanced information systems. I. Masser, & H. J. Onsrud (eds.), Diffusion and use of geographic information technologies (pp. 27-44). NATO Advanced Science Institutes Series D, Behavioural and Social Sciences, vol. Vol. 70. Dordrecht, Netherlands: Kluwer.

Eastman, J., & Castner, H. (1983). The meaning of experience in task-specific map reading. D.R.F. Taylor (ed.), Graphic communication and design in contemporary cartography (pp. 115-148). Chichester: John Wiley.

Eaton, R. M. (1993). Designing the electronic chart display. The Cartographic Journal, 30(2), 184-187.

Edwards, G. (1991). Spatial knowledge for image understanding. D. M. Mark, & A. U. Frank (eds.), Cognitive and Linguistic Aspects of Geographic Space (pp. 295-307). Dordrecht, Netherlands: Kluwer.

Egan, D. E. (1988). Individual differences in human-computer interaction. M. Helander (ed.), Handbook of Human-Computer Interaction (pp. 543-568). Amsterdam: Elsevier (North-Holland).

Ekstrom, R. B., French, J. W., Harman, H. H., & Dirmen, D. (1976). Manual for kit of factor-referenced cognitive tests. Princeton NJ, USA: Educational Testing Service.

Elder, J., & Zucker, S. (1993). The effect of contour closure on the rapid discrimination of 2-dimensional shapes. Vision Research, 33(7), 981-991.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. Behavior Research Methods, Instruments, & Computers, 28(1), 1-11.

Ericsson, K. A., & Simon, H. A. (1993). Protocol Analysis: verbal reports as data 2nd edition. Cambridge MA, USA: MIT Press.

Eysenck, H. J. (1953). Uses and Abuses of Psychology. Harmondsworth, Middlesex, UK: Penguin.

Eysenck, H. J. (1990). Biological dimensions of personality. L. A. Pervin (ed.), Handbook of personality: theory and research (pp. 244-276). New York: Guilford Press.

Firns, P. G. (1990). A survey of current and potential GIS users in New Zealand. GIS 1990 International Symposium: GIS for property and environmental management - current applications and future directions 16-20 September 1990, Brisbane, Australia.(p. 5.1.1-5.1.10). Dept. of Lands, Australia.

Fisher, G. H. (1976). Measuring ambiguity. American Journal of Psychology, 80, 541-557.

Fletcher, S. J. (1999). Mapping, GIS and the armed forces. Mapping Awareness, 13(6, June/July), 37-38.

Frank, A. U. (1993). The use of geographical information systems: the user interface is the system. D. Medyckyj-Scott, & H. M. Hearnshaw (eds.), Human Factors in Geographical Information Systems (pp. 3-14). London: Belhaven Press (now part of John Wiley, Chichester).

French, J. W. (1965). The relationship of problem-solving styles to the factor composition of tests. Educational and Psychological Measurement, 25, 9-28.

Freundschuh, S., & Gould, M. (1991). Empirical user testing for validation of GIS design. GIS/LIS '91 Proceedings 28 Oct-1 Nov 1991, Atlanta, Georgia, USA.(pp. 922-931).

Gärling, T., & Golledge, R. G. (1993). Behavior and Environment: psychological and geographical approaches. Amsterdam: Elsevier.

Garnham, A. (1989). A unified theory of the meaning of some spatial relational terms. Cognition, 31, 45-60.

Gilhooly, K., Wood, M., Kinnear, P., & Green, C. (1988). Skill in map reading and memory for maps. Quarterly Journal of Experimental Psychology, 40A(1), 87-107.

Gilmartin, P. (1981). The interface of cognitive and psychophysical research in cartography. Cartographica, 18(3), 9-20.

Gilmartin, P., & Patton, J. (1984). Comparing the sexes on spatial abilities: map-use skills. Annals of the Association of American Geographers, 74(4), 605-619.

Giraudo, M., & Péruch, P. (1992). Analyses of the cognitive processes involved in the spatial representation of unskilled adults. Travail Humain, 55(2), 135-153.

Goel, V., Gold, B., Kapur, S., & Houle, S. (1998). Neuroanatomical correlates of human reasoning. Journal of Cognitive Neuroscience, 10(3), 293-302.

Goh, S., & Coury, B. (1994). Incorporating the effect of display formats in cognitive modelling. Ergonomics, 37(4), 725-745.

Golledge, R. (1992). Do people understand spatial concepts: the case of first-order primitives. Frank, A. U., Campari, I., and Formentini, U. (eds.) Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. International Conference 'GIS - from space to territory: theories and methods of spatio-temporal reasoning' 21-3 September, 1992, Pisa, Italy.Volume 639: (pp. 1-

21). New York: Springer-Verlag.

Golledge, R. (1995). Primitives of spatial knowledge. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive aspects of human-computer interaction for geographic information systems (pp. 29-44). NATO ASI Series D, Behavioural and Social Sciences, vol. Vol. 83. Dordrecht, Netherlands: Kluwer.

Gooding, C. (1993). Putting assets on the map. Financial Times, pp. 3-6.

Gothard, K. M., Skaggs, W. E., Moore, K. M., & McNaughton, B. L. (1996). Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. Journal of Neuroscience, 16(2), 823-835.

Gould, M. (1994). Defining GIS tasks: when you say what you do, do you mean what you say? GIS Europe, 3(7, Sept), 7.

Grant, S., & Mayes, T. (1991). Cognitive task analysis? G. R. S. Weir, & J. L. Alty (eds.), Human-computer interaction and complex systems (pp. 147-167). London: Academic Press.

Green, D. R. (1993). Map output from geographic information and digital image processing systems: a cartographic problem. The Cartographic Journal, 30(December), 91-96.

Griffin, T. (1983). Problem solving on maps - the importance of user strategies. Cartographic Journal, 20(2), 101-109.

Harts, J., Ottens, H., & Scholten, H. (1993). EGIS '93: Proceedings of the 4th European conference and exhibition on geographical information systems 29 March - 1 April, 1993, Genoa, Italy.Utrecht/Amsterdam, The Netherlands: EGIS Foundation.

Harvey, P. D. A. (1993). Maps in Tudor England. London: Public Record Office and The British Library.

Harwood, K. (1989). Cognitive perspectives on map displays for helicopter flight. Perspectives: Proceedings of the Human Factors Society 33rd Annual Meeting 16-20 Oct 1989, Denver Colorado, USA.(pp. 13-17). Santa Monica CA, USA: Human Factors Society.

Haunold, P., & Kuhn, W. (1994). A keystroke level analysis of a graphics application: manual map digitizing. CHI94: Human Factors in Computing Systems 24-28 April 1994, Boston MA, USA.(pp. 337-343). New York: Association for Computing Machinery.

Hayward, W., & Tarr, M. (1995). Spatial language and spatial representation. Cognition, 55, 39-84.

Hearnshaw, H., & Medyckyj-Scott, D. (1990). Psychological aspects of using geographical information systems. (Midlands Regional Research Laboratory

Research Report No. 19). Leicester, UK: MRRL.

Hirtle, S. J. J. (1985). Evidence of hierarchies in cognitive maps. Memory and Cognition, 13(3), 208-217.

Hirtle, S., & Heidorn, P. (1993). The structure of cognitive maps: representations and processes. T. M. Gärling, & R. G. Golledge (eds.), Behavior and Environment: psychological and geographical approaches (pp. 170-192). Amsterdam: Elsevier.

Hockey, G. (1990). Styles, skills and strategies: cognitive variability and its implications for the role of mental models in HCI. D. Ackermann, & M. J. Tauber (eds.), Mental models and human-computer interaction I (pp. 113-129). Amsterdam: North-Holland.

Hopkin, V. D., & Taylor, R. M. (1979). Human Factors in the Design and Evaluation of Aviation Maps. (NATO AGARDograph no. AG225). Neuilly-sur-Seine, France:

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using omega-squared and d. American Psychologist, 36, 892-901.

ISO. (1996). Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs) - Part 10: Dialogue Principles. (ISO 9241-10). Geneva, Switzerland: International Standards Organization.

Jacobson, R. (1995). The natural representation of mapping and surveying data: applying the virtual worlds paradigm. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive aspects of human-computer interaction for geographic information systems (pp. 239-246). NATO ASI Series D, Behavioural and Social Sciences, Vol. 83. Dordrecht, Netherlands: Kluwer.

Jastrow, J. (1900). Fact and fable in psychology. Boston and New York: Houghton Mifflin.

Johnson-Laird, P. (1983). Mental Models. Cambridge, UK: Cambridge University Press.

Johnson-Laird, P. N. (1996). Images, models, and propositional representations. M. de Vega, M. J. Intons-Peterson, P. N. Johnson-Laird, M. Denis, & M. Marschark Models of visuospatial cognition (pp. 90-127). New York and Oxford: Oxford University Press.

Kant, I. (1934). The Critique of Pure Reason (2nd edition). London: Dent.

Kealy, W. A., & Webb, J. M. (1995). Contextual influences of maps and diagrams on learning. Contemporary Educational Psychology, 20(3), 340-358.

Keates, J. S. (1989). Cartographic Design and Production 2nd edition. London: Longman.

Kirsch, R. A. (1994). Fifty Years of Computer Graphics and Imaging since 1957! Open lecture (unpublished) 6th May 1994. Loughborough University of Technology, Loughborough, England.

Kirwan, B., & Ainsworth, L. K. (1992). A Guide to Task Analysis. London: Taylor & Francis.

Kitchin, R. M. (1996). Are there sex differences in geographic knowledge and understanding? Geographical Journal, 162(3), 273-286.

Knapp, B. G., & Moses, F. L. (1982). Information highlighting on complex displays. Human Factors and New Technology: Proceedings of the 15th annual meeting of the Human Factors Association of Canada (pp. 13-16). Rexdale, Ontario, Canada: Human Factors Association of Canada.

Knapp, B. G., Moses, F. L., & Gellman, L. H. (1982). Information highlighting on complex displays. A. Badre, & B. Shneiderman (eds.), Directions on Human/Computer Interaction (pp. 195-215). Norwood NJ, USA: Ablex.

Knapp, L. (1995). A task analysis approach to the visualization of geographic data. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive aspects of human-computer interaction for geographic information systems (pp. 355-371). NATO ASI Series D, Behavioural and Social Sciences, Vol. 83. Dordrecht, Netherlands: Kluwer.

Kulhavy, R. W., & Stock, W. A. (1996). How cognitive maps are learned and remembered. Annals of the Association of American Geographers, 86(1), 123-145.

Landau, B., & Jackendoff, R. (1993). What and where in spatial language and spatial cognition. Behavioural and Brain Sciences, 16(2), 217-265.

Lansdale, M. (1998). Modelling memory for absolute location. Psychological Review, 105(2), 351-378.

Lansdale, M., Scrivener, S., & Woodcock, A. (1996). Developing practice with theory in HCI: applying models of spatial cognition for the design of pictorial databases. International Journal of Human-Computer Studies; 44, 777-799.

Laurini, R. (1995). Towards visual interfaces for multi-source updating of geographic information. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive aspects of human-computer interaction for geographic information systems (pp. 147-162). NATO Advanced Science Institutes Series D, Behavioural and Social Sciences, vol. Vol. 83. Dordrecht, Netherlands: Kluwer.

