Map-Induced Journey-Planning Biases for a Simple Network: A Docklands Light Railway Study

Maxwell J. Roberts,

Department of Psychology, University of Essex, UK

Doug Rose

FWT, London, UK

RUNNING HEAD: MAP-INDUCED JOURNEY-PLANNING BIASES

28/08/2015

PREPRINT OF:
Roberts, M. J., & Rose D. (2017). Map-induced journey-planning biases for a simple network: A Docklands Light Railway study.
Transportation Research A, *94*, 446-460.

Correspondence to the first author at: Department of Psychology, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK, email mjr@essex.ac.uk.

Map-Induced Journey-Planning Biases for a Simple Network: A Docklands Light Railway Study

ABSTRACT

A usability study was conducted to identify the most effective prototype Docklands Light Railway map for installation on trains. This comprised a series of tasks that required station finding and also planning of routes between pairs of stations, with response time and accuracy as measures of performance. In addition, subjective ratings of map design were collected via questionnaire-based evaluations, and also ranked preferences between designs. A clear best-option was easily identifiable as a result of this research. The existing design was associated with the most journey planning errors, and two of the prototypes were associated with inefficient journey choices. The latter finding suggested that respondents were using unsophisticated planning strategies that were put at a disadvantage by certain route depictions. This has wider implications for suggestions that schematic maps should maintain topographical relationships in order to facilitate appropriate journey choices, with the danger that the inevitable increased complexity of line trajectories for such designs would simultaneously reduce the ability of passengers to identify the most appropriate routes.

KEYWORDS

Schematic maps, Metro maps, Route diagrams

Journey planning, Planning biases, Usability testing

Map-Induced Journey-Planning Biases for a Simple Network: A Docklands Light Railway Study

Urban transit networks, including rail-based ones, are steadily increasing in complexity worldwide, owing both to lines and extensions being added within existing networks (particularly in Asia), and also to the increasing tendency to integrate different transport modalities within cities, made possible by common fare arrangements. Ovenden (2003, 2009, 2015) provides an excellent time-lapsed overview of this phenomenon worldwide. With this increasing complexity, the user has to cope with ever more options when journey planning – a task that requires identifying appropriate departure and arrival points, and determining the most efficient route for travel from one to the other. The challenge for designers – the creation of legible, effective network maps – likewise increases year-by-year. Indeed, a recent mathematical analysis (Gallotti, Porter, & Barthelemy, 2015) suggests that there is a cognitive limit to the understandability of complex transport networks, and that a number of these worldwide have already exceeded this threshold. Despite the proliferation of journey-planning software, the network map remains an important source of information. Indeed, many applications for hand-held devices are merely network maps with extra functionality added, such as options to add additional layers of information to a base design.

In response to this complexity, many network maps are *schematized,* which involves simplification via at least some of the following techniques.

1) *Omission of surface details other than the most important landmarks*. The intended use of such a design is therefore to plan routes between pairs of *already-known* start and destination stations, rather than the selection of target stations.

2) *Simplification of line trajectories*. At the very least, complex twists and turns may be smoothed, but often the network is depicted using straight lines; typically horizontal, vertical, and 45° diagonals linked by tightly radiused corners. The technical name for this is *octolinearity*, as used for over eighty years on the London Underground diagram.

3) *Global and/or local scale distortion*. For many networks, at least some expansion of the centre and compression of the suburbs is required in order to produce a compact design that is legible. However, in order to accommodate station labels and simplify line trajectories, local distortions may also be applied, with the consequence of altering relative station positions, line directions, and distances between stations, both within and between lines.

Roberts (2012) discusses a number of issues surrounding topographical distortion on schematic maps. Individuals vary massively in their tolerance of this (see also Forrest, 2014; Roberts 2014b), and network maps differ considerably in this regard. For example, at the time of writing, Berlin, Boston, and London have far higher levels than Chicago and Paris. In London and Berlin, distortion is necessitated by extensive networks with widely spaced stations in the suburbs, and dense central areas with closely-spaced stations. In Boston, distortion has been driven by the numerous closely-spaced stations on the ‘Green Line’. For Chicago, the network structure closely follows the city’s grid street layout, and stations are named by intersecting streets – resulting in numerous duplicated station names. Topographical distortion would run the risk of confusing users and conflicting with their mental models of the city. For Paris, there is a very compact dense network; the correct relative positioning of stations is likely to be more important for users when these are nearby than when these are distant.

In general, for any network schematic, if high topographical fidelity is specified for one particular location in order to assist users, then accuracy is required for all locations, because it is logically impossible for a naive user to know, *a-priori*, which parts of a map have been configured to provide journey planning hints, and which have been distorted, for example, in order to accommodate long station names. However, for many networks, a requirement to minimize topographical distortion is in conflict with the necessity to provide a compact legible design with simple line trajectories – the main goal of creating a schematized map to begin with. Roberts (e.g. 2012, 2014a, 2014b; Roberts *et al*. 2013) has argued that merely creating an octolinear schematic depiction of a network is not sufficient by itself to ensure enhanced usability. It is also necessary to simplify line trajectories so that the major system elements, their relatedness, and the underlying network structure can be readily identified. Roberts *et al*. (2013) argue that in the case of Paris, the network properties – dense inter-connected lines with complex trajectories – coupled with the requirement for correct relative positioning of stations, means that the current schematized map fails this prime requirement of presenting the user with a simplified depiction of the network. The twisting meandering lines have merely been converted to short straight line segments and numerous zig-zags, so that the complexity of line trajectories has been reshaped rather then reduced.

