Schematic Maps in the Laboratory

Maxwell. J. Roberts
Department of Psychology
University of Essex
Colchester, UK
mjr@essex.ac.uk

***Abstract*—Empirical research into the usability of schematic maps is currently in its infancy. This paper justifies the necessity for research, reviews the most relevant recently-published studies, gives an overview of potential pitfalls of this topic, and suggests future directions.**

***Keywords*—schematic mapping; journey planning; empirical work; evidence-based design**

I. Why we need Empirical Research into Schematic map usability

Travel by public transport in almost any city in the world, and the chances are that sooner or later you will be confronted with a highly stylised map of the routes taken by trains, buses, or trams. Straight lines will dominate, usually horizontal, vertical, and 45-degree diagonals, and there may be considerable topographical distortion [12]. The intention is that such *schematic maps* will be easier to use than topographically correct designs, with their disorganised undisciplined lines meandering across the page. However, the evidence for their superior usability is generally unconvincing, (see Roberts *et al.*, [16] for a discussion). Indeed, many users object strongly to such depictions, arguing that the topographical inaccuracy is, at best, disorientating, and may even be misleading. Worse still, although there are many pronouncements and beliefs about the criteria necessary to optimise schematic maps, empirical research to test their validity is almost completely absent. Against this backdrop, it is therefore urgent to identify what research is needed, and why, in order to determine the principles of best practice. The alternative is that transport undertakings will run the risk of publishing expensive mistakes that are rejected by the public, such as the controversial Madrid Metro map of 2007.

Fig. 1. The *linearity* of a map refers to the number of angles permitted in a design. The basis of this is the number of directions of travel at any given point. Hence, the traditional *octolinear* layout, as exemplified by the London Underground diagram, has *four* angles (horizontal, vertical, and 45˚ diagonals, giving *eight* directions. A *hexalinear* map has three angles spaced at 60˚ intervals, giving six directions. Any angle is permitted for a *multilinear* design, and a *curvilinear* design has no straight lines, only Bézier curves.

A. Preference is not Performance

Unfortunately, in evaluating schematic map design, many researchers simply ask members of the general public which versions they prefer, what they like and dislike about them, and what additional features they would find desirable. Psychology research has a long history of demonstrating that canvassing people's opinions in this way is a very poor method for identifying the best designs and establishing good-practice. In general, people have very little insight into their own performance [4, 8] and, for any task where they lack expertise and are effectively novices, they tend to base their judgements on the identification of superficial surface features rather than a deep understanding of the concepts involved [5]. For schematic maps, this translates into judgements based upon expectations and prejudices about design (e.g., octolinearity – see Figure 1 – is the best format for a schematic map). Even in

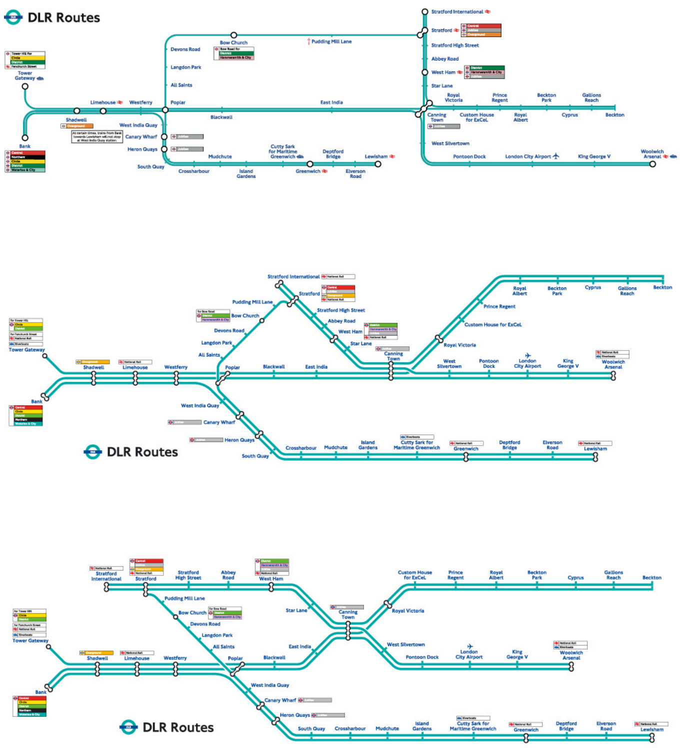


Fig. 2. Three Docklands Light Railway maps subjected to usability testing. The top map was preferred, but was associated with the most journey planning errors. The centre map resulted in inefficient journeys: those from the Lewisham to Woolwich/Beckton branches tending to be roundabout via Stratford rather than the direct route via Blackwall. The bottom map was least popular but had neither of these problems.

the absence of prejudices, aesthetic judgements may *substitute* for any assessment of underlying criteria for good usability. This dissociation between *map effectiveness* and *map engagement* has been strikingly demonstrated in work by Roberts *et al.* [16], in which a novel curvilinear design was shown to be superior to the official octolinear schematic in terms of the time necessary to plan complex journeys (see Figure 5). The objective measures of planning performance were, however, not correlated in any way with subjective ratings of map usability (obtained via questionnaires) or map choices. Retrospective reports by many subjects indicated that even when they had a suspicion that the curvilinear version was easier to use, it was still rejected because it was believed to be an inappropriate or unfamiliar way to design a map.