Lee, J. (1995). Map design and GIS - a survey of map usage amongst GIS users. The Cartographic Journal, 32(1, June), 33-44.

Lloyd, R. (1993). Cognitive processes and cartographic maps. T. Gärling, & R. G. Golledge (eds.), Behavior and Environment: Psychological and Geographical

Approaches, 141-169. Amsterdam: Elsevier.

Logan, G. (1995). Linguistic and conceptual control of visual spatial attention. Cognitive Psychology, 28, 103-174.

Lohman, D., Pellegrino, J., Alderton, D., & Regian, J. (1987). Dimensions and components of individual differences in spatial abilities. S. H. Irvine, & S. E. Newstead (eds.), Intelligence and Cognition: Contemporary Frames of Reference (pp. 253-312). NATO ASI Series D: Behavioural and Social Sciences, vol. No. 38. Dordrecht, Netherlands: Nijhoff.

Lowe, R. K. (1994). Selectivity in diagrams: Reading beyond the lines. Educational Psychology, 14(4), 467-491.

Luria, S. M., Neri, D. F., & Jacobsen, A. R. (1986). The effects of set size on colour matching using CRT displays. Human Factors, 28(1), 49-61.

Lynn, H. (1998). Start mapping. The Women's Environmental Network Newsletter, (35 (Autumn 1998)), 1.

MacEachren, A. (1995). How Maps Work: representation, visualization, and design. New York: The Guilford Press.

Makkonen, K., & Sainio, R. (1991). Computer aided cartographic communication. Proceedings, 15th International Cartographic Association Conference Bournemouth, UK.(pp. 211-222).

MapInfo®. (1997). MapInfo Product Family; 50 Ways to Use MapInfo.

Mark, D. (1993). Human spatial cognition. D. Medyckyj-Scott, & H. M. Hearnshaw (eds.), Human factors in geographical information systems (pp. 51-60). London: Belhaven Press.

Mark, D. (1993). Toward a theoretical framework for geographic entity types. Frank, A. U. and Campari, I. (ed.)COSIT '93: Spatial Information Theory (pp. 270-283). Berlin: Springer-Verlag.

Mark, D., & Egenhofer, M. (1994). Calibrating the meanings of spatial predicates from natural language: line-region relations. Waugh, T. C. and Healey, R. G. (eds.)SDH '94: Sixth International Symposium on Spatial Data Handling 5-9 September 1994, Edinburgh, Scotland. Advances in GIS Research Proceedings: (pp. 538-553). London: Taylor and Francis.

Mark, D., & Frank, A. (1990). Language, Cognitive Science, and Geographic Information Systems. (NCGIA Report 90-10). University of California, Santa Barbara CA 93106 USA: National Center for Geographic Information and Analysis.

Mark, D., & Frank, A. (1996). Experiential and formal models of geographic space. Environment and Planning B: Planning and Design, 23, 3-24.

Mark, D., & Freundschuh, S. (1995). Spatial concepts and cognitive models for geographic information use. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive aspects of human-computer interaction for geographic information systems (pp. 21-28). NATO Advanced Science Institutes Series D, Behavioural and Social Sciences, Vol. 83. Dordrecht, Netherlands: Kluwer.

Marschark, M., & Hunt, R. (1989). A reexamination of the role of imagery in learning and memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 15(4), 710-720.

Marschark, M., Richman, C., Yuille, J., & Hunt, R. (1987). The role of imagery in memory: on shared and distinctive information. Psychological Bulletin, 102(1), 28-41.

McCann, C. (1982). Classification of Map-Reading Tasks. (DCIEM Report No. 82-R-59). Canada: Dept. of National Defence.

McDonald, T., & Pellegrino, J. (1993). Psychological perspectives on spatial cognition. T. Gärling, & R. G. Golledge (eds.), Behavior and Environment: psychological and geographical approaches (pp. 47-82). Amsterdam: Elsevier.

McGuinness, C. (1994). Expert/novice use of visualization tools. A. M. MacEachren, & D. R. F. Taylor (eds.), Visualization in Modern Cartography (pp. 185-199). Modern Cartography Series, vol. Volume 2. Oxford: Pergamon - Elsevier Science Ltd.

Medyckyj-Scott, D. (1991). "What have I done now?". J. Cadoux-Hudson, & I. Heywood (eds.), Geographic Information 1991: The Yearbook of the Association for Geographic Information (pp. 235-241). London: Taylor and Francis.

Medyckyj-Scott, D., & Blades, M. (1990). Users' Cognitive Representations of Space: Relevance to the Design and Use of GIS. (Midlands Regional Research Laboratory Report No. 13, December 1990). Leicester, UK: MRRL.

Medyckyj-Scott, D., & Blades, M. (1991). Cognitive representations of space in the design and use of geographical information systems. People and Computers VI: Proceedings of the HCI '91 Conference 20-23 August 1991, British Computer Society Conference Series : Vol. 3. (pp. 421-433). Cambridge, UK: Cambridge University Press (on behalf of BCS).

Medyckyj-Scott, D., Davies, C., & Byrne, V. (1993). Discovering the trials of GIS use: the USIS study of GIS usability (Unpublished paper presentation). GIS Research UK Conference March 1993, Keele University, England.

Medyckyj-Scott, D., & Hearnshaw, H. (1993). (eds.), Human factors in geographical information systems . Chichester, UK: Belhaven, now John Wiley.

Monmonier, M. (1991). How to Lie with Maps 1st ed.,). Chicago and London: University of Chicago Press.

Monmonier, M. (1993). Mapping It Out: expository cartography for the humanities and the social sciences. Chicago, USA: University of Chicago Press.

Monmonier, M. (1994). [Letter to Medyckyj-Scott, D.].

Monmonier, M. (1996). How to Lie with Maps 2nd ed. Chicago and London: University of Chicago Press.

Montello, D. (1993). Scale and multiple psychologies of space. Frank, A. U. and Campari, I. (eds.)COSIT '93: Spatial Information Theory (pp. 312-321). Berlin: Springer-Verlag.

Montello, D. R., & Golledge, R. G. (1998). Summary Report: "Scale and Detail in the Cognition of Geographic Information" (Project Varenius Workshop, May 14-16, 1998). Santa Barbara CA, USA: National Center for Geographic Information and Analysis.

Montello, D. R., Lovelace, K. L., Golledge, R. G., & Self, C. M. (1999). Sex-Related Differences and Similarities in Geographic and Environmental Spatial Abilities. Annals of the Association of American Geographers, 89(3), 515-534.

Moscovitch, M., Kapur, S., Kohler, S., & Houle, S. (1995). Distinct neural correlates of visual long-term-memory for spatial location and object identity - a positron emission tomography study in humans. Proceedings of the National Academy of Sciences of the United States of America , 92(9), 3721-3725.

Moses, F. L., & Maisano, R. E. (1978). User performance under several automated approaches to changing displayed maps. Computer Graphics, 12(3), 228-233.

Muehrcke, P. (1978). Map use: reading, analysis and interpretation. Madison WI, USA: J.P. Publications.

MUSiC. (1992). The Performance Metrics Toolkit User Guides. Teddington, Middlesex, UK: National Physical Laboratory HCI Group, DITC.

Neisser, U. (1976). Cognition and Reality. San Francisco: W.H. Freeman and Co.

Nielsen, J., & Mack, R. L. (1994). Usability Inspection Methods. Chichester, UK: John Wiley.

Nunes, J. (1991). Geographic space as a set of concrete geographical entities. D. M. Mark, & A. U. Frank (eds.), Cognitive and Linguistic Aspects of Geographic Space (pp. 9-33). Dordrecht, Netherlands: Kluwer.

Nyerges, T. (1991). Analytical map use. Cartography and Geographic Information Systems, 18(1), 11-22.

Nyerges, T. L. (1993). How do people use geographical information systems? D. Medyckyj-Scott, & H. M. Hearnshaw (eds.), Human factors in geographical information systems (pp. 37-50). London: Belhaven (now part of John Wiley, Chichester).

Nyerges, T. L. (1995). Cognitive issues in the evolution of GIS user knowledge. T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (eds.), Cognitive aspects of human-computer interaction for geographic information systems (pp. 61-74). NATO Advanced Science Institutes Series D, Behavioural and Social Sciences, vol. Vol. 83. Dordrecht, Netherlands: Kluwer.

Nyerges, T., Mark, D., Laurini, R., & Egenhofer, M. (1995). Cognitive aspects of human-computer interaction for geographic information systems. Dordrecht, Netherlands: Kluwer.

Oracle®. (1997). Oracle8™ Spatial Cartridge: Advances in Relational Database Technology for Spatial Data Management. An Oracle Technical White Paper, June 1997. Redwood Shores, CA, USA: Oracle Corporation.

Orne, M. (1962). On the social psychology of the psychological experiment: with special reference to demand characteristics and their implications. American Psychologist, 17, 776-783.

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? Psychonomic Science, 11(4), 137-138.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). EXPE: an expandable programming language for on-line psychology experiments. Behavior Research Methods, Instruments and Computers, 29(3), 322-327.

Pederson, E. (1993). Geographic and manipulable space in two Tamil linguistic systems. COSIT '93 proceedings (pp. 294-311).

Pederson, E. (1995). Language as context, language as means: spatial cognition and habitual language use. Cognitive Linguistics, 6(1), 33-62.

Peterson, M. (1987). The mental image in cartographic communication. Cartographic Journal, 24(1), 35-41.

Peterson, M. A., Kihlstrom, J. F., Rose, P. M., & Glisky, M. L. (1992). Mental images can be ambiguous: reconstruals and reference-frame reversals. Memory and Cognition, 20(2), 107-123.

Pick, H. J., & Lockman, J. (1983). Map reading and spatial cognition: discussion. H. L. Pick, & L. P. Acredolo (eds.), Spatial Orientation: theory, research, and application (pp. 219-224). New York and London: Plenum Press.

Pinker, S. (1998). How the Mind Works. London: Allen Lane.

Pinto, J. K., & Onsrud, H. J. (1993). Correlating adoption factors and adopter characteristics with successful use of geographic information systems. I.

Masser, & H. J. Onsrud (eds.), Diffusion and Use of Geographic Information Technologies (pp. 165-194). NATO Advanced Science Institutes Series D, Behavioural and Social Sciences, Vol. 70. Dordrecht, Netherlands: Kluwer.

Potash, L. (1977). Design of maps and map-related research. Human Factors, 19(2), 139-150.

Raper, J. F., & Bundock, M. S. (1991). UGIX: A layer based model for a GIS user interface. D. M. Mark, & A. U. Frank (eds.), Cognitive and Linguistic Aspects of Geographic Space (pp. 449-475). Dordrecht, Netherlands: Kluwer.

Rasmussen, J. (1986). Information processing and human machine interaction: an approach to cognitive engineering. New York: North-Holland.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). Cognitive Systems Engineering. New York/ Chichester: John Wiley.

Ravden, S., & Johnson, G. (1989). Evaluating usability of human-computer interfaces: a practical method. Chichester, UK: Ellis Horwood.