The psychological concept of *cognitive load* is of prime importance (e.g., Carpenter, Just, & Shell, 1990). The more information that must be processed in order to perform a task, the greater the cognitive load and therefore the greater its difficulty. This can easily be demonstrated, for example, with items used to test intelligence, where the elements of the hardest items are not only greater in quantity, but also have low *structural salience*: their visual complexity makes their identification difficult (Meo, Roberts, & Marucci, 2007; Roberts *et al*., 2000). An appropriately designed schematic map should have simple line trajectories giving it high structural salience, facilitating both journey planning and learning, so that a virtuous circle is set up, with performance getting better as more is learnt. For such a design, we would expect fast journey planning, few errors, better remembered plans, and more easily reconstructed plans in the event of a failure to remember. In comparison, a poorly designed schematic will not have these benefits, and may even have little to offer compared with a topographical map, other than the simplification entailed in removing street details and most other landmarks.

It has been argued earlier that the criteria for effective schematization – simplifying line trajectories while attempting to maintain topographical accuracy – are potentially in conflict. It is therefore important to consider how topographical information might assist users in journey planning and, conversely, how distortion might hinder their efforts. If a compact schematic map has most surface detail omitted, then its utility for identifying target stations is limited. However, in the event of service disruption, a map that preserves relative station positions and distances will enable users to identify nearby alternatives. Hence the need for topographical accuracy is greater in areas of high station density. Even so, the variable scale of many schematic maps may mislead users, and Mijksenaar and Vroman (1983) proposed a hybrid map of the London Underground, with curvilinear lines at a topographically accurate centre to show accurate station placement in the region of the map where this information is likely to be of most use, and octolinear schematized suburbs, for simplicity, where topographical accuracy and scale is less important.

Guo (2011) discusses two further ways in which topographical distortion can influence journey planning. These are particularly important because the entire basis of the schematic map – facilitation of the planning of efficient journeys – may be undermined. Firstly, by exaggerating distances between stations, it is possible that users may be induced into taking inappropriate journeys in which there is little time benefit, even compared with walking. Secondly, by altering relative distances and directness of competing options, inappropriate roundabout routes may be made to appear viable.

The basis of Guo's (2011) work is an in-depth analysis of actual London Underground journeys. The major finding in this study was that map-depicted journey distance is a better predictor of journey choice than actual journey distance, even for experienced users (see also Hochmair, 2009; Raveau, Muñoz, & Grange, 2011), and that map-implied interchange quality is also influential. The official London Underground schematic does distort topography, but it is difficult to devise global measures of this that would have a valid transparent bearing on usability, such that the consequences for different maps with different distortion scores could be predicted. For example, Guo (2011) calculated distances between adjacent stations on the map with distances between them in reality, finding a correlation of .22 between these values. Any global function that systematically compresses suburbs and expands the centre will depress such a correlation, despite preserving local spatial relationships faithfully. Hence it is difficult reconcile the global journey planning tendencies identified by the study with global measures of distortedness in order to identify global loss of utility as a result of sub-optimal journey choices. It would therefore be an over-reaction to conclude thus far that any topographical distortion on schematic maps should be forbidden, because the resulting increased complexity and reduced legibility of such designs could defeat the intentions of schematization (Roberts *et al.*, 2013). Hence it might be considered that the benefits of a simple schematic with topographical distortion (ease of use, publicity value) outweigh the costs (a minority of passengers choosing a less direct option). The findings of Guo (2011) therefore provide guidance for designers on how map configurations provide the potential for (1) misinformation that should be avoided, such as the types of layouts that might lead to inappropriate journey choices, and (2) demand management, such as discouraging use of an over-used option, shifting preference towards an under-used one.

One such case study highlighted by Guo (2011) is on the official London Underground map, where there is considerable distortion around Paddington station. The distance between Queensway and Bayswater station entrances is around 100 metres, but the map exaggerates this, and around four actual journeys per day are made between these stations, despite requiring a change of trains at Notting Hill Gate (personal communication from TfL statistical team). In addition, two equivalent alternatives routes from Paddington to Bond Street are implied because they are depicted as being roughly equidistant on the map: either (1) District/Circle Lines to Notting Hill Gate, then Central Line, or (2) Bakerloo Line to Baker Street, then Jubilee Line. In fact, the first option is considerably longer than the second, and estimated to take up to 15% more time to complete. Guo (2011) reports that this option is used for approximately 30% of journeys. Presumably, the remediation of these highlighted problems would be to show this region of the map more accurately, so that the nearby stations are depicted correctly, along with a more appropriate depiction of route lengths, so that people are deterred from travelling from Paddington to Bond Street via Notting Hill Gate.