This dissociation between subjective versus objective measures is present even for non-controversial designs. In a study commissioned by the Docklands Light Railway to evaluate car line diagrams, the most popular prototype was associated with the most journey planning errors, the 2nd/3rd most popular designs resulted in inefficient journeys, and the least popular designs were associated with the fewest errors and the most efficient journeys [14] (see Figure 2). The overall message from these findings is clear. People can dislike designs that are easy to use, and prefer designs that are difficult to use. Basing a theory of effective design on votes and opinion polls is, therefore, a dangerous pastime.

B. *Evidence-Based Design*

One of the most comprehensive set of criteria and guidelines for effective design has been assembled by Ovenden as a result of canvassing various graphic designers [13, page 151]. These range from very general and wide-ranging (e.g. straighten line trajectories) to very subtle criteria which would be unlikely to impinge on usability in any measurable way – such as *do not change directions twice between two stations* – although the subtle criteria might well have an effect on the overall impression of neatness that a map gives. The problem with all the listed criteria though is that none of the principles has any cited evidence supporting its contribution to usability.

A similar problem can be identified in more scientifically directed writing. Researchers who devise automated map layout algorithms have devised usability criteria partly out of necessity, in order to constrain the space of possibilities and enable their software to generate solutions. These have been developed over the years by a number of researchers in the field, with a particularly clear exposition by Nöllenburg and Wolff [10]. These include requirements for: octolinearity; straightened line trajectories; preservation of relative spatial positions of nearby stations; equalizing inter-station distances (with the consequence of enlarging the centre of the map if stations are more closely spaced in this region); and orderly station name placement. All of these sound reasonable, and yet many published official maps can be found that break one or more of these criteria. At the very least, if empirical evidence could be collected to demonstrate that obeying the principles would always yield a more effective design than disobeying them, then this would provide a more powerful basis – than the mere assertions and whims of academic researchers – for rejecting offending maps.

Of course, it is possible that at least some of these principles can be neglected, and yet an effective design can still be created. Even so, it is hard to imagine that a map with

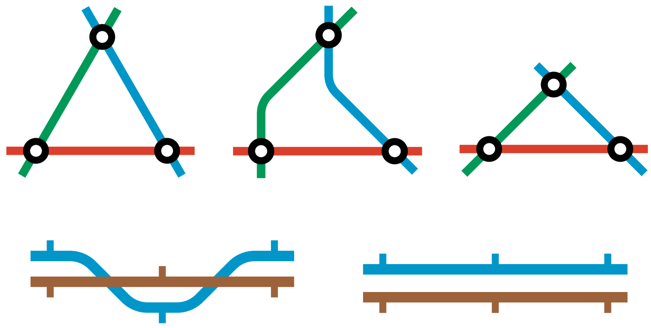


Fig. 3. Upper: If the reality is three lines intersecting at 60˚ to each other (left), it is hard to see what is gained by forcing the lines to a 45˚ grid, degrading the line trajectories (centre) or distorting topography (right). Lower: Conversely, it is not always essential to preserve relative positions of stations (left), if this might simplify line trajectories (right). A distorted layout need not be misleading if the schematic map omits all street details, and local signage is good. Unless the precise bearing of a station exit is known, it will not be possible to locate a nearby station even on a schematic map that has exceptional topographical fidelity. A compass will be always required, or a separate street map.

endless zigzagging line trajectories could be easier to use than one with simple straight ones, but part of the problem is that map design is a complex activity, and many of the criteria for effective design can conflict with each other. For example, straighter line trajectories might be possible for a map with hexalinear angles rather than octolinear ones (Figures 1 and 3), or else a network might be so complicated that no linear design can yield simple straight trajectories, so that a curvilinear version might be more appropriate instead; both of these examples indicate that different design rules suit different networks. Lines might also be straightened by relaxing the *relative spatial layout preserved* criterion for station placement (see Roberts [15] for a discussion of all of these examples, and many others). Empirical evidence can help, therefore, not just by confirming that the criteria for effective design are valid, but also by indicating whether any of the criteria should be given a higher priority than others where trade-offs might be required.