Remington, R., & Williams, D. (1986). On the selection and evaluation of visual display symbology: factors influencing search and identification times. Human Factors, 28(4), 407-420.

Rittschof, K. A., & Kulhavy, R. W. (1998). Learning and remembering from thematic maps of familiar regions. ETR&D - Educational Technology Research and Development, 46(1), 19-38.

Robinson, A. H., Sale, R. D., Morrison, J. L., & Muehrcke, P. C. (1984). Elements of Cartography Fifth ed. New York: John Wiley and Sons.

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

Rossano, M. J., & Hodgson, S. L. (1994). The process of learning from small-scale maps. Applied Cognitive Psychology, 8(6), 565-582.

Rubin, E. (1915). Synsoplevede Figurer. Copenhagen: Glydendalska.

Schlichtmann, H. (1991). Plan information and its retrieval in map interpretation: the view from semiotics. D. M. Mark, & A. U. Frank (eds.), Cognitive and linguistic aspects of geographic space (pp. 263-284). Dordrecht: Kluwer.

Schofield, N. J., & Kirby, J. R. (1994). Position location on topographical maps - effects of task factors, training, and strategies. Cognition and Instruction, 12(1), 35-60.

Scott, D. (1987). Mental imagery and visualisation: their role in map use. Unpublished doctoral dissertation, Dept. of Geography, London School of

Economics and Political Science, London.

Self, C., Gopal, S., Golledge, R., & Fenstermaker, S. (1992). Gender-related differences in spatial abilities. Progress in Human Geography, 16(3), 315-342.

Sharit, J., & Czaja, S. J. (1994). Ageing, computer-based task performance, and stress: issues and challenges. Ergonomics, 37(4), 559-577.

Sharps, M., Welton, A., & Price, J. (1993). Gender and task in the determination of spatial cognitive performance. Psychology of Women Quarterly, 17(1), 71-83.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science, 171, 701-703.

Shepard, R., & Hurwitz, S. (1985). Upward direction, mental rotation and discrimination of left and right turns in maps. S. Pinker (ed.), Visual Cognition (pp. 161-193). Cambridge, MA, USA: MIT Press.

Shepherd, A. (1989). Analysis and training in information technology tasks. D. Diaper (ed.), Task analysis for human-computer interaction (pp. 15-55). Chichester, England: Ellis Horwood.

Sherry, D. F., & Hampson, E. (1997). Evolution and the hormonal control of sexually-dimorphic spatial abilities in humans. Trends in Cognitive Sciences, 1(2), 50-56.

Shneiderman, B. (1992). Designing the User Interface: strategies for effective human-computer interaction Second ed. Reading MA, USA: Addison-Wesley.

Sholl, M., & Egeth, H. (1982). Cognitive correlates of map-reading ability. Intelligence, 6, 215-230.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. London: McGraw-Hill.

Spearman, C. (1927). The Abilities of Man: their nature and measurement. New York: Macmillan.

Spencer, C., Blades, M., & Morsley, K. (1989). The Child in the Physical Environment. Chichester: John Wiley.

Spiker, A., Rogers, S. P., & Cicinelli, J. (1986). Selecting colour codes for a computer-generated topographic map based on perception experiments and functional requirements. Ergonomics, 29(11), 1313-1328.

Spooner, R. (1999). GIS is dead. GEOEurope, 8(4, May), 20-21.

Springer, S. P., & Deutsch, G. (1998). Left brain, right brain: Perspectives from cognitive neuroscience (5th edition ed.). New York, NY, USA: W. H. Freeman and Co, Publishers.

Stumpf, H., & Eliot, J. (1995). Gender-related differences in spatial ability and the k factor of general spatial ability in a population of academically talented students. Personality and Individual Differences, 19(1), 33-45.

Sutherland, S. (1992). Irrationality: the enemy within. London: Constable.

Temple, Barker, & Sloane (Inc.). (1990). The Benefits of the Graphical User Interface: a Report on New Primary Research. Lexington, MA, USA:

Theeuwes, J. (1993). Visual selective attention - a theoretical analysis. Acta Psychologica, 83(2), 93-154.

Thomson, N. (1985). Human memory: different stores with different characteristics. A. Monk (ed.), Fundamentals of Human-Computer Interaction (pp. 49-56). Computers and People Series, vol. London: Academic Press.

Thorndyke, P., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. Cognitive Psychology, 12, 137-175.

Thurstone, L. L. (1938). Primary Mental Abilities. (Report No. Psychometric Monographs, No. 1).

Tinbergen, N. (1948). The study of instinct . Oxford: Oxford University Press.

Tkacz, S. (1998). Learning map interpretation: skill acquisition and underlying abilities. Journal of Environmental Psychology, 18(3), 237-249.

Toon, M. (1998). The object of its attention. Mapping Awareness, 12(8, October), 12.

Tufte, E. (1990). Envisioning Information. Cheshire, CT, USA: Graphics Press.

Turk, A. (1990). Towards an understanding of human-computer interaction aspects of geographic information systems. Cartography, 19(1), 31-60.

Turk, A. (1992). GIS cogency: cognitive ergonomics in geographic information systems. Unpublished doctoral dissertation, Dept. of Surveying and Land Information, University of Melbourne, Melbourne, Australia.

Turk, A., & Mackaness, W. (1993). The cognitive ergonomics of computer-assisted visualization design. Proceedings of National Center for Geographic Information and Analysis, I-8 Specialist Meeting October 1993, (pp. 211-222). Santa Barbara CA, USA: NCGIA.

Turkle, S. (1984). The Second Self: computers and the human spirit. New York: Simon and Schuster.

Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Tversky, B. (1992). Distortions in cognitive maps. Geoforum, 23(2), 131-138.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. Frank, A. U. and Campari, I. (eds.)COSIT '93: Spatial Information Theory (pp. 14-24). Berlin: Springer-Verlag.

Tversky, B. (1993). Distortions in memory for visual displays. S. R. Ellis, M. K. Kaiser, & A. C. Grunwald (eds.), Pictorial Communication in Virtual and Real Environments 2nd edition ed. (pp. 61-75). London: Taylor and Francis.

Tversky, B. (1999) [personal communication]

Unwin, D. J. (1997). Graphics, visualisation and the social sciences. Graphics, Visualization and the Social Sciences: workshop report 8-9 May 1997, Loughborough University, U.K.AGOCG Technical Report Series: Number 33, June 1997. (pp. 2-103-2/108). Loughborough: Advisory Group on Computer Graphics.

Vecchi, T., & Girelli, L. (1998). Gender differences in visuo-spatial processing: the importance of distinguishing between passive storage and active manipulation. Acta Psychologica, 99(1), 1-16.

Venturino, M., & Geiselman, E. (1992). Quantifying the goodness of mental representations of spatial relationships. Innovations for Interactions: Proceedings of the Human Factors Society 36th annual meeting 12-16 Oct 1992, Atlanta, Georgia, USA.(pp. 1363-1367). Santa Monica CA, USA: Human Factors Society.

Visvalingam, M. (1992). Cartography and geographical information system. The Cartographic Journal, 29(1), 51-54.

Wagenaar, W. A., Keren, G., & Lichtenstein, S. (1988). Islanders and hostages: Deep and surface structures of decision problems. Acta Psychologica, 67(2), 175-189.

Ware, C. (1988). Color sequences for univariate maps: theory, experiments, and principles. IEEE Computer Graphics and Applications, 8(5, September), 41-49.

Whitefield, A., Esgate, A., Denley, I., & Byerley, P. (1993). On distinguishing work tasks and enabling tasks. Interacting With Computers, 5(3), 333-347.

Whitefield, A., & Hill, B. (1994). Comparative analysis of task analysis products. Interacting With Computers, 6(3), 289-309.

Whittlesea, B. W. A., & Wright, R. L. (1997). Implicit (and explicit) learning: acting adaptively without knowing the consequences. Journal of Experimental Psychology: Learning, Memory and Cognition, 23(1), 181-200.

Widdel, H. (1990). Subjective evaluation of color design in digital map displays. Berlinguet, L. and Berthelette, D. (eds.)Work with Display Units 89 Amsterdam: Elsevier (North-Holland).

Wiegand, P., & Stiell, B. (1997). Mapping the place knowledge of teachers in training. Journal of Geography in Higher Education, 21(2), 187-198.

Williams, R. (1989). Geographic information: aspects of phenomenology and cognition. Proceedings of Auto Carto 9, 9th International Symposium on Computer-Assisted Cartography 2-7 April 1989, Baltimore, MD, USA.(pp. 557-566). Falls Church VA, USA: American Society for Photogrammetry and Remote Sensing/American Congress on Surveying and Mapping.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). A manual for the Embedded Figures Tests. Palo Alto CA, USA: Consulting Psychologists Press.

Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. Vision Research, 34(9), 1187-1195.

Wolfe, J., Cave, K., & Franzel, S. (1989). Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception and Performance, 15(3), 419-433.

Wong, G.-K., & Rengger, R. (1990). The Validity of Questionnaires Designed to Measure User-Satisfaction of Computer Systems. (NPL Report DITC 169/90). Teddington, Middlesex, England: National Physical Laboratory.

Wood, M. (1993). Interacting with maps. D. Medyckyj-Scott, & H. M. Hearnshaw (eds.), Human factors in geographical information systems (pp. 111-123). London: Belhaven Press.

Worth, C. (1988). Subjective and objective in map design research. Proceedings, Austra Carto III, 7th Australian Cartographic Conference (pp. 592-600). Sydney: Australian Institute of Cartographers.

Worth, C. (1992). Core concepts in map design of relevance to geographical information systems. Proceedings of the seminar on teaching surveying and mapping aspects of visualisation and GIS 6-8 August 1992, University of Maryland, USA.International Cartographic Association Commission on Education and Training.

Wright, R. L., & Whittlesea, B. W. A. (1998). Implicit learning of complex structures: active adaptation and selective processing in acquisition and application. Memory and Cognition, 26(2), 402-420.

Young, H. H., & Miller, J. T. (1991). Visual discrimination on colour VDTs at two viewing distances. Behaviour and Information Technology, 10(3), 191-205.

Appendix 1: questions and categories in the original USIS survey

Below is a list of the questions included in the original USIS survey (see Chapter 2), which was developed not by the present author but by David Medyckyj-Scott and Hilary Hearnshaw (see Davies and Medyckyj-Scott, 1993 for full details of its analysis). No formatted copies survive, so the response format for each question is explained in plain italics below. The first page of instructions has been omitted; it basically asked that END-USERS complete the survey (rather than management), and encouraged them to copy it to other users. A number of users did so, particularly in continental European countries; some passed it to people outside their own organisation or even their own country, who then also responded. Note that anything in plain italics below was not actually printed in the survey questionnaire, and is for clarification.

Part A: You and Your GIS

Something about you the user

Q1 How many different types of computer have you used (e.g. Sun Workstation, IBM PC, Apple Macintosh etc.)?

Multi-choice response: Less than 3; 3-6; More than 6

Q2 How do you rate your experience of using computers?

Multi-choice response: Low; Medium; High

Q3 How do you rate your knowledge of GIS?