In order to capitalize fully on the potential for map configuration to influence journey choice, it is helpful to have some insight into the factors that determine this. Thus far, implied distance has been identified as a major variable, with a shorter route being preferred. However, other factors have also been identified over and above this, in relation to route characteristics. For example, in a study of journey planning comparing the official Paris Metro map with an alternative curvilinear design, Roberts *et al*. (2013) found that for one particular journey, for the curvilinear map, many more users suggested a route that comprises a long section of Line 10, compared with the official octolinear map. The various competing routes were virtually identical in length within each design, but on the curvilinear map, the trajectory of Line 10 was considerably less convoluted than on the octolinear version. This is compatible with findings by Bailenson, Shum, and Uttal (2000) who found that, where there were route choices, the option with the longest and straightest initial segment is preferred. Hence it appears that people will prefer a shorter route, but where there is little to choose between them in this respect, people will choose a route that is *initially* simpler. In terms of schematic map design, this means the straightest route. Another influencing factor, from the research of Christenfeld (1995) is that, again, when there is little to choose between routes in terms of length; people will prefer an option in which the greatest distance is travelled before there is a change of direction. For schematic maps, we might infer from this that people prefer to change between lines late rather than early, so that of all the options facing in roughly the correct direction, the one with the longest straightest *uninterrupted* initial segment is preferred. The motivation behind this would be that users prefer to identify the most direct route available, while deferring the cost of making an interchange, if at all, to as late as possible in the journey.

Usability Experiments

Thus far, we have identified some potential benefits of adhering to topographical accuracy for schematic maps. This may assist in journey choice because people are sensitive to implied map distance when planning. However, the usability study reported here, comparing car line diagrams for the Docklands Light Railway, will suggest that this is an oversimplification, and that the costs of topographical accuracy could, in certain circumstances, fail to yield the expected benefits of improved journey choice. This is because, if increased line trajectory complexity is necessitated by topographical accuracy, then this will raise the cognitive load of journey planning, reducing the quality of planned journeys.

The Docklands Light Railway (DLR) is a self-contained network linking east and south-east London to the centre. The network comprises a number of different routes, and one important factor in journey planning is that not all destinations are directly available from all stations. Docklands Light Railway diagrams on trains and stations show this information, so that it is possible for a user to determine whether a direct journey is available, or whether it will be necessary to change trains. However, individual routes are not colour-coded on the diagram, and so it is necessary to track them accurately, visually in order to identify this.

The purpose of the study reported here was to identify which, of various prototypes, was the easiest to use for a variety of journey planning tasks, including (1) finding named stations; (2) identifying whether direct routes were available between pairs of stations and, if not, which station would be the most appropriate interchange; and (3) identifying the fare zone of stations.

These tasks were presented by displaying maps on computers, with responses collected via touch-screens. Each person received just one prototype. Response times, error rates, and error types were objective measures of performance. Subjective responses to the designs were also collected by soliciting preferred versions from the various designs tested, rank ordered, and also questionnaire ratings of individual designs. The tasks were intended to gain a comprehensive picture of usability differences between prototypes, and also people’s responses to them. Two experiments were conducted. The first compared the existing design of 2011 with three prototypes. The second experiment re-investigated two of the prototypes from Experiment 1, enabling its major findings to be replicated.

For the purpose of this current paper, the error data gave a clear picture on the influence of line configuration on journey selections. These data also provided an opportunity to investigate further the previously-identified lack of any relationship between performance versus preference via objective and subjective measures of usability (e.g. Roberts, 2014b; Roberts *et al*., 2013). A recurring feature of investigations into schematic map usability is that the correlation between these is effectively zero: user ratings of map usability are independent of actual objective measures of usability. Previously, this has been demonstrated using journey planning times for complex journeys. For the current study, the large number of trials, along with the accurate timing possible using the touch-screen methodology, permits a finer grained analysis than previously.

METHOD

Materials

Experiment 1

Four DLR route diagrams, intended for display on trains, were investigated via various objectively-measured computer-presented tasks (see Figure 1). These designs were: (1a) the *Existing map*; (1b) the *Major-Axis map* – which showed a clear straight east-west axis; (1c) the *Major-Axis map (CTc)* – which was similar but showed the important interchange at Canning Town in a compacted format using just three interchange circles; and (1d) the Non-Axis map – which was intended not to imply dominance for any one particular route, and hence was deliberately designed with no clear east-west axis. After the computer-presented tasks, a seventeen-item questionnaire was administered which was intended to identify subjective opinion on the tested design, similar to the format used by Roberts *et al*. (2013). Questions one to thirteen comprised various statements which were rated for agreement/ disagreement on a seven point scale, with seven indicating *strong agreement* with the statement, and one indicating *strong disagreement*. The statements were as follows:

a)

b)

c)

d)