II. EMPIRICAL RESEARCH STRATEGIES

There are numerous methods that can be used to determine the most usable map from a set of options, and perhaps, from this, form the basis of the creation of a general theory of effective design. Even if we pick one single straightforward task, such as *station finding*, there are many different ways in which we can assess how easily visual information can be extracted from a map. Add to this that we can study different networks, real and imaginary, and different people, and the range of options multiplies.

A. *Tasks and Methodologies*

If we restrict ourselves to planning a journey, there are several sub-components that can be investigated. For example, first it is necessary to identify two stations on the map (start and end), then it is necessary to identify possible routes between them, and then (if necessary) choose between various options. For analysing data, it is preferable to disconnect the station finding and route identification stages. Both components are relatively difficult, and subject to high levels of variability – both within people (one station is found quickly by luck in a few seconds, another takes up to a minute) and between people (one person, experienced with maps and public transport, takes 10 to 20 seconds to plan each journey, another person, less able, takes one to two minutes). High data variability means that large numbers of volunteers must be tested in order to have sufficient *statistical power* to identify differences between designs, and combining sequences of complex tasks – each subject to high variability by themselves – compounds this problem.

Station finding (e.g. *point to the following station on the map: Oxford Circus*) is straightforward to test, but still can be implemented in many ways, for example by a simple paper map and stopwatch timing, or else automatically with a computer-presented sequence of trials and, either responses made using a mouse, or a touch-screen (perhaps the latter is closest to real life and is a more natural method of responding). Mean search time for identifying stations is the easiest measure of performance to analyse. A more ambitious project would use gaze-tracking methodology as well to see whether some map designs result in a more systematic and orderly search than others.

Journey planning is a richer task and can be tested in at least two different ways. One method is to provide people with pairs of highlighted stations, and simply ask them to identify a route between the two. The time taken to plan each journey can be recorded, and the directness/likely duration of the journey itself can be analysed. For this format, complex journeys must be planned in order to maximise performance differences due to map usability differences. In practice this means that the researcher should select journeys that require two changes of trains in order that an efficient, direct journey can be planned.

Responses to route identification tasks are harder to collect if computer-implementation is preferred, and an alternative method for measuring the speed and accuracy of this is to highlight pairs of stations, each pair on different lines, and ask people to click on (mouse) or point to (touch screen) the interchange station necessary to change between the two in order to complete the journey. This gives a performance measure in terms of mean time to complete the task and, for a complex map, numbers of errors might also be a useful measure. Gaze tracking might again supply additional information about the search process, and another interesting variant is to persuade people to plan quickly by giving them time deadlines [16].

Traditionally, when presenting a task on a computer, psychologists try to devise these such that simple responses are necessary, such as *yes-no* decisions. These can be implemented for maps, but at the expense of making tasks less realistic. For example, a person could be asked *does a journey from Brixton to Morden pass through Balham? Yes/No*. Another possibility is to count stations en-route. Both methods provide two measures of performance: response times and numbers of errors, but new technology such as touch screens reduces the need for the more contrived tasks. Other possibilities include asking people to draw elements of the networks from memory after the planning tasks in order to see whether any designs promote better learning than others. Finally, seeking subjective evaluations via rating questions and preference tasks ensures that *map engagement* as well as *map usability* is measured.

For all these tasks, an important consideration is whether they should be run *between-subjects* or *within-subjects.* For a between-subjects design, each person experiences just one single map, different groups of people are given different versions. The disadvantage of this format is the lower statistical power, requiring more people to be tested in order to identify differences between maps. However, only planning using one single map guarantees that the experience of using one map will not influence performance when using another. Within-subjects designs are more dangerous in this respect. Here, every person experiences every map (with different but equivalently-difficult sets of journeys assigned to each version). However, the advantages are greater statistical power, and clearer evidence for each individual as to whether a particular map was personally easier or more difficult to use.

B. Individual Difference Variables

It is important to understand and select carefully not just the tasks that are administered, but also the people who will be taking them. Many individual difference variables are likely to affect journey planning performance, such as general cognitive abilities (e.g. intelligence and spatial ability), both indicating a persons' ability to cope with a complex task that requires a considerable quantity of information to be assimilated. Expertise is another important consideration. People will differ not just in their experience of using a particular network, but also their experience at using public transport in general. Expertise will obviously affect the speed of planning a journey although, as expertise with a network increases, so the importance of using a map (and indeed, perhaps the importance of the design of a map) for planning a journey decreases. It could therefore be argued that the aim of a map is to provide assistance to people who are not experts of a city (e.g. tourists), and so all people who are experienced at using public transport in a city whose maps are being tested are excluded. Similarly, a researcher who wishes to identify *general* principles of effective design might wish to minimise *specific* network effects by, for example, devising a complex fictitious network. However, this methodology loses an important source of information: whether a novel, unusual design might put people who are used to conventional ones in difficulty. A more powerful way of controlling for specific network expertise would be to conduct an international study, for example by testing London and Berlin designs comparing residents of London and Berlin.