Multi-choice response: Low; Medium; High

Something about the system

Q4 What is the name of the Geographical Information System you currently use? (If you currently use more than one OR one in conjunction with another, please fill in a questionnaire for each)

free-text response

In the remainder of this questionnaire please base your responses and any comments on the use of the system you have named above.

Q5 What language does the GIS use (e.g. English, French, etc.)?

free-text response

Q6 How fluent do you consider yourself in the language of the GIS?

multi-choice response: Beginner; Intermediate; Fluent

Q7 In which language is the documentation written?

free-text response

Q8 How regularly do you use the GIS?

multi-choice response: Every day; At least once a week; At least once a month; Less than once a month

Q9 How is your GIS implemented?

multi-choice response: On a mainframe; On a minicomputer; On a workstation (e.g. a Sun); On a PC; On a distributed network system; Other (Please give details)

Q10 How long has the GIS been installed in your organisation?

respondent prompted for years and months

Q11 How long have YOU been using the system?

respondent prompted for years and months

Q12 Which version of the GIS software you are using?

free-text response

Q13 For what sort of tasks do you use the GIS at present ?

free-text response

Q14 What form of training, if any, did you receive?

free-text response

Q15 Who provided the training you received?

multi-choice response: The GIS supplier; Other staff from within your organisation; Self-taught; Other (Please specify)

Q16 Please describe anything about the training that you feel was lacking.

free-text response

Q17a Has the user interface of the GIS been customised (specially modified) in any way?

yes/no boxes

Q17bi If yes, what was done?

free-text response

Q17bii Why was the GIS customised?

free-text response

Part B: Using the GIS B1 - Rating scale items

Instruction page omitted - explained use of Likert scale and use of 'Comments' column which was to the right of the response scale. The numbers 1 to 5 were given to the right of each question, and 'Strongly disagree' and 'Strongly agree' were marked above 1 and 5 at the top of each page.

NB In the list below, + and - signs indicate whether statements were left in the original positive, or reversed to negative to prevent response generalisation by users (this process was applied randomly). Where an item was adapted from Digital's System Usability Scale, this is indicated in italic square brackets after the statement, but was not distinguished on the actual questionnaire. 'GIS' was substituted for the original word 'system' where appropriate in the SUS items. Also please note that the headings in square brackets indicate our groupings of items by topic; the unbracketed headings are those actually shown on the questionnaire.

The Training

[Training quality:]

Q18 + The training was very useful in helping me understand the GIS

Q19 + The training was very relevant to the work I now do with the GIS

Q20 - The training was not useful in helping me to use the GIS

[Training documents:]

Q21 + The terms used in the training documentation were ones with which I was familiar

Q22 - The training documentation was not helpful

Q23 - I was unfamiliar with the ideas described in the training documentation

The Documentation

[User guide:]

Q24 + The User Guide contains all the information necessary to use the GIS

Q25 - The information in the User Guide is hard to find

Q26 + The information in the User Guide is easy to understand

Q27 - The information in the User Guide contains inaccuracies

[System manual:]

Q28 + The System Documentation provides all the information to set up and run the GIS

Q29 + The information in the System Documentation is easy to find

Q30 - The information in the System Documentation is ambiguous

Q31 + The information in the System Documentation is accurate

[All documentation:]

Q32 - There were general inconsistencies between the different documents provided

The Hardware

[Hardware:]

Q33 + The initial installation of the GIS was problem free

Q34 - It was difficult to make the GIS software work on our hardware

It was easy to link up each of the following devices with the GIS:

Q35 + A Digitiser

Q36 + A Scanner

Q37 + A Plotter

Q38 + A Printer

Q39 - Producing hardcopy (paper) output is difficult

Using the GIS

[Interaction:]

Q40 + I like to use the GIS frequently *[SUS 1]*

Q41 + I think the GIS is easy to use *[SUS 3]*

Q42 - The GIS does not do all that I need it to

Q43 + I find the various functions of the GIS well integrated *[SUS 5]*

Q44 - I find the GIS very cumbersome to use *[SUS 8]*

Q45 + I feel very confident when using the GIS *[SUS 9]*

[Controllability:]

Q46 + The GIS always keeps me informed about what it is doing

Q47 - I worry that the GIS will crash while I am working with it

Q48 + I can easily undo an action I just started

Q49 - I sometimes find myself in parts of the GIS I hadn't expected

Q50 + I can generally understand what the GIS is doing

[Error tolerance:]

Q51 + I understand all error messages when they appear

Q52 - Error messages are not helpful

Q53 - Additional explanations about an error are not easily available

Q54 + I can escape out of any situation I want to

Q55 - Messages from the GIS never seem to mean what they say

[Suitability:]

Q56 + The screen always displays information in the way I expect

Q57 + The output from the GIS matches my task requirements

Q58 - I find the GIS unnecessarily complex *[SUS 2]*

Q59 + The standard defaults values are appropriate

Q60 + Tasks can be performed in a straightforward manner

[Predictability:]

Q61 - I think there is too much inconsistency in the GIS *[SUS 6]*

Q62 + The response times of the system meet my expectations

Q63 + I always know what input is expected

Q64 - The position of messages on the screen is inconsistent

Q65 + Output on the screen matches the hardcopy when it is printed/plotted

[Flexibility:]

Q66 - It is difficult to switch from one task to another when I want to

Q67 + I can change the way I input information to suit the way I want to work

[Self-descriptiveness:]

Q68 - The GIS does not warn me about performing a task with severe consequences

Q69 + The GIS allows me to vary the amount of information I receive in messages

Q70 - When using the GIS I frequently forget what I have to do next

Q71 - I frequently have to refer to the user or system documentation

[Learnability:]

Q72 - The GIS was difficult to learn to use

Q73 - I needed to learn a lot of things before I could get going with the GIS *[SUS 10]*

Q74 - Exploring new features by trial and error is difficult

Q75 + I would imagine that most people learn to use this GIS very quickly *[SUS 7]*

Q76 - I needed to learn a lot of new ideas before I could use the GIS

[Individualisation:]

Q77 + The GIS is such that I can change the user interface to suit my preferences

Q78 + It is possible to construct new functions (macros) if I require them

Q79 - The macros I constructed should have been part of the system functions

Q80 + The GIS allows me to rename commands to suit my preferences

Q81 - The GIS does not allow me to adapt it to suit my preferred language (e.g. German)

The Help Facilities

[Help:]

Q82 + Information in the Help facilities is easy to understand

Q83 - The information in the Help facilities is hard to find

Q84 + The help I receive from the Help facilities is related to what I am trying to do

Q85 - I feel I need the support of a technical person when I use the GIS SUS 4

B2 - Final questions

[NB each of these had several lines provided for free-text responses]

Q86 What are the best usability aspects of your GIS?

Q87 What are the worst usability aspects of your GIS?

Q88 Please add any further comments you wish to make.

Q89 Finally, we will be looking for a small number of organisations to take part in a more detailed survey. This will involve one of our researchers talking to you in depth and observing you at work with the GIS. If you would be willing to take part please complete the details below. Please return the completed questionnaires to us in the envelope provided. Thank you very much for your help.

Appendix 2: Copy of draft cartographic quality checklist

The checklist on the next several pages was developed jointly by the present author, Erica Milwain (who produced its formatting), Peter Robinson and David Medyckyj-Scott. See Chapter 2 for details of its construction and evaluation.

Context Questions

User I.D...

1.1	What hardware is being used?					
1.2	What GIS is being used?	Name	Version			
1.3	How many monitors are being used?					
1.4	What size is/are the screen(s)?	Width	Height			
1.5	What is the resolution of the screen(s)?	High	Medium	Low	Value (if known)	
1.6	How many screen colours does the monitor support?					
1.7	What input devices are being used? P = primary S = secondary	Keyboard	Mouse	Puck	Other	
1.8	What is the application domain (e.g. roads, gas supply, etc.)?					
1.9	What is the task domain? (What is the user doing?)					
1.10	What stage has the system implementation reached ?	Pilot	Data input	Fully in use (data in use)		
1.11	How experienced in GIS is the user?	V. experienced	Experienced	Intermediate	Novice	Comment
1.12	How experienced in cartography is the user?	V. experienced	Experienced	Intermediate	Novice	Comment

System Questions

		Yes/ Always	Mostly	Sometimes	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
2	**Working Environment**									
2.1	Is the working area correctly lit?		▨	▨	▨				▨	
2.2	Does the GIS user have enough work (desk) space?		▨	▨	▨				▨	
2.3	Is there enough room for the GIS user to move around in the work area?		▨	▨	▨				▨	
2.4	Is the user's desk suitable?		▨	▨	▨				▨	
2.5	Is the user's seating suitable?		▨	▨	▨				▨	
3	**Screen/Monitor**									
3.1	Is the size of the screen great enough to show an adequate portion of the image for comfortable working?		▨	▨	▨				▨	
3.2	Is there enough space on the screen for all the windows that are necessary, without causing clutter?								▨	
3.3	Is the amount of screen flicker likely to cause eye strain?	▨	▨	▨	▨	▨				
3.4	Is the user able to adjust screen brightness?	▨	▨	▨	▨	▨				
4	**System Usability**									
	• Response Times:									
4.1	Is the speed of zooming-in sufficient? (less than 4 seconds)								▨	

System Questions

		Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
4.2	Is the speed of zooming-out sufficient? (less than 4 seconds)								▨	
4.3	Is the speed of panning sufficient? (less than 4 seconds)								▨	
4.4	Is the speed of a requested redraw sufficient? (less than 12 seconds)								▨	
4.5	If the map is dynamic, is its speed of change adequate?								▨	
4.6	Does the response time of the system match the expectations of the user?								▨	

- **Information Organisation:**

4.7	Does the display include only information which is essential to the user's needs?								▨	
4.8	Is there a facility to select different types of information for display? (e.g. layers)		▨	▨	▨				▨	
4.9	If 'yes' to the above, is there clear naming of selected information (e.g. in the legend)?		▨	▨	▨				▨	
4.10	Is there any clear indication of the order of selected information?		▨	▨	▨				▨	
4.11	Is the user able to eliminate irrelevant items from the display (e.g. by turning off data layers) and to reverse these decisions?								▨	
4.12	Are there any floating windows or tool palettes?		▨	▨	▨				▨	
4.13	Are windows tiled (avoiding overlapping one another)?								▨	
4.14	Does the layout avoid windows overlapping the map?								▨	

System Questions

		Yes/ Always	Mostly	Some- times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
4.15	Is the user able to display grid lines as a background feature?		▨	▨	▨				▨	
4.16	Where grid lines are available for display, do they provide an illusion of continuity?		▨	▨	▨				▨	
4.17	Is it easy to correct mistakes and work flexibly, rather than rigidly pre-planning tasks?	▨	▨	▨	▨	▨				

- **Navigation**

4.18	Can the appropriate map easily be located and retrieved?								▨	
4.19	Can the user easily ascertain where he/she is on the map? (Use of Viewports?)		▨	▨	▨				▨	
4.20	Where related information is in seperate windows is that relationship obvious?		▨	▨	▨				▨	
4.21	Is panning of the screen smooth?		▨	▨	▨				▨	
4.22	Does the system support automatic pan when the user reaches the edge of the screen?		▨	▨	▨				▨	
4.23	When moving or panning between geographic areas, is there an overlap between successively displayed views?								▨	
4.24	Where **speed of user performance** is important, and the user pans the display across between geographic areas, is fixed-amount panning used with a 25% overlap, rather than continuous user-controlled panning?(should be fixed with 25% overlap as this improves response times)		▨	▨	▨				▨	