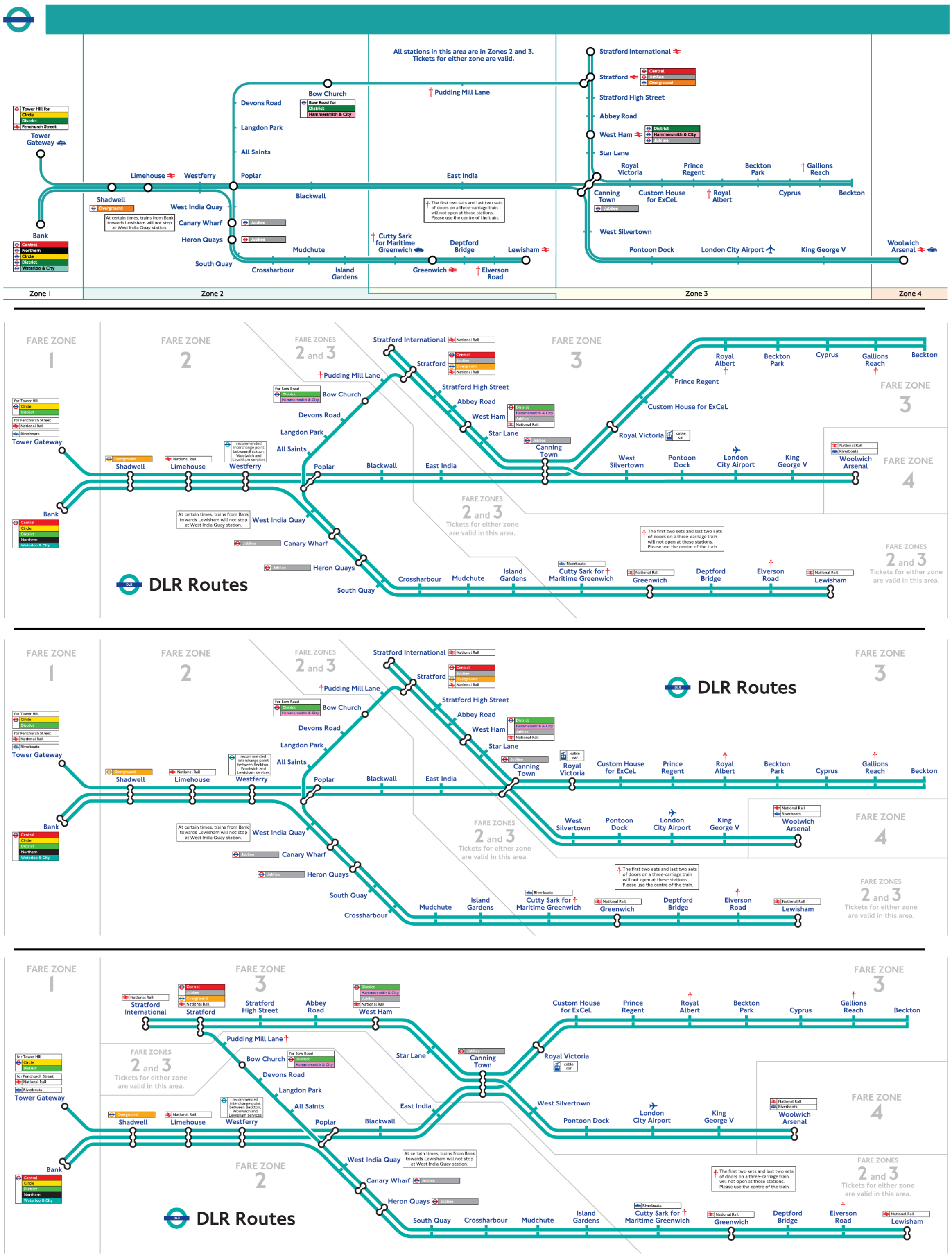

Figure 1. The four maps used in Experiment 1, from the top: (a) Existing map; (b) Major-Axis map; (c) Major-Axis map (CTc); and (d) Non-Axis map.

1) I found stations easy to find on this map

2) Routes were difficult to identify using this map

3) Zones were easy to identify using this map

4) Station interchanges were difficult to negotiate using this map

5) All parts of this map were equally easy to comprehend

6) I found this map disorientating to use

7) I could track the individual DLR routes easily from end to end

8) Some parts of this map looked overly complicated

9) I think that this map is geographically accurate

10) The design of this map felt rather intimidating

11) I found this map clean and uncluttered

12) I found this map visually disturbing

13) I would be happy to see this map used on Docklands Light Railway trains

These questions deliberately had varied polarities, so that agreement with some indicated a favourable rating, and agreement with others indicated an unfavourable rating. This was to counter potential response biases, such as a tendency to agree with statements. Questions fourteen and fifteen invited respondents to give brief written comments concerning likes and dislikes about their test map; people who take part in usability studies often desire an opportunity to make further comments, although these seldom yield useful insights. Question sixteen queried frequency of DLR usage on a seven point scale with options ranging from *every day* to *never/not for years*. For Question seventeen, respondents were asked to identify preferred designs. They were shown all maps in the experiment, A4 sized, laminated, and asked to rank order them from most to least preferred.

Experiment 2

Experiment 2 repeated the tests conducted for Experiment 1 on two of the prototypes: the *Major-Axis* map and the *Non-Axis* map. A slight change was made to each of these: a text box with interchange advice was removed from Westferry Station. The questionnaire was identical in format to Experiment 1, with all four maps ranked for the final preference task.