It is difficult to identify differences between males and females in their ability to plan using schematic maps. Age differences influence not just planning ability, but also map choice. For example, surprisingly, older people preferred the unconventional curvilinear Paris Metro map in the study by Roberts *et al.* [16]. Understanding reasons for individual differences in map engagement are important: insights into why people might reject a perfectly usable version could mean that designers are better able to accommodate their preferences in future designs. However, research into personality and aesthetic preference so far have given few insights, other than the suggestion that people who have strong preferences for maps in a certain style might prefer all maps for all cities to be in that same style [9, 18]. Overall though, understanding an individual's response to a map is clearly an important element of understanding map usability and map engagement.

III. Results of Research Into Schematic Map Design

Very few studies have directly investigated schematic map usability. The older ones that have generally been less than informative. For example, Bronzaft and colleagues [3] found that every single one of twenty novice participants made at least one error when using the then-current New York Subway schematic map to plan a series of journeys. However, after failing to find any clear improvements for new topographical designs, the researchers switched to gathering user ratings instead [2]. In contrast, Bartram [1] found faster planning for a schematic bus map, but the topographically accurate version had considerably more detail, i.e. streets as well as bus routes. The disadvantage for the topographical map might have been owing to the considerable quantity of potentially distracting supplementary information – irrelevant to the set tasks – rather than complex route trajectories. Overall, older research is lacking in both quantity and sophistication, so that the high rate of adoption of schematic maps during the 1970s and 1980s can scarcely said to be evidence-based. More recent research is more promising in this respect.

A. Perceiving a Map

When a person first encounters any map, the crucial first few seconds will convey an impression as to whether this is likely to be of assistance, or even threatening. Quite apart from aesthetic considerations, visual properties of designs can affect many aspects of engagement and usability. A number of researchers in psychology have investigated topics that, in theory might have some sort of bearing on people's reactions to the appearance of the map. For example, *visual stress* (where the qualities of certain images can be shown to result in adverse responses neurologically) may be an important consideration, although, so far, it is not obvious whether findings have implications for schematic map design (e.g., [6, 20]).

Of particular relevance here is work by Rosenholtz and colleagues [17]. This research team has developed a computer analysis technique to create so-called *mongrels* from visual scenes. These are intended, given a chosen visual focus, to assess the integrity of the peripheral vision around the focus. Hence, for a user who is focusing on, say, Oxford Circus on the London Underground map, how clear are the adjacent areas in the periphery of vision. A number of different official and unofficial network maps have been analysed, and it does appear possible to identify designs which are more-or-less clear in this respect, but such analyses have not yet gone beyond the *intriguing findings* stage.

The important next step for such research is to identify whether behavioural measures are related in any way to the analysis of perceptual qualities. For example, when people plan a journey, they generally focus on different locations of a map, and so the importance of peripheral vision to this task is not clear. Given that this analysis can be used to identify unclear parts of a map, the following possibilities are all plausible, but evidence in their support is important before this methodology is used to condone or condemn individual designs.

- In a station-location task (*find station x and point to it on the map*), are stations that are located in clear regions of the map easier to find compared with unclear regions?

Fig. 4. A 'mongrel' prepared by Rozenholtz and colleagues for the London rail map (all railways in London, in addition to the Underground network). Centred on Oxford Circus, the integrity of the east side of the periphery appears particularly poor.

- When planning a journey, do people tend to prefer travelling through regions of the map that are clear rather than unclear?
- In general, when people plan journeys, are maps easier to use if they have substantial clear regions rather than unclear regions?
- When tracking gaze direction, is it easier for people to shift their gaze from the centre of the map to peripheral regions that are clear versus unclear?

My own suspicion is that the qualities of maps identified by the researchers will impinge more directly on map engagement rather than map usability, so that they affect aesthetic judgement, immediate emotional reaction, and overall general impressions, positively or negatively, leading to a map's acceptance or rejection. This is obviously an important aspect of design, but to go further than this requires more evidence, if it is forthcoming then this will prove to be a very useful means of identifying and weeding out poor designs.