System Questions

		Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
4.25	Where frequency of redrawing the displayed map needs to be minimised (e.g. where redrawing is very slow), is continuous panning or fixed-amount panning with a 0% overlap used, rather than fixed-amount panning with an overlap? (should be continuous under user control, or fixed at 0%, as either minimises redraw)		▨	▨	▨				▨	
4.26	Where **speed of user and system performance** are both unimportant, is continuous panning available?		▨	▨	▨				▨	
4.27	Is the identity of an area ever lost when zooming in? (When area boundaries are beyond edge of screen).								▨	

- **Interaction**

4.28	Is a sufficient level of magnification possible to perform the task?	▨	▨	▨	▨	▨				
4.29	Is the working scale (amount of zoom) appropriate to the task?								▨	
4.30	Are marked differences in the appearance of the display at different levels of magnification avoided?								▨	
4.31	Is the cursor easily visible?								▨	
4.32	Is cursor size adjustable?	▨	▨	▨	▨	▨	▨	▨		
4.33	Is there a cross-hair if high accuracy is required?								▨	
4.34	Does the cursor avoid obscuring information?								▨	
4.35	Can the snap distance be altered?	▨	▨	▨	▨	▨	▨	▨		
4.36	Are there appropriate macros to speed things up?								▨	

System Questions

		Yes/ Always	Mostly	Sometimes	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
4.37	Is there a facility to easily place point symbols?		▨	▨	▨				▨	
4.38	Is there a facility for attaching information to a point symbol?		▨	▨	▨				▨	
4.39	Is it easy to design styles for lines, symbols etc.?	▨	▨	▨	▨	▨	▨	▨		
4.40	Is it easy to make polygons from existing lines?	▨	▨	▨	▨	▨	▨	▨		
4.41	Is it easy to place text and fit it to a space?	▨	▨	▨	▨	▨	▨	▨		
4.42	Is it easy to change line type/symbol type etc.?	▨	▨	▨	▨	▨	▨	▨		
4.43	Is it easy to correct mistakes, immediately or later, without redrawing the object?								▨	

- **Highlighting**

4.44	Are special areas of the display distinguished using either brightness/contrast, colour coding or surrounded with a box?								▨	
4.45	Where the user is warned of the consequences of a potential action, is this highlighted using either brightness/contrast, different character sizes, colour coding or surrounding with a box?								▨	Which?
4.46	Where the user's attention is to be drawn to a particular point on a display filled with information, is a moving or flashing cursor used?								▨	
4.47	Are high priority messages and codes highlighted using either brightness/contrast, different character sizes, colour coding or surrounding with a box?								▨	Which?
4.48	Is 'alarm' information indicated by blinking or pulsating of relevant symbols/objects? (e.g. object selection)								▨	

System Questions

		Yes/ Always	Mostly	Sometimes	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
4.49	Are unusual values highlighted using either brightness/contrast, different character size or colour coding?								▨	
4.50	Are command or data entry errors highlighted by brightness/contrast, different character sizes, or colour control?								▨	Which?
4.51	Where specific information has been changed or is about to be changed, is this highlighted using either brightness/contrast, different character size or colour coding?								▨	Which?
4.52	Where the user has performed a search, are the identified targets highlighted using either brightness/contrast, inverse video display, or colour coding?								▨	Which?
4.53	Is highlighting of any kind restricted to only one or two types of object at all times?								▨	
4.54	Where the user's attention is being drawn to a message, is this done without making the whole message flash?								▨	
4.55	Where flashing is used for anything (e.g. highlighting etc.) can the user always turn this off?	▨	▨	▨	▨	▨	▨	▨		

- Hard Copy

4.56	If a hard copy is printed or plotted does the map on the screen look the same as the hard copy (WYSIWYG)?		▨	▨	▨				▨	

Map/Task Questions

5 How did you choose the map(s) to evaluate?

6 Task Suitability

		Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
6.1	Is the GIS system appropriate to the intended task?		▨	▨	▨					
6.2	Is the user adequately trained in the use of the system to be able to perform the task required?		▨	▨	▨					
6.3	Does the user have sufficient domain knowledge to perform his/her task?		▨	▨	▨					
6.4	Has the user had sufficient cartographic training to perform his/her task?		▨	▨	▨					
6.5	Is the map on the screen in a form familiar to the user?		▨	▨	▨					
	• **Information Availability**									
6.6	Is all information required for the task present in the work area (screen) at all times (including legend)?								▨	
6.7	Does interim data produced during a process or calculation disappear from the screen once it is no longer needed?								▨	

Map/Task Questions

7 Cartographic Variables

- **General**

7.1	What type of map is being used?	Topographic	Plan	Thematic	Other
7.2	Is a base map used?	Yes	No	Comment	
7.3	What is the intended final product of this task?	Working map (on screen)	Screen presentation for secondary users	Draft hard copy	Final hard copy

		Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.4	Does the user understand how the map is set up?		▨	▨	▨					
7.5	Is the cartographic method being employed appropriate to the task?									
7.6	Is the user aware of limitations to design like available format size, line widths, lettering styles and sizes, colours etc. and how they will print?	▨	▨	▨	▨	▨				
7.7	Is the user aware of how much system memory his/her choices (in question above) will use?	▨	▨	▨	▨	▨				

Map/Task Questions

7.8 Does the map score highly (with our evaluators) on the 3 factor scores of aesthetics, potency and quality derived from the scale below?

Aesthetics	Potency	Quality	Comments

	1	2	3	4	5	6	
discernable							indiscernable
unfriendly							friendly
obtrusive							unobtrusive
dislikeable							likeable
unpleasant							pleasant
muted							bright
beautiful							ugly
tasteless							tasteful
rich							sparse
ordered							confused

	1	2	3	4	5	6	
loud							discreet
unacceptable							acceptable
discordant							harmonious
clear							muddled
usable							unusable
dull							exciting
subdued							vivid
exact							fuzzy
recommendable							inadvisable
aesthetic							unaesthetic

Map/Task Questions

- **Presentation/Layout**

		Yes/ Always	Mostly	Some- times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.9	Does the arrangement of components on the map (legend, scale bar, title etc.) seem logical?								▨	
7.10	Does the arrangement of components on the map avoid being disturburbing to the eye?		▨	▨	▨				▨	
7.11	Does the map layout acheive a visual balance and order?								▨	
7.12	Is there access to guidance on map design in on-line help or in the manual?	▨	▨	▨	▨	▨				

- **Supporting Information**

Accuracy of Information

7.13	Does the collected data match the requirements of the GIS user in order to produce a good map?	▨	▨	▨	▨	▨				
7.14	Can the user easily get some indication as to the accuracy and quality of the information portrayed? (No. of sampling points etc.)		▨	▨	▨					
7.15	Does the user understand how this might affect his/her task?	▨	▨	▨	▨	▨				
7.16	Is the correct number of statistical classes used?		▨	▨	▨				▨	Comment
7.17	Is uncertain data indicated appropriately (dashed line, etc.)?		▨	▨	▨				▨	
7.18	Have correct names of features been established?	▨	▨	▨	▨	▨				

Map/Task Questions

Scale and Orientation

		Yes/ Always	Mostly	Some- times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.19	Is the map (data) scale appropriate to the task? (Should be the smallest scale to allow legible presentation)		▨	▨	▨				▨	
7.20	Is the scale of the source data known by the user?	▨	▨	▨	▨	▨				
7.21	Is the accuracy of the source information correct within the scale limits?	▨	▨	▨	▨	▨				
7.22	Is there a scale bar?		▨	▨	▨				▨	
7.23	Is the scale bar a sensible size?		▨	▨	▨				▨	
7.24	Is it divided up appropriately?		▨	▨	▨				▨	
7.25	Is there an indication of orientation?		▨	▨	▨				▨	
7.26	Where the map is three dimensional, is the vertical exaggeration appropriate?		▨	▨	▨				▨	
7.27	For 3-D maps, is there a choice of vertical scales?		▨	▨	▨				▨	

Legend

7.28	Is a legend available?		▨	▨	▨				▨	
7.29	Is the legend accurate?								▨	
7.30	Are the symbol descriptions in the legend accurate?								▨	
7.31	Are the symbols in the legend the same as on the map?								▨	

Map/Task Questions

		Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.32	Can the user work without having to refer to the legend very frequently?		▨	▨	▨					
	• Map Projection									
7.33	Is the map projection in use appropriate to the task?		▨	▨	▨				▨	
	• Symbols									
7.34	Does the user understand the range of symbols and line types in use?	▨	▨	▨	▨	▨				
7.35	Is the level of contrast between symbols/lettering and their background sufficient to recognize all symbols correctly?								▨	
7.36	Do all symbols conform to well-established habits or population stereotypes? (i.e. should conform to locally familiar conventions) If they don't conform, why not?								▨	
7.37	Have the essential characteristics of a class of objects been captured (symbolisation) (i.e. Information content)								▨	
7.38	Have related symbols been used for related phenomena?								▨	
7.39	Have symbols which could mislead the user been avoided?		▨	▨	▨				▨	What?
7.40	Does the prominence of features reflect their significance to the task? (E.g. Are equally important symbols used for equally important features?)								▨	
7.41	Are map symbols a suitable size? E.g. Are symbols large enough to be easily legible? Are symbols small enough to avoid clutter? (Reason for poor readability of symbols may be poor resolution)								▨	

Map/Task Questions

		Yes/ Always	Mostly	Some- times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.42	If the screen is used at normal office viewing distance, are symbols at least 4mm high, or larger if sometimes embedded among text?								▨	
7.43	If the viewing distance may vary (e.g. when display screen may be viewed briefly from a few feet away), are symbols at least 8 mm high?								▨	
7.44	Are proportional symbols sensibly sized (where they are used)?								▨	
7.45	Are relative symbol sizes perceived as relating correctly to their real sizes?								▨	
7.46	Are too many of the features given emphasis (competing for visual attention)?		▨	▨	▨				▨	
7.47	Does the map avoid being cluttered? (Too high a density of information)								▨	
7.48	Is it possible to distinguish between continuous and discontinuous features?								▨	
7.49	If lines are used to outline filled areas, are they actually necessary?								▨	
7.50	Are lines used to outline filled area as fine as possible?								▨	
7.51	Are lines used on the map where area shading would have shown the information more clearly?								▨	
7.52	Do interrupted line-styles meet neatly at intersections?								▨	
7.53	Are overlapping sets of data apparent visually?		▨	▨	▨				▨	