Respondents

Experiment 1

Testing was conducted on students at the University of Essex, recruited via a departmental subject panel system. There was a payment of £6.00 for taking part. Age and sex compositions for the four groups were as follows: Existing Map; 29 females, 12 males, mean age = 21.8 years, *SD* 3.7; Major-Axis map; 28 females, 12 males, mean age = 21.8 years, *SD* 3.8; Major-Axis map (CTc); 31 females, 10 males, mean age = 23.8 years, *SD* 5.3; Non-Axis Map; 28 females, 12 males, mean age = 21.2 years, *SD* 4.8. In terms of experience at using the DLR, for each group the median response was the second option: *once a year or less*.

Experiment 2

Sampling and payment were identical to Experiment 1. Age and sex compositions for the two groups were as follows: Major-Axis map; 23 females, 18 males, mean age = 22.9 years, *SD* 7.5; Non-Axis map; 22 females, 18 males, mean age = 20.5 years, *SD* 2.9. In terms of DLR usage, the median rating was *once a year or less* for the Major-Axis map, but exactly between *once a year or less* and *never/not for years* for the *Non-Axis* map. In general, all group composition for both experiments tended towards little experience at using the DLR.

Procedure

All experiments commenced with a sequence of five computer tasks. These were conducted in a laboratory setting and supervised by experimenters. Tasks were presented using Apple Macintosh computers running *Superlab 4.0.7b* software on ACER T231H touch-screens at 1920 x 1020 resolution. Each map made full use of the width of the screen, and this resolution was sufficient for the station names to be easily legible. For all tasks the touch-screen fields were centred on station markers, around 2cm by 2cm where possible, and respondents were asked to touch the station marker to register a response, rather than the station name. The first and fifth tasks are not described in detail here. For the first, individual stations were highlighted in yellow, and the task was to touch the station as quickly as possible – the purpose of this was simply to familiarize respondents with the use of the touch-screen and does not yield insights into differential usability between designs. For the fifth, respondents were given station names, and were asked to identify their fare zones. No differences between maps were found for this task and so it is not reported here. No feedback regarding accuracy or time performance was given during any of the tasks. After the computer-presented tasks, the testing session was completed by administering the questionnaire.

Computer Task 2: Station finding task

For this task, for each trial, a respondent was given a station name at the top of the screen and asked find and touch the correct station on the map itself. All respondents received the same stations, which were chosen to ensure that all branches were represented. There were twenty trials, and their order was randomized for each presentation. The instructions were given on-screen and included illustrative diagrams based on a simple fictional network comprising three lines and five stations, whose names did not appear on the experimental maps. Respondents were instructed as follows:

For this task you will see the DLR map. For each trial, you will see a a station name displayed above the map. Every time a station is named, touch the station as quickly as possible. Try to touch the actual station symbol, not the station name.

After reading this, a practice trial commenced using the fictional example, followed by an opportunity to ask questions.

The sequence of events for each trial was as follows: a blank screen was displayed for one second. This was then replaced by the DLR map with a station name displayed above it. This remained on-screen until the respondent made a response, which cleared the screen and commenced the next trial. After all trials were completed, the next task was introduced.

Computer Task 3: Simple route identification task

Respondents were asked to identify routes, but the task was devised such that the need to search for stations was minimized. For each trial, a station was highlighted in yellow and, simultaneously, a name of a terminus was given above the map. The respondent was asked to determine whether there was a direct route (touching a button underneath the map) or else touch an appropriate interchange station (see Figure 2). There were twenty-eight trials, and their order was randomized. The instructions were given on-screen and included illustrative diagrams using the same fictional network as before. The wording was as follows:

The DLR map shows the routes that the trains take, so that you can see whether or not you need to change. For each trial, you will see the name of a terminus, and a highlighted station. Your task is to decide whether there is a direct route from the station to the named terminus. If there is not a direct route, you should identify where you need to change trains. Every journey will be possible with just one change of trains.

After reading this, two practice trials were administered using the fictional network, one with a direct connection, one with no direct connection. There was an opportunity to ask questions.

The sequence of events for each trial was as follows: a blank screen was displayed for one second. This was then replaced with the DLR map. A terminus name was displayed above this and a station was highlighted in yellow. Respondents were asked to decide whether or not there was a direct route between the two, and touch a button at the bottom of the screen once decided. Touching the button displayed a second screen with the highlighted station and terminus repeated, and another button underneath the map. For a direct route, the button was to be touched. For an indirect route, the interchange station was to be touched. The trial remained on display until a response was made, which cleared the screen and commenced next trial. After all trials were completed, the next task was introduced.