B. Using a Map

There is no substitute for real usability data, comparing maps designed in different ways in order to identify versions that are more-or-less usable than others. Roberts and colleagues investigated journey planning using the official octolinear Paris Metro map versus a novel curvilinear design. The underlying intention of this was that the complexity of the Paris network was resulting in the octolinear schematic failing to offer any benefits as regards simplification, so that the complex meandering trajectories of reality were merely being converted into complex zig-zagging straight lines instead. An alternative valid design approach, therefore, might be to smooth away the corners to create gentle curves instead, with the added intention of making the underlying network structure more salient. The research was a between-subjects design, with individuals given pairs of stations for complex cross-Paris journeys, and planning these on laminated maps. The curvilinear design was around 30% to 50% faster for journey planning across three studies, with no differences in *journey quality* overall. [Journey quality was estimated using a simple station and interchange counting

Fig. 5. Sections from Paris Metro maps investigated by Roberts and colleagues [16]. The upper curvilinear map is up to 50% faster for journey planning than the lower official octolinear map. However, note the trajectory of Line 10. On the curvilinear map, the smoother routing makes the line more direct and enticing (the trajectory on the octolinear map exaggerates reality, and is almost as misleading from a topographical point of view). Data suggest that for certain journeys, more people chose line 10 for the curvilinear map than the octolinear map. Line 10 has many station stops, which can result in a slower journey, but Line 10 is one of the least busy in Paris, so the journey experience would have been more pleasant.

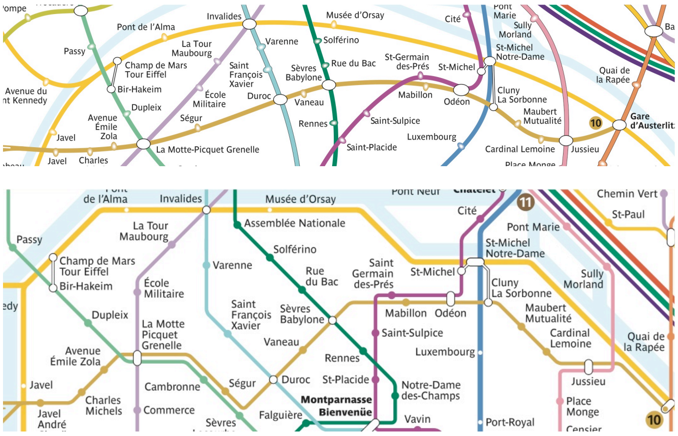
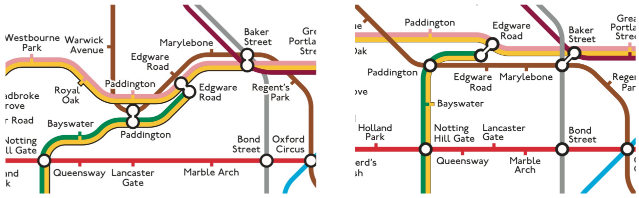


Fig. 6. Two octolinear schematic maps of London, one optimised for simple line trajectories, the other is topographically correct. Note that for the optimised map, the two alternative options from Paddington to Bond Street appear equally satisfactory. The topographically correct map shows that one option is considerably less direct.



heuristic]. In one task, a deadline procedure was applied, limiting the amount of time available for planning. Interestingly, people were generally able to cope with this, but with no measurable decline in journey quality, suggesting that, in the other studies, the planning times were inflated, perhaps because of prevarication between alternative journey options with little to choose between them in terms of time. Overall, the authors acknowledge the limitations of what can be concluded from the findings (see later) but one strong valid conclusion can definitely be drawn, the hypothesis: *octolinearity is a gold standard, maps designed in this way will always be superior to equivalent maps created using other design rules* is refuted.

Map design does not only affect usability, it also affects people's journey choices. Roberts *et al.* [16] sometimes found differences in option preference, between maps, on a journey by journey basis. These appeared either to be related to different maps implying that different lines were taking more-or-less direct routes (see Figure 5), or else because certain complex regions of the map gave the impression of being difficult to navigate and so were avoided. Guo [7] has demonstrated journey choice effects in real life, with ticketing data showing that 30% of journeys between Paddington and Bond Street were inefficient, presumably because of the misleading official map (see Figure 6). Guo seems to interpret this finding as evidence that schematic maps with topographical distortion should not be created. However, Roberts and colleagues [15, 16] argue that this response is too extreme. Even for this specially chosen journey: most people took the more efficient option; the time penalty was small – probably around seven minutes; the fare was identical; and splitting people between lines could alleviate overcrowding. In general, the costs of forcing a map to be topographically accurate could jeopardise the ostensible usability benefits of schematic designs, a not insignificant consideration given that the majority of all journey choices on the Underground are, presumably, appropriate.

In fact, with a poorly designed map, people can plan inefficient journeys even when this might seem obvious visually. In Figure 7, the roundabout journeys planned on the lower map imply that people tend to use a *hill climbing* strategy: *continue travelling on the same line for as long as possible if it seems to be heading roughly in the right direction* [14]. In this case, the tendency to behave inefficiently may be a result of the lack of colour-coding. Although the individual service pattern is shown, these routes are not colour-coded, making them harder to identify and follow, perhaps triggering unsophisticated planning strategies. Hence, a map may cause inefficient journey behaviour, even in cases where the line

Fig. 7. Enlarged sections from the DLR maps from Figure 2. People were asked to plan journeys from, for example, Canary Wharf to West Silvertown. For the upper map, almost all people suggested changing at Poplar. For the lower map, a substantial minority of people suggested changing at Stratford, a considerable diversion.

trajectories appear to be configured in a non-misleading way. In contrast to the suggestion by Guo, making a map topographically correct might fail to improve journey planning effectiveness, because the more complicated line trajectories could trigger less efficient journey planning strategies.