Map/Task Questions

		Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.54	If a graphic display contains a high density of display elements, particularly overlapping and continuously updated features, does it segregate the information using a coding (i.e. represent information to maximise symbol differentiation) or sequencing (show information in segments or groups over time) technique?								▨	Which?
7.55	If the user needs to rapidly **extract information,** is double or redundant coding (i.e. redundant extra coding, e.g. to emphasise things that have changed) used to assist this?								▨	
7.56	Is double or redundant coding needed (where it is used)?		▨	▨	▨				▨	
7.57	If redundant coding is used, does it augment symbolic information, rather than add new symbols?								▨	
7.58	If coding is used to **discriminate** among different classes of simultaneously-displayed items, is it meaningful?								▨	
7.59	If coding is used to **discriminate** among different classes of simultaneously-displayed items, is it consistent across different display configurations?								▨	
7.60	If the user performs **search** and **identification** on displayed symbols, is shape coding used to distinguish between types?								▨	
7.61	If shape coding is used to **distinguish** between symbol types, are there less than 15 different shapes?		▨	▨	▨				▨	
7.62	If shape coding is used to **distinguish** between symbol types, are the shapes used compatible with and associated to the objects they're representing?								▨	

Map/Task Questions

		Yes/ Always	Mostly	Sometimes	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.63	If shape coding is used to **distinguish** between symbol types, are the shapes highly discriminable, avoiding confusion between two similar shapes?								▨	
7.64	If shape coding is used to **distinguish** between symbol types, are the shapes symmetrical?								▨	
7.65	If the user must make **absolute identification** of particular objects, is alphanumeric coding displayed?								▨	
7.66	If the user must make **absolute identification** of particular objects, and if speed and accuracy are essential, are numeric codes displayed?								▨	
7.67	Can the user hide alphanumeric coding where it reduces legibility?		▨	▨	▨				▨	
7.68	Can the user hide alphanumeric coding where it increases transmission/display drawing time?		▨	▨	▨				▨	
7.69	Can the user hide alphanumeric coding where it would draw attention to the wrong part of the screen, if it is important to avoid this distraction?		▨	▨	▨				▨	
7.70	If symbols are used which resemble an arrow or other indicator of direction, and these symbols may indicate movement or move themselves, are they always oriented to 'point' in the direction of movement?								▨	
7.71	Where **speed** and **accuracy of symbol identification** are important, are familiar alphanumeric symbols used rather than similar geometric ones?								▨	
7.72	Are letters used in preference to numbers where the numbers would have similar digits?								▨	

Map/Task Questions

- **Colour**

		Yes/ Always	Mostly	Some- times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.73	Is colour used?		▨	▨	▨				▨	
7.74	Where colour is used is it necessary?								▨	
7.75	Are the colours used aesthetically pleasing?		▨	▨	▨				▨	
7.76	Have familiar conventions for colours been followed?								▨	Which convention has been followed? Paper map/public or cultural convention/inhouse style
7.77	Is each major element of the map easily distinguished by its colour (where necessary) (e.g. roads, buildings)?								▨	
7.78	If colour coding is used, is it used solely to assist the user at the primary or first level of interest?		▨	▨	▨				▨	
7.79	If colour coding is used, is it used conservatively, with less than 11 colours?		▨	▨	▨				▨	
7.80	Where accuracy of **identification** is important, is the number of colours used kept at 7 or less?		▨	▨	▨				▨	

Background Colour

		Yes/ Always	Mostly	Some- times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.81	Is there an appropriate background colour (black/white)?		▨	▨	▨				▨	
7.82	Is white used as the background colour where the screen map has to match the paper map?		▨	▨	▨				▨	
7.83	Is black used as a background where the screen map does not need to match a paper map?		▨	▨	▨				▨	
7.84	Is a medium blue avoided as a background colour?		▨	▨	▨				▨	

Map/Task Questions

		Yes/ Always	Mostly	Some- times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.85	If the viewing distance may vary (e.g. when the display screen may be viewed briefly from a few feet away), is black used as the background colour, in preference to white or blue?		▨	▨	▨				▨	
7.86	Are foreground colours distinguishable from one another and background?								▨	
7.87	Where black is used as a background, are relatively luminous symbol colours used, e.g.green, orange or yellow?								▨	
7.88	Where white is used as a background, are green, red and light blue favoured as symbol colours?								▨	
7.89	Where both the foreground and background are coloured, are black occlusion zones used around all point symbols to maintain colour contrast?								▨	
7.90	If there is important information in the background, are black occlusion zones around point symbols avoided where they would obscure too much background?								▨	
	Strong Colours									
7.91	Are the strongest colours used for the principal information?								▨	
7.92	Are strong, fully saturated colours used for lines, lettering, point symbols and solids?								▨	
7.93	Where speed of **recognition** is important, are saturated colours used for colour-coding?								▨	
7.94	Are a maximum of 8-9 saturated colours used? (if there is no legend)		▨	▨	▨				▨	
7.95	Are red and green symbols larger than 15 arc minutes (2.25 mm at 50 cm)?									

Map/Task Questions

	Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.96 Are more than 8 tints of black avoided?		▨	▨	▨				▨	
7.97 Are more than 4/5 tints of dark colours (black, red, purple) avoided?		▨	▨	▨				▨	
7.98 If grey scales are used are they divided into what are perceived to be regular steps?								▨	

Light Colours

7.99 Are the 'lighter' (duller) colours used for the secondary information? (E.g. Visual hierarchy should not be inverted.)								▨	
7.100 Are colours which interfere with one another avoided?								▨	
7.101 If the user has to **search** for target objects, is colour coding used to facilitate this?		▨	▨	▨				▨	
7.102 If the user has to **search and identify** objects are the most effective colours used for the most important 'targets' (i.e. red is most effective, then blue, yellow, green, black and white)?								▨	
7.103 If the user has to perform rapid and precise **identification** of objects, is colour coding (alone) avoided?		▨	▨	▨				▨	
7.104 If a display is to be used for multiple tasks, or if the system designer could not have known what information would be of prime importance to the user, is the use of colour coding under user control?		▨	▨	▨					If 'yes', is it: Direct (on screen)/Indirect (in menu)
7.105 Where colour coding is used, does it avoid using both of any of these pairs of easily-confused colours: magenta/red, yellow/white, white/grey, aqua/blue, aqua/grey, aqua/green (only applicable to aqua/green when against a grey background)?		▨	▨	▨				▨	Which?

Map/Task Questions

		Yes/ Always	Mostly	Sometimes	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.106	If **searching** speed is important, are the following foreground/background colour combinations avoided: red or white against aqua, yellow on grey or blue or pink?		▨	▨	▨				▨	Which?
7.107	If searching speed is important and colour coding is used, are these recommended colour combinations used: red or white on grey, red on pink, yellow on aqua, white on blue?		▨	▨	▨				▨	Which?
7.108	If colour coding is used, does it avoid possible confusion by red-green colourblind users?								▨	
7.109	Is 'clean' blue avoided for text, narrow lines, small details [where this would make the features illegible]?								▨	
7.110	Are blue symbols larger than 20 arc minutes (3 mm at 50 cm distance)?									
7.111	Are more than 2 tints of light colours avoided (e.g. yellow, light green)?								▨	
7.112	For a univariate map (with data available over a continuous plane using colour coding to show value of one variable), if a colour key is being used to read metric quantities, does the colour sequence avoid varying monotonically with the colour opponent channels? (e.g. spectrum approximation supposedly good)		▨	▨	▨				▨	
7.113	For a univariate map, if it is important to reveal the shape or form of features such as parallel ridges, is the luminance varied monotonically rather than using a chromatic sequence? (e.g. spectrum approximation supposedly bad)		▨	▨	▨				▨	
7.114	For a univariate map, if it needs to reveal shape and metric values, does the colour sequence increase monotonically in luminance while cycling through a range of hues? (i.e. should have a sequence of colours which is repeated with increasing luminance, so quantity can be accurately identified on colour key but luminance shows shape of surface).		▨	▨	▨				▨	

Map/Task Questions

- **Patterns**

		Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.115	Are patterns used?		▨	▨	▨				▨	
7.116	Are patterns necessary (where they are used)?								▨	
7.117	Are patterns easily distinguished from one another?								▨	
7.118	Have conventions for patterns (e.g. geological) been followed as far as possible?								▨	
7.119	If 'yes' to the above, what are they?	▨	▨	▨	▨	▨			▨	Paper map/public or cultural convention/inhouse style
7.120	Have patterns which interfere with the legibility of other information been avoided?								▨	
7.121	Have line and area patterns which appear jagged on the screen been avoided?								▨	
7.122	Where line hatching is employed are horizontal lines used, where possible, in preference to vertical?								▨	
7.123	Has diagonal cross-hatching been used in preference to horizontal/vertical?								▨	
7.124	Do patterns give an impression of a continuous surface (even though the pattern is perceptible)?								▨	

- **Names**

7.125	Have conventions for lettering been followed as far as possible (e.g. no more than two font styles used, serif for physical features, san serif for cultural features, italic for water features, all capitals for oceans and lakes, capitals and lower case for rivers)?									

Map/Task Questions

		Yes/ Always	Mostly	Some-times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.126	Are the fonts chosen simple and easy to read (large x-height, short ascenders & descenders)?								▨	
7.127	Is lettering large enough to be readable on screen? (probably should be 9 pt. or larger)								▨	
7.128	If differences in type size are significant, are the size differences easily perceptible?								▨	
7.129	Has upper case text been used sparingly?		▨	▨	▨				▨	
7.130	Has bold type been used sparingly?		▨	▨	▨				▨	
7.131	Has wider than necessary letter/word spacing been avoided?								▨	
7.132	Is lettering orientated with the structure of the map [where appropriate]? (parallel with upper and lower edges - large scale; parallel with parallels - small scale).								▨	
7.133	Does lettering for linear features always follow the feature?								▨	
7.134	Has all angled lettering been set on a slight curve?								▨	
7.135	Does lettering avoid being unnecessarily curved?								▨	
7.136	Do names attached to point symbols follow these rules as far as possible: placed at top right if possible, if not - bottom right, top left, bottom left, directly above, then directly below?								▨	
7.137	Is text always placed as close to the relevant feature as possible?								▨	
7.138	Is it always clear to which feature a name refers?								▨	

Map/Task Questions

	Yes/ Always	Mostly	Some- times	Rarely	No/ Never	Don't know	Not applic.	Would ask user	Comments
7.139 Has the interruption of lettering by other features been avoided?		▨	▨	▨				▨	
7.140 Has inverted lettering been avoided?		▨	▨	▨				▨	
7.141 Do names avoid crossing coastlines?		▨	▨	▨				▨	
7.142 Do places on the shoreline have their names wholly in the water?								▨	
7.143 Have name duplications been avoided (e.g. 'Rio Grande River'?)		▨	▨	▨				▨	

- Generalisation

7.144 Is the level of generalisation of detail appropriate to the map scale and/or purpose?								▨	
7.145 If computer-assisted simplification is taking place, has the effect of projection been taken into account?	▨	▨	▨	▨	▨				
7.146 On special subject maps is the detail in the base map kept to a level appropriate to the level of detail of the specialised information?								▨	
7.147 Have the essential characteristics of a class of objects been captured?								▨	
7.148 Has the correct sequence for symbol displacement been employed? 1. Hydrographic lines (coast and drainage) 2. Contours and heights 3. Railways 4. Main roads 5. Minor roads 6. Buildings 7. Limits of vegetation, land use etc.	▨	▨	▨	▨	▨				
7.149 Do features shown by point symbols retain their correct relationship to topography? E.g. Water mill should remain adjacent to stream .	▨	▨	▨	▨	▨				

Appendix 3: Questions used in experiment

The tables below show the question sets used in the different experimental conditions (see also Chapter 6). Each set's 20 questions are shown in the order in which they were presented, which was randomised during preparation but kept the same for all subjects within each condition. The other columns show the left-hand and right-hand response options which were presented, and which one of those (left or right) was the correct one (note that the number of L and R correct responses is equal in all conditions).