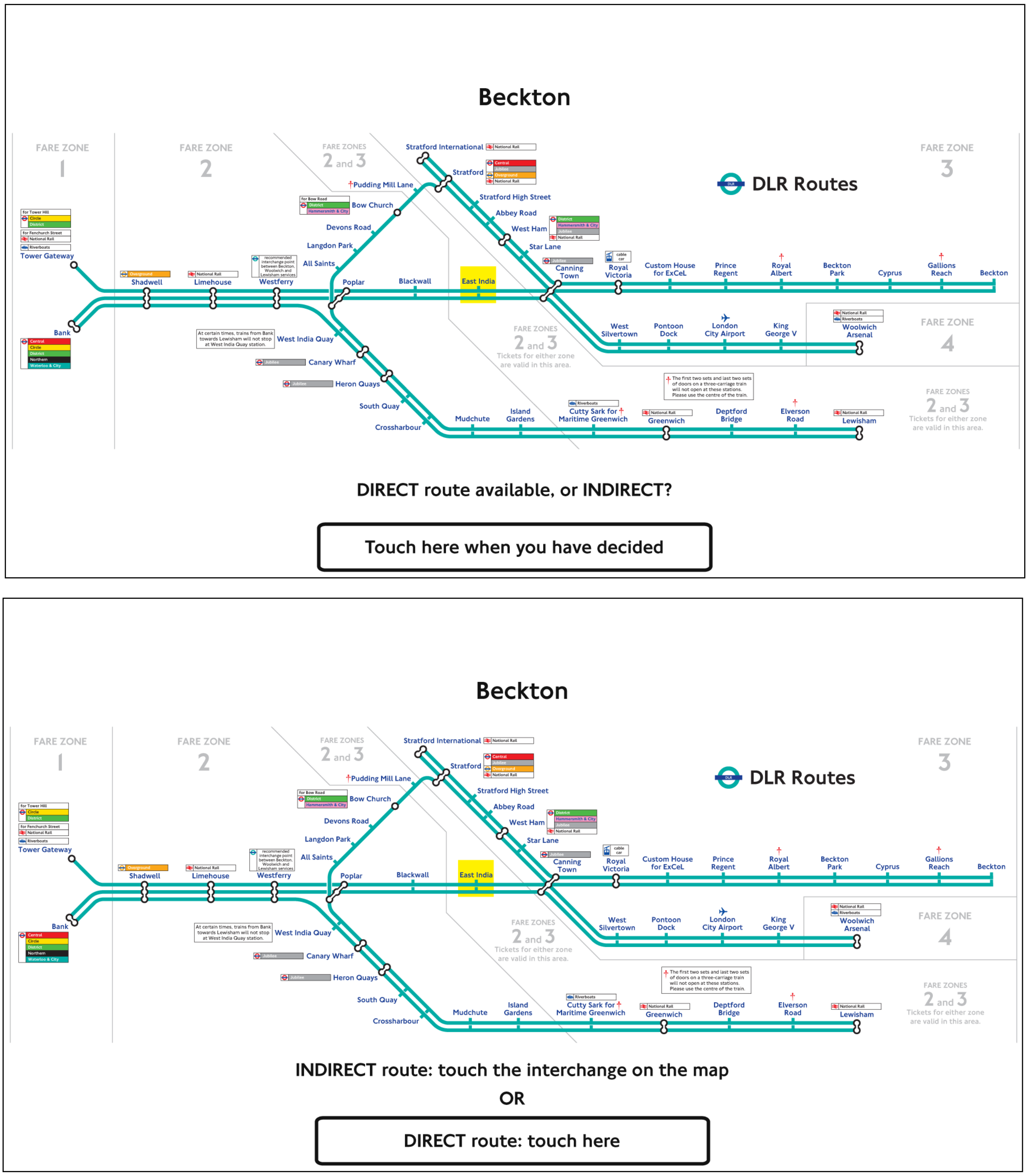

Figure 2. A pair of response screens for a single trial from Task 3. There is a direct route available between the highlighted and the named terminus station.

Computer Task 4: Combined station finding and route identification task

For this, respondents were asked to identify routes between pairs of named stations. The location of these on the map had to be identified with no assistance. The respondent was again asked to determine whether there was a direct route (touching a button underneath the map) or else touch an appropriate interchange station (see Figure 3). There were thirty-two trials, and their order was randomized. The instructions were given on-screen as before. Respondents were instructed as follows:

For the next task, you will be given pairs of station names. Your task is to decide whether there is a direct route between the two named stations. If there is not a direct route, you should identify where you need to change trains. Every journey will be possible with just one change of trains.

After reading this, two practice trials were administered using the fictional network, one with a direct connection, one with no direct connection. There was an opportunity to ask questions.

The sequence of events for each trial was as follows: a blank screen was displayed for one second. This was then replaced by the DLR map with two names displayed above it (e.g., *All Saints* to *Gallions Reach*) and a button underneath the map. Respondents were asked to decide whether or not there was a direct route between the two, and touch the button once decided. Touching the button displayed a second screen with the station names and another button underneath the map. For a direct route, the button was to be touched. For an indirect route, the interchange station was to be touched. The trial remained on display until a response was made, which cleared the screen and commenced the next. After all trials were completed, the station zone identification task was administered, followed by the final questionnaire.