C. Choosing a Map

The final element of a usability study will, ideally, look at peoples' opinions concerning the maps that they have used, and ask them to choose between different designs. For the Paris Metro map study [16], the dissociation between subjective ratings and choices versus actual performance measures was striking. It is not hard to see why this should be the case for basic planning tasks. Our subjects are primarily based near London, and will have a strong cultural bias towards believing that London's Henry Beck-style octolinear map is an example of design excellence. If they then experience a Paris Metro map designed in the same way, and plan a few journeys, they will only reject this design if their own perception of their performance indicates that the map is failing. However, for the planning deadline task [16], although more people experienced time-outs for the official octolinear design then the curvilinear one, this did not seem to translate into a more adverse assessment overall. Also, in our more recent Paris Metro map research, we have been piloting within-subjects designs. Despite people experiencing both curvilinear and octolinear maps, and directly finding the curvilinear map easier to use, their tendency to prefer it remains unchanged: roughly 50% of people select it in preference to the octolinear design, a similar proportion to the studies in which only one of the two designs is directly experienced. This *metacognitive gap* between people's performance and their preference seems to be almost insurmountable: peoples' expectations and prejudices about design are so strong that they overwhelm their far weaker perceptions of their level of performance. Even so, we might take some solace in the finding that preference for the curvilinear map tends to be at around the 50% mark. People are not so indoctrinated in the octolinearity gold standard that they are incapable of choosing and using alternatives. However, looking at less controversial designs, the clear preference for the worst performing Docklands Light Railway maps [14] should strike a considerable note of caution for any researcher contemplating basing design decisions on public opinion surveys.

IV. Research Pitfalls

In an ideal world, we could develop and test theories of usability, at the very least identifying a general toolkit for best practice. This might indicate how to choose the best design rules for a particular network (hexalinear, octolinear, multilinear, curvilinear, concentric circles, etc.) and within the chosen set of rules, identifying the design principles necessary to maximise the advantages of the schematic depiction. It would also be useful to know how much topographical distortion is permissible, and the best symbols to use for stations, interchanges, and so on. The problem is that, in devising a research program to achieve these goals, there are a number of pitfalls that we need to be aware of. If not addressed, these can limit considerably the strength and generalisability of the conclusions that can be drawn.

A. We need general theories of design, but we can only test specific individual designs

In devising a set of maps to test theories, it is really only individual designs that are being evaluated directly. Hence, if a curvilinear map of a network is easier to use than an octolinear one, we cannot know whether octolinear maps in general are inappropriate for that network. The finding that *map X is easier to use than map Y* is limited, therefore, in how much it can be generalised from in order to make prescriptions for best-