The two 'r' (random) conditions used a randomly selected and mixed selection from the sets of 'g' and 'm' questions, for each village. Again L and R responses were balanced, and the same randomised order was used for all subjects within each condition.

Condition: gn (geometry, Newington)

Question	**L response**	**R response**	✓
On the left, does a white line cross the deep blue lines?	Yes	No	L
Where is there a pair of dark blue wavy lines?	Centre	Far left	R
What colour is used for most oblongs?	Red	Purple	L
What colour line crosses a red rectangle near top right?	White	Purple	L
How many white lines are there?	Three	One	L
Is there a yellow letter 'N' anywhere?	Yes	No	L
Can you move from a blue triangle to a deep blue line without crossing a line?	Yes	No	R
Is the green object near the bottom closer to a blue or red line?	Blue	Red	L
Which of these pink letters is crossed by a blue line near the bottom right?	w	g	R
What green shape is near the centre-bottom red one?	Tree	Cross	L
How many shapes are inside the central blue triangle?	Three	One	R
How many lines cross the red object near the central blue triangle?	Ten	Five	R
What's the highest number shown?	6	3	L
Which vertical white line crosses a purple line?	The left one	The right one	R
Which of these words is shown more than once?	Church	Home	L
Where is an upside-down red L shape shown?	Top right	Middle left	R
How many colours are there in total?	Four	Eight	R
What colour object is just above the bottom-right red one?	Blue	Yellow	L
Which lines tend to curve more, the red or the purple?	Red	Purple	R
What colour lines are next to the two small green circles?	White	Blue	R

Condition: gw (geometry, Saltwood)

Question	**L response**	**R response**	✓
What colour triangle's inside a bigger purple one?	Red	Blue	R
Does the blue triangle have any sharp corners, or are they all rounded?	Round	Sharp	R
How many colours are there?	Seven	Three	L
What word is crossed by the white line?	Saltwood	Newington	L
Do the red lines tend to curve more or less than the purple ones?	More	Less	R
How many tree names are shown?	Two	Four	L
Which of these two words is shown?	Church	Hotel	R
Where is a perfect small blue square shown?	At top right	In blue triangle	R
Which of these words appears twice?	Forge	Castle	L
What colour line is nearest the top left hand corner?	Yellow	Blue	R
How many words are shown in yellow?	Eight	Two	L
What two words are above the top-left small blue rectangle?	Two Firs	Oak House	L
Where's the largest red object?	Top centre	Bottom right	L
Can you move from the blue triangle to a red oblong and not cross a line?	No	Yes	L
What's the lowest pink number that's shown?	60.1	60.0	L
Does the straight white line cross any purple lines?	No	Yes	R
Moving from a green circle to the white line, do you cross any lines?	Yes	No	L
Is the white line longer than all the purple ones?	Yes	No	R
Do the two long, close, parallel bright blue lines meet any other lines?	No	Yes	R
What two words are below the central red rectangle?	Castle Hotel	Forge Cottage	R

Condition: mn (map, Newington)

Question	**L response**	**R response**	**✓**
Can you walk from the church to the Old Vicarage without crossing a road?	No	Yes	L
If you walk from the War Memorial to the church, is the yew tree on your right?	Yes	No	L
Which building exists behind number 6, The Street?	Church Cottages	The Old Vicarage	R
Is the Telephone Call Box (TCB) across the road from the Pound, or next to it?	Across the road	Next to it	L
What's the name of this village?	Saltwood	Newington	R
When the village had a school and a vicarage, which was nearer the church?	Vicarage	School	R
Home Farm has two barns with open fronts. Do they face the track or the stream?	Stream	Track	R
How many farms appear on the map?	Three	None	L
Through which farm does the stream flow?	Pound Farm	Home Farm	R
How many road names are given?	Four	Two	R
How many of Pound Farm Cottages are visible?	One	Three	L
Which farm exists between Pound Farm Lodge and The New House?	New Pound Farm	Home Farm	L
Does Newington Road run past the church?	Yes	No	L
Is The New House fully visible when standing beside the War Memorial?	Yes	No	R
Do any houses have a small front porch?	No	Yes	R
How many Church Cottages are there?	Two	Five	R
Are all the road names definitely shown, or could any be missing?	Some possibly missing	All definitely shown	L
What's at the southern end of the village?	Village church	Village hall	L
Will the track within Home Farm join Newington Road or The Street?	The Street	Newington Road	R
What's next door to The Old School?	The New House	Pound Farm Cottages	L

Condition: mw (map, Saltwood)

Question	L response	R response	✓
Can you walk from Fountain Stores to Forge House without crossing any roads?	No	Yes	L
What do you suppose the pink numbers indicate?	Height above sea level	Miles distant from London	L
How many road names are visible?	Three	Six	L
What faces the Village Hall from across The Green?	Castle Hotel	Oak House	L
What's in the middle of this village?	The village church	The village green	R
Which building lies due north of Forge Cottage?	The Homestead	Village Hall	R
What's the name of this village?	Saltwood	Newington	L
What's the name of the village shop?	Fountain Stores	Mountain Stores	L
Which of these two house names is shown?	Oak Cottages	Rose Cottages	R
Do any houses have a small porch at the front?	No	Yes	R
What has had two houses named after it?	The forge	The castle	L
How many houses are named after trees?	Two	Six	L
What's the main landmark situated on the green?	Village well	War memorial	R
Is the electricity substation in Grange Road or School Road?	Grange Road	School Road	R
Is the War Memorial visible when standing outside the Castle Hotel?	No	Yes	R
Which house exists between number 4 Grange Road, and The Green?	The Homestead	Forge Cottage	R
Which house is next door to Romney House in School Road?	Oak House	Two Firs	L
Does the Castle Hotel have a garden in front, or face directly onto the road?	On road	Has garden	R
Are all the road names definitely shown, or could any be missing?	All definitely shown	Some possibly missing	R
Can you walk from Oak House to the Castle Hotel without crossing any roads?	No	Yes	L

Comparison of question sets on readability statistics

Microsoft Word™ calculates a number of readability statistics for text, and these were used to double-check that there was no difference between conditions in linguistic difficulty. The Flesch Reading Ease statistic, based on the average number of syllables per word and words per sentence, is a score out of 100 and is higher the greater number of people who can readily understand the text. The Flesch-Kincaid, Coleman-Liau and Bormuth statistics are all based on different calculations (generally to do with number of characters or syllables per word, and words per sentence, combined in some way), and all suggest a grade of difficulty, so that the *lower* the number the easier it is to read the text (the Flesch-Kincaid supposedly equates to

American school grades). The statistics for the question sets as a whole are shown in the table below; it will be seen that in general the language used was graded as relatively simple in all four conditions.

Summary table across the 4 question sets

Set	Chars per word	Words per question	No of passive-voice questions	Flesch Reading Ease	Flesch-Kincaid Grade Level	Coleman-Liau Grade Level	Bormuth Grade Level
gn	4.2	9.8	5	91.1	3.0	6.1	8.2
gw	4.3	9.3	6	91.4	2.8	6.2	8.2
mn	4.2	9.6	2	89.9	3.1	6.0	8.2
mw	4.4	9.4	4	83.5	4.0	7.1	8.2

Results of statistical analyses

The above statistics were also calculated for each individual question in all four sets, and statistical tests (nonparametric, owing to normality deviations in some sets) were applied to check there was no difference between (a) question type (map versus geometry) or (b) village (Saltwood versus Newington). For all the above scores except number of passive-voice questions, Mann-Whitney U tests were used for map/geometry and Saltwood/Newington comparisons; for passive-voice, chi-square crosstabulations were calculated (since this data is nominal yes/no for individual questions). The results were all insignificant, so the conditions were assumed to be matched for linguistic difficulty of questions.

Photograph questions: Saltwood

Question	L response	R response	✓
What's behind you in photo D?	El Sub Sta	Forge Cottage	L
What's behind you in photo A?	Romney House	Fountain Stores	R
What's behind you in photo C?	The Green	Forge House	L
What's behind you in photo B?	Castle Hotel	Grange Road	L

Photograph questions: Newington

Question	L response	R response	✓
What's behind you in photo D?	Home Farm	Pound Farm	L
What's behind you in photo A?	Pound Farm Cottages	St Nicholas's Church	R
What's behind you in photo C?	TCB	The Old School	L
What's behind you in photo B?	Church Cottages	The Old Vicarage	L

Appendix 4: full experimental procedure

As stated in Chapter 6, the experimental procedure consisted of ten parts, which are detailed below. All subjects had been previously told that the experiment would last for well under an hour, that the only physical requirements were good enough eyesight to see small details on a screen, and no colourblindness, and that they would simply be responding to things on a computer screen by hitting one of two buttons. They had also been told that they would be paid £5 (some, being personal contacts of the experimenter, declined this), and that a variety of ages and backgrounds was sought with special interest in people who had experience of using visual information.

Completion of demographic/expertise questions (on paper).

Subjects were welcomed (if appropriate; the experiment was performed in various places but always in a quiet rom with minimal distractions). They were asked to start by filling in both sides of the expertise questionnaire (see Appendix 3), while the experimenter set up the laptop (if necessary).

Brief verbal instructions from experimenter.

The experimenter then showed the subject how to position the laptop and adjust its screen if desired, checking the brightness and contrast settings were appropriate. She then explained that they would simply follow through the instructions on the computer screen, that if they had coffee they should drink it only during the breaks which they'd be allowed to have (so that it didn't affect response timings), that the computer would ask them to hit one of two buttons to answer questions that came up on the screen, and that they should answer as quickly and accurately as possible. The card with the photographs on it (either Saltwood or Newington) was placed face-down on the table to one side, and the subject was told that its use would be explained at a later point. The experimenter then sat down, behind and slightly to the side of the subject, with a clipboard on her lap, and explained that she would be keeping an eye on the screen to check for progress and problems, but wouldn't be able to see the subject's responses.

Working through on-screen instructions, training Ss to make responses.