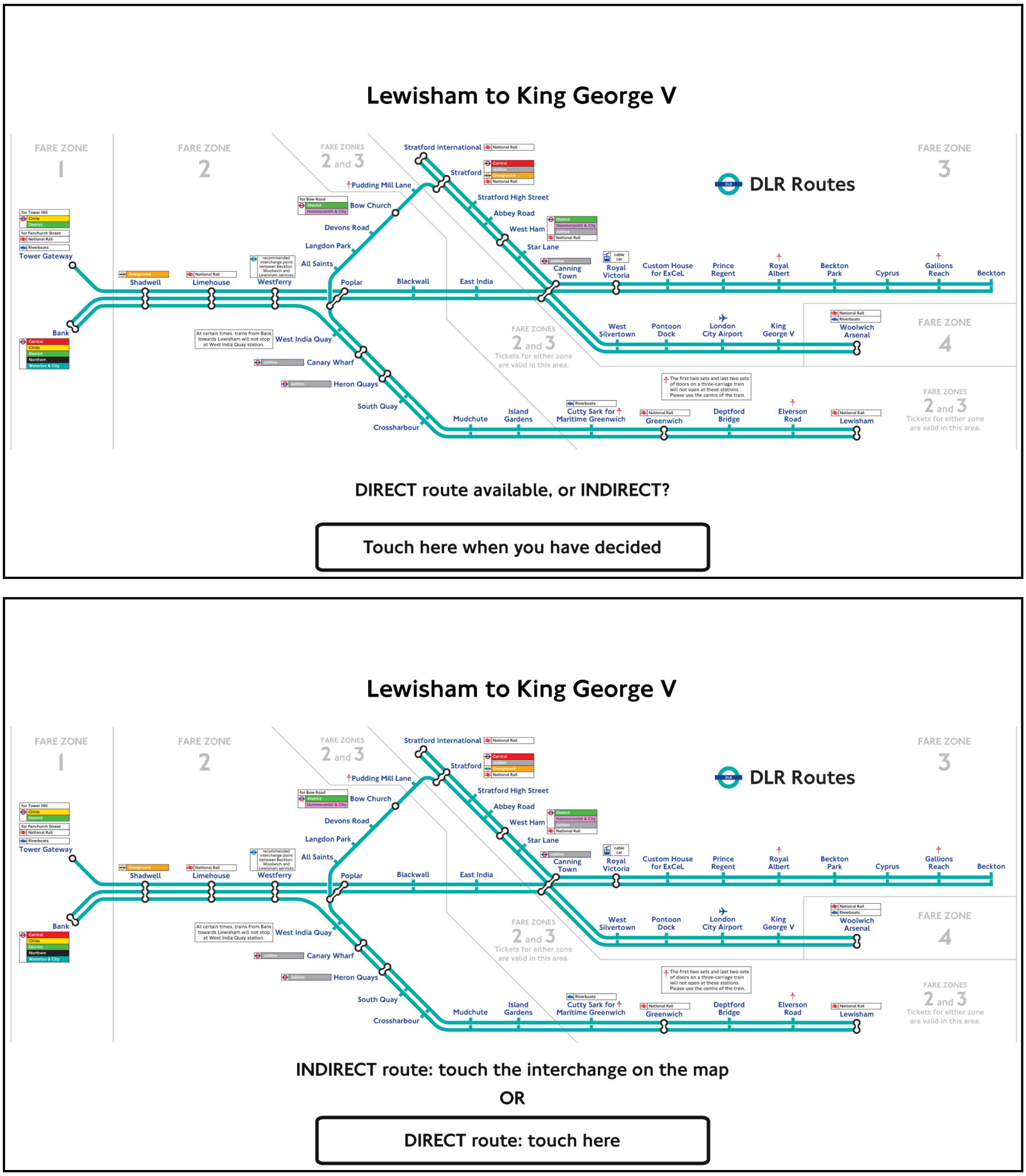

Figure 3. *A pair of response screens for a single trial from Task 4. There is not a direct route available between the two named stations and changing at Poplar gives the most direct journey.*

Design

Each of the computer tasks presented here comprised a between-subjects design experiment: each respondent was tested with just one design of map. Hence, map type was the independent variable, with four levels for Experiment 1 and two levels for Experiment 2. The dependent variables which gave objective indications of usability comprised response times and accuracy/error rates, and well as journey choices, for the journey planning tasks, for indirect routes. Subjective evaluations of usability were identified via the questionnaire scores. For the map ordering task, all respondents ranked all four maps, hence this aspect of the study was a repeated measures design, with map ranking (1st, 2nd, 3rd or 4th) as the dependent variable.

RESULTS

Basic task performance and ratings

Overall response times were calculated by identifying the median response time for each person for each task, excluding times for incorrect responses. Respondents occasionally had extended times for individual trials, for example because of difficulty in identifying the location of an individual station. Calculating medians on an individual basis prevents these outliers from having undue influence on summary scores. Mean response times for each task for each map were calculated from the medians of individuals and are displayed in Table 1. Accuracy rates were calculated simply on the basis of whether a direct route was correctly identified as such, likewise an indirect route, and these are also displayed in Table 1. A more fine-grained analysis of route choice performance is given in the next section. Aggregate questionnaire rating scores were calculated by summing individual responses, taking account of the polarity of each question. For example, with question one, *I found stations easy to find on this map*, a response of seven (*strongly agree*) indicates a favourable evaluation, but for question two, *routes were difficult to identify using this map*, a response of seven (*strongly agree*) indicates an unfavourable evaluation. Scores were reversed as necessary so that high scores always indicated a favourable response. Question nine (concerning geographical accuracy) was excluded because it did not relate directly to an assessment of map usability. Hence aggregate questionnaire scores for individuals ranged from 12 to 84, with 48 representing a neutral value. Previous analysis of a similar questionnaire (Roberts *et al*., 2013) found high internal consistency, with scores subsequently predicting whether a design was subsequently likely to be accepted or rejected. For the map preference task, means were calculated from ranks (from 1 to 4), with a lower score indicating a more favourable ranking.