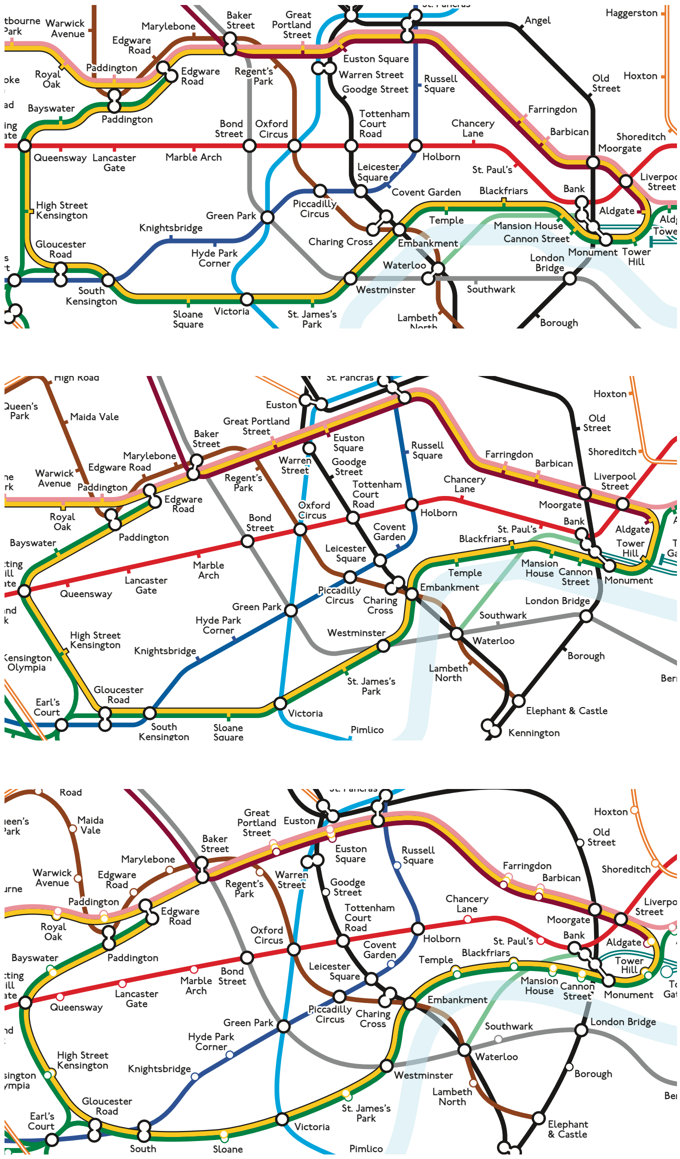


Fig. 8. Three topographically informative schematic maps showing Central London. The octolinear map (upper) requires London to be rotated by roughly 12°, otherwise the Central Line (red) would resemble a flight of stairs, and ruin the coherence of the design at a key location. The multilinear (centre) and curvilinear (lower) maps do not require this transformation. Should it nonetheless be applied, so that the maps are designed as similarly as possible, or not be applied, given that the design priority for each map is the greatest possible topographical accuracy while trying to maintain reasonably simple line trajectories?

practice. This is a problem generally faced in applied psychology/ergonomics, where readers looking for general guidance discover numerous sensible platitudes (*remember the user, don't overload with information etc.*), and some very interesting case histories, but little specific design guidance [11]. There is no watertight solution here but, at the very least, the design of maps should reflect the theories under test. For example, as noted by Roberts *et al.* [16] the finding that the curvilinear map is easier to use than the octolinear design cannot lead to the conclusion that *curvilinear maps are always better than octolinear maps,* nor even *curvilinear maps are always better than octolinear maps for the Paris Metro.* Instead, the only reasonable conclusion is that *octolinear designs are not always easier to use than their alternatives*. Clearly, in order to interpret results a researcher needs to be aware of the scope of what can be concluded from findings for a particular set of designs. For more general and far-reaching conclusions, closely-matched sets of maps would be required, equally optimised within their own design rules, and preferably for multiple cities, but this leads to the next pitfall.

B. *The more we try to match maps, the less well-matched they become*

The problem with matching a set of maps for any study is that, although this can be achieved at a basic level (size of page, font and font size, line thickness, interchange symbols, and so on), and an attempt can be made to optimise in similar ways between designs (e.g. simplest possible line trajectories at the centre of the map, similar size defining feature such as a circle line orbiting the central business district), the different design rules will result in maps that simply *are* different. For example, with the three topographically informative London schematic maps in Figure 8, because of London's tilt, any octolinear design will result in a particularly poor trajectory for the Central Line (red), right at the centre of the map. The solution for an octolinear design is to tilt London by 12.5 degrees, but the two matching designs (curvilinear and multilinear) do not suffer from this problem: their design rules are more compatible with the network structure. Even so, should they likewise be tilted, or not? Different design rules might affect the layout of a complex interchange, so that it is the interchange clarity that is affecting usability, rather than the angles used to lay out the lines. Again, as long as a researcher is aware of these issues, appropriate conclusions will be drawn, and an analysis of individual journeys comparing maps will highlight the extent to which individual design idiosyncrasies are contributing to overall results. Obviously, the more conscientiously that designs are matched, the less this will be a problem, but this leads us to the next pitfall.

C. *Double-blind experiments are easy, double-blind design is harder*

Ideally, research is conducted by people who have no vested interest in the outcomes of their studies, so that they neither influence subject performance nor interpret the findings in such a way that is favourable to their ideas. The phenomenon of the *experimenter bias effect* is well-known in psychology, even if its practitioners are not always entirely careful to avoid this. However, once the designs themselves are created by the theorist, then the problem is compounded. A designer who believes that octolinear maps are always the best might try harder to optimise them, compared with a designer who disagrees (although, conversely, the octolinearity-supporting designer might be more careless, assuming that the design will be good no matter what). The obvious solution would be to take away the human element, and have computers design the maps, but we have not yet reached the point at which these can tackle complex networks with station names fully labelled – the sort of maps where the effectiveness of design really matters – without human intervention [10, 19]. More feasible would be the development of automated scoring criteria, so that there is some objective means of evaluating design, and whose output could be directly analysed in conjunction with performance data. Also important is to enlist as wide a research community as possible, so that a diverse collection of people are creating, testing, and exchanging designs in order to identify general principles of usability.

V. THE WAY FORWARD

A number of techniques have been outlined for investigating map engagement and map usability in sophisticated ways, and a number of interesting findings concerning the effects of map design on perception, journey planning, and route choice have been identified. Given the importance of research diversity, it would certainly not be appropriate to attempt to identify a strict research agenda here. However, a number of pressing issues concerning best-practice for schematic map design can be identified, and the tools for addressing these are available.