The initial set of instructions to the subject were already present on the screen when the laptop was turned towards them:

```
This experiment is very easy: you don't need to know anything
about the computer, because all you have to do is hit one of
two buttons on the keyboard each time you're asked a question.

Just follow the instructions on the screen, and when you're asked
a question please answer as quickly and accurately as you can.
If you're stuck or you take too long, after a while the computer
will just go on to the next question.

When you're ready to start, please press any letter on the
keyboard...
```

After pressing a key, this was displayed:

```
To make this work OK, please use either your index or middle
finger to hit each button, and try to keep your other fingers out
of the way (e.g. curled up under your hand) so they don't hit any
buttons by mistake.

Please place your left index or middle finger over the letter X on
the keyboard (it's near the bottom left), with your other fingers
and thumb tucked out of the way. When you're comfortable, please
press the X button.
```

Once the subject had pressed the X (NB at this stage and during the practice session, no other key was allowed to substitute for X or M, although later in the experiment any surrounding key was treated the same way as X/M so that finger slips made no difference to results):

```
OK, now please do the same with your right hand. Use the same
finger that you did for the left, and place it over the letter
M, tucking your other fingers and thumb out of the way. When
you're ready, please press the M button.
```

On pressing the M, this appeared for 3 seconds before clearing the screen and continuing:

```
OK, keep your fingers on the keys - here we go...
```

Practice session.

```
OK, next you'll be shown some images on the screen, asked a
question about each one, and shown two possible answers. Hit your
left (X) or right (M) button to say which answer is correct. Please
answer as quickly and accurately as you can.

This bit is only a practice session, to get you used to things, so
if you have any problems with it we'll just run through it again.

When you're ready to carry on, put your two fingers on the left and
right buttons (X and M), and press one of them.
```

Then (for 3 seconds again as before):

```
OK, keep your fingers on the keys - here we go...
```

The practice session was then run, with its five images shown twice each (one question of each pair being about the visual appearance of the image, and the other about whatever it represented; the questions were randomly mixed but in such a way that each pair was separated by at least one question about a different image). For this session and all remaining sessions of the experiment, the images were centrally placed on the screen (if they didn't fill it), and the question always appeared at exactly the same point near the bottom, with its two possible responses displayed underneath (again at exactly the same points each time, chosen to roughly align with the X and M keys).

Provided that more than 7 correct responses were made (which in the event was achieved by every subject), the screen then cleared for this message:

```
OK, you've got the hang of it.

When you're ready to continue, press either key...
```

Reversible figures session.

```
For the next few images you see, you'll be asked two questions
about each one: first you'll be asked to say what you see in the
image, and then you'll be asked if, and how soon, you can spot
something different in it.

As before, you just use the same two keys to answer the questions,
and please answer as quickly and accurately as you can. From now
on your answers will be timed and recorded by the computer.

When you're ready to carry on, put your two fingers on the left and
right buttons (X and M), and press one of them.
```

The 'here we go' message was shown for 3 secs as before, and then the reversible figures task began. This time, after subjects responded to the first question about each image, the image remained and only the question and answers at the bottom changed. (One side-benefit of this was in giving Ss extra practice at the two-choice response task, and at answering a series of questions about one image).

For each image, the same two questions were asked but substituting the appropriate objects in each case, for example:

```
What do you see first in this drawing?
Duck                          Rabbit
```

followed by:

```
Look again: can you see a duck?
Yes                           No
```

(or rabbit, if duck had been the first response)

First main-task session (either M, G or R questions, and either Saltwood or Newington).

```
The next set of questions will all be about the same image, which
will stay on the screen the whole time. You answer in the same way
as before, hitting either the left (X) or right (M) buttons. Once
again, please answer as quickly and accurately as you can.

When you're ready to go on, put your two fingers back on the X and
M keys, and press one of them.
```

The 3-second 'here we go' message followed, then the screen cleared, the first map (either Saltwood or Newington) was displayed, and subjects answered the first series of questions depending on the condition to which they had been (randomly) assigned (the questions are listed in Appendix 1).

Pause and opportunity for short break if required.

```
Right, now if you want you can rest for a while before doing some
more questions. These will also all be about one image, but it's a
different one. Don't start this set until you feel ready, but then
as before please answer as quickly and accurately as you can.

When you're ready to carry on, put your two fingers on the X and
M buttons, and press one of them.
```

(Then the 3-second 'here we go' message again, before the second session started.)

Second main-task session

(As the first session but G if previously M and vice versa, still R if previously R; Saltwood if previously Newington and vice versa).

Photograph task (with reappearance of second map).

```
OK, next you'll get just a few more questions, but this time you'll
be looking at some photos as well as the map you just saw. They show
the same village, from different positions.

In each question you'll be told which photo to look at, and asked to
imagine that you're the person taking the photo. You have to say
what's BEHIND you. As before, you choose one of two answers. Once
again, please answer as quickly and accurately as you can.

When you're ready:
-  turn over the big card, but DON'T look at the photos just yet;
-  place it beside the computer where you'll see it clearly;
-  put your two fingers back on the X and M keys;
-  press one of them to carry on.
```

Then the 3-second 'here we go' message. The photograph questions were identical other than the letter identifying the photo, and the two possible answers, e.g.:

```
What's behind you in photo A?
Romney House                  Fountain Stores
```

After the fourth photo, the screen cleared and this message appeared:

```
OK, now you're finished with the computer. Thanks for your help!
```

Debriefing and brief post-experiment interview.

At this point the experimenter rose and took the laptop to one side, explaining that she wanted to just ask a few questions about "how you found it". The questions were asked informally and the experimenter typed notes on subjects' answers directly into their results file on the laptop. The questions asked:

- Did the subject have any problems with seeing or responding to the questions at any time? Specifically, no problems with seeing colours or details, or with coordinating their responses and matching them to the answers on the screen?

- The reversible figures: had the subject managed to see both pictures in all the images? Had they seen any of them before?

- The main task: had the subject realised that there were two different types of question being asked about the maps? (If not, and most hadn't, the experiment explained about the 'map' versus 'geometry' distinction.) Which type had they found harder to respond to, or had they all seemed about the same?

- How had they found the photograph questions?

- Had there been any question at any point that they'd been unable to answer or had had to guess?

- What did they think was the purpose of the experiment?

The experimenter then debriefed the subject about the actual purpose of the experiment (most subjects' suggestions were inaccurate, and often assumed some kind of evaluation of the map designs). The subject was then paid (if appropriate) and allowed to leave.

Appendix 5: Demographic/expertise questionnaire

The questionnaire follows on the next page; see Chapters 6 and 7 for a discussion of the questions and of which/how data was used in the analysis.

Name:..

Age....................... Gender (M/F)............................

Occupation (please give details of subject, specialism, etc.).....................................

....

1. How much education or training, or extensive informal learning, have you done in the subjects listed below?

   Please tick the boxes as appropriate. If any of it has been within the past year, please tick the circle on the right as well.

	Learned without formal qualification	School-level qualification	Post-school qualification	Some learning in past year?
Computing	☐	☐	☐	○
Technical drawing or CAD	☐	☐	☐	○
Geometry	☐	☐	☐	○
Photography	☐	☐	☐	○
Art	☐	☐	☐	○
Graphic/visual/fashion design	☐	☐	☐	○
Architecture or planning	☐	☐	☐	○
History of any visual arts	☐	☐	☐	○
Geography	☐	☐	☐	○
Cartography/GIS	☐	☐	☐	○
Psychology	☐	☐	☐	○

2. When you're creating or using information, do you prefer to work 'verbally' (words, descriptions, reading) or 'visually' (drawings, diagrams), or neither?

   ☐ Verbal ☐ Visual ☐ Neither

3. When you're going somewhere unfamiliar, or helping someone else go somewhere, do you prefer a list of verbal/written directions, or a map/sketch?

   ☐ Directions ☐ Maps ☐ Neither

PLEASE TURN OVER

4. For each of the following things, please think about how much experience you've had of using or looking at them, and circle a number between 0 and 5. Write any comments or clarifications on the right if you want to.

	0 = never looked at					5 = extensive experience
Flowcharts, family trees, similar diagrams	0	1	2	3	4	5
Blueprints/designs/diagrams of objects/machines	0	1	2	3	4	5
Medical (or vetinary) images/diagrams	0	1	2	3	4	5
2D computer graphics or games	0	1	2	3	4	5
3D virtual reality environments (e.g. in games)	0	1	2	3	4	5
Graphic designs (e.g. for publicity materials)	0	1	2	3	4	5
Computer-aided design (e.g. engineering)	0	1	2	3	4	5
Layout plans for an office, garden, kitchen, etc.	0	1	2	3	4	5
Layout plans of whole buildings or properties	0	1	2	3	4	5
Street maps (e.g. London A-Z)	0	1	2	3	4	5
Railway/Underground/metro route maps	0	1	2	3	4	5
Other network maps (e.g. pipelines, cables)	0	1	2	3	4	5
Ordnance Survey-style maps, or road atlases	0	1	2	3	4	5
More detailed countryside maps (e.g. for hiking)	0	1	2	3	4	5
Digital maps, geographic information systems	0	1	2	3	4	5
Route planning/navigation systems (computer)	0	1	2	3	4	5
Drawings or paintings of urban or rural scenery	0	1	2	3	4	5
Photographs of urban or rural scenery	0	1	2	3	4	5
Aerial photographs (taken from planes etc.)	0	1	2	3	4	5
Abstract drawings, paintings or other images	0	1	2	3	4	5
Any visual images that *you've* created yourself	0	1	2	3	4	5
Doodles (yours or other people's)	0	1	2	3	4	5
Satellite images or weather maps	0	1	2	3	4	5
Unfamiliar countryside (travelling/visiting)	0	1	2	3	4	5
Unfamiliar towns/villages (travelling/visiting)	0	1	2	3	4	5
Unfamiliar large buildings (travelling/visiting)	0	1	2	3	4	5
Unusual objects/images	0	1	2	3	4	5
Photographs of unfamiliar places	0	1	2	3	4	5
Any other kind of visual image (write below)	0	1	2	3	4	5

Appendix 6: Practice stimuli and questions

The order of the images and questions was pseudo-randomised, such that the two questions about each image were separated by at least one other question about a different image. For each image, one question focuses on the visual appearance, avoiding interpreting its semantic content, while the other focuses on the image's meaning. Thus, without realising it, subjects were being acclimatised to the two types of task, while avoiding too obvious a focus on maps (which could bias the results).

The image files have since been lost, but were:

campsite.bmp - a simple coloured drawing of a few tents grouped around trees and a river.

animals.bmp - a cartoon of a pig, a rabbit and a squirrel (all crudely coloured and not drawn particularly to scale)

ferrymap.bmp - a portion of a map of western Scotland, showing some ferry routes to the Western Isles as blue lines.

kent.bmp - a portion of a similar map of part of Kent, showing the M25/M26/M20 road junctions and a few towns

upstairs.bmp - a crude 'blueprint' of the upstairs floor of a house (an aerial view)

Image file	**Question**	**L response**	**R response**	✓
campsite.bmp	Is the blue line down the middle straight?	No	Yes	L
animals.bmp	Which animal is the squirrel looking at?	Pig	Rabbit	L
ferrymap.bmp	Which village has a ferry to Oban?	Benderloch	Achnacroish	R
kent.bmp	How many little squares are shown inside the dark blue shape?	Four	Two	R
upstairs.bmp	What colour are the lines?	Pink	Blue	R
ferrymap.bmp	How many dark blue lines are there?	Seven	Three	L
campsite.bmp	Are there more trees on the campsite side of the river?	Yes	No	L
animals.bmp	What colour is the smallest object?	Brown	Pink	L
upstairs.bmp	Does this house have a bathroom?	Yes	No	L
kent.bmp	Which town is east of Westerham?	Biggin Hill	Sevenoaks	R