The first issue concerns the importance of topographical accuracy. Given that two of the key criteria for many designers – simplicity of line trajectories versus the need for topographical accuracy – can be in considerable conflict with each other, the need for identifying the relative importance of these for journey planning are paramount. For example, it would be straightforward to devise otherwise-matched maps with different degrees of topographical distortion – in key areas and far-flung suburbs – and then conduct station finding and journey planning tasks, testing city experts and novices, in order to identify the extent to which topographical distortion disrupts performance for these people. If the answer to this is *not very much* then designers can safely prioritise simplifying line trajectories over the strictest topographical accuracy.

The next issue concerns the so-called *metacognitive gap*. Why is people's self-awareness of their own journey planning performance so poor, and what can be done to improve this. If market researchers are going to continue to insist on asking the general public to vote for their favourite designs, rather than conduct usability studies, are there at least any ways of conducting opinion polls and focus groups that are likely to lead to a closer match between preference and performance?

Of course, understanding the influence of design on planning performance itself is another important route. In particular, should any design rules (octolinear, hexalinear, curvilinear, multilinear) really be prioritised over any others, and do people's preferences, expectations and experience have any influence on the actual usability of a design (as opposed to their likelihood of accepting it)? What are the most important design principles for effective maps, and how should they be prioritised? If people's route choice is to be influenced, how is this best achieved? Only by taking an evidence-based approach to these issues and others can be we sure that schematic map design in the future will be improved, and the repetition of the mistakes of the past will be avoided

REFERENCES

[1] Bartram, D. J. (1980). Comprehending spatial information: The relative efficiency of different methods of presenting information about bus routes. Applied Cognitive Psychology, 65, 103–110.

[2] Bronzaft, A. L., & Dobrow, S. B. (1984). Improving transit information systems. *Journal of Environmental Systems*, *13*, 365–376

[3] Bronzaft, A. L., Dobrow, S. B., & O'Hanlon, T. J. (1976). Spatial orientation in a subway system. *Environment & Behavior*, *8*, 575–594.

[4] Chabris, C., & Simons, D. (2010). *The invisible gorilla*. New York: Crown Publishing.

[5] Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

[6] Fernandez, D., & Wilkins, A.J. (2008). Uncomfortable images in art and nature. *Perception*, *37*, 1098–1113.

[7] Guo, Z. (2011). Mind the Map! The Impact of Transit Maps on Travel Decisions in Public Transit. *Transportation Research Part A*, *45*, 625–639.

[8] Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it. *Journal of Personality & Social Psychology*, *77*, 1121–1134.

[9] Lindell, A.K., & Mueller, J. (2011). Can science account for taste? Psychological insights into art appreciation. *Journal of Cognitive Psychology*, *23*, 453–475.

[10] Nöllenburg, M., & Wolff, A. (2011). Drawing and labeling high-quality metro maps by mixed-integer programming. *IEEE Transactions on Visualization and Computer Graphics*, *17*, 626–641.

[11] Norman, D.A. (1998). *The psychology of everyday things*. Cambridge MA: MIT Press.

[12] Ovenden, M. (2003). *Metro maps of the world* (2nd edn). Harrow Weald: Capital Transport Publishing.

[13] Ovenden, M. (2008). *Paris Metro style in map and station design*. Harrow Weald: Capital Transport Publishing.

[14] Roberts, M.J. (2011). DLR Route Diagrams: A usability study. Unpublished technical report, University of Essex, Colchester, UK.

[15] Roberts, M. J. (2012). *Underground maps unravelled, explorations in information design*. Wivenhoe, Essex: Published by the author.

[16] Roberts, M.J., Newton, E.J., Lagattolla, F.D., Hughes, S., & Hasler, M.C. (2012). Objective versus subjective measures of Paris Metro map usability: Investigating traditional octolinear versus all-curves schematics. *International Journal of Human-Computer Studies*, *71*, 363–386.

[17] Rosenholtz, R. (2011). What your visual system sees when you are not looking. *Proceedings of the SPIE: Human vision and electronic imaging*, *XVI*, San Francisco, California.

[18] Silvia, P.J., & Barona, C.M. (2009). Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empirical studies of the arts*, *27*, 25–42.

[19] Stott, J. M., Rodgers, P. J., Martinez-Ovando, J. C., & Walker, S. G. (2011). Automatic Metro Map Layout Using Multicriteria Optimization. *IEEE Transactions on Visualization and Computer Graphics*, *17*, 101-114.

[20] Wilkins, A.J. (2012). Origins of visual stress. In J. Stein & Z. Kapoula (eds.) *Visual aspects of dyslexia*. Oxford: Oxford University Press.