Where reported, unless stated otherwise, F values are calculated using single-factor between-subjects *Analysis of Variance* to determine whether differences amongst means are significant. The maps were generally closely matched in terms of basic measures of usability, but it is possible to identify some overall patterns. In Experiment 1, the Existing map yielded the lowest accuracy rate of all four designs for both of the journey planning tasks, and the effect was significant for Task 3, $F(3,158) = 3.08$, MSe = 122.2, $p < .05$. Hence there is some basis for rejecting the Existing map despite it having the shortest response times for Task 3, $F(3,158) = 2.88$, MSe = 9.58, $p < .05$ – the one-second per-trial advantage compared with the slowest design is unlikely to compensate for its poor accuracy rate. Without the Existing map in Experiment 2, none of the (small) differences in mean performance between designs was significant, greatest $F(1,79) = 1.74$, MSe = 8.04, $p > .05$, for the response times for Task 3, confirming the close match in performance between prototypes, at least for the basic measures. Hence, the results so far give no basis from which to choose between the prototypes. Of these three, the Major-Axis (CTc) map was always the worst performer in Experiment 1, but there is little to choose from between the other two. However, a finer analysis of journey planning choices reveals a far clearer picture concerning the differing usability of the designs.

Table 1. Means of basic objective and subjective measures of usability for each map for Experiments 1 and 2. Standard deviations are in parentheses.

	Existing map	**Major-Axis map**	**Major-Axis map (CTc)**	**Non-Axis map**
Experiment 1: mean response times (seconds)				
Task 2: Station finding	6.5 *(2.1)*	6.1 *(1.7)*	7.3 *(2.6)*	7.3 *(2.7)*
Task 3: Journey planning	10.0 *(2.8)*	10.3 *(2.6)*	11.6 *(3.2)*	11.4 *(3.7)*
Task 4: Combined task	15.6 *(4.0)*	15.6 *(4.9)*	16.7 *(4.0)*	16.5 *(5.2)*
Experiment 1: accuracy rate (% items correct)				
Task 2: Station finding	96.2 *(5.1)*	96.1 *(4.3)*	95.5 *(6.0)*	96.1 *(4.7)*
Task 3: Journey planning	85.4 *(11.8)*	92.1 *(8.5)*	89.2 *(12.0)*	91.7 *(11.5)*
Task 4: Combined task	88.4 *(12.8)*	92.7 *(7.5)*	91.4 *(8.7)*	93.3 *(8.3)*
Experiment 1: means of subjective opinions on usability				
Questionnaire rating (range 12-84)	56.9 *(12.2)*	57.2 *(13.2)*	59.1 *(11.2)*	55.9 *(10.4)*
Rank order of preference (range 1-4)	2.1 *(1.4)*	2.6 *(1.0)*	2.5 *(0.9)*	2.8 *(1.1)*
Experiment 2: mean response times (seconds)				
Task 2: Station finding	---	6.6 *(2.2)*	---	6.8 *(2.5)*
Task 3: Journey planning	---	9.5 *(2.4)*	---	10.3 *(3.2)*
Task 4: Combined task	---	15.6 *(4.3)*	---	15.1 *(3.2)*
Experiment 2: accuracy rate (% items correct)				
Task 2: Station finding	---	96.0 *(5.9)*	---	95.9 *(6.8)*
Task 3: Journey planning	---	92.4 *(8.0)*	---	94.6 *(9.3)*
Task 4: Combined task	---	94.0 *(9.4)*	---	96.3 *(7.4)*
Experiment 2: means of subjective opinions on usability				
Questionnaire rating (range 12-84)	---	59.8 *(9.8)*	---	57.3 *(12.1)*
Rank order of preference (range 1-4)	2.6 *(1.4)*	2.5 *(1.0)*	2.4 *(1.0)*	2.6 *(1.1)*

Respondents were generally not sensitive to the usability difficulties associated with the Existing map. Indeed, when looking at the preference task, this was more likely to be ranked favourably than the remainder, with a significant difference in means overall: For Experiment 1, using a repeated measures ANOVA, $F(3,463) = 8.82$, MSe = 1.60, $p < .01$. Inspecting the responses for the individual groups showed that the prime driving force for this effect was the Existing map group, who ranked this design, that had been administered to them in the usability tests, much more favourably (mean = 1.5) than any of their non-experienced designs. There was little evidence for any similar exposure effect for the other three groups. With the Existing map group deleted, the rank ordering differences between maps was no longer significant, $F(3,360) = 1.80$, MSe = 1.67, $p > .05$. In Experiment 2, with no groups receiving the Existing map for computer-testing, there were no significant differences in rank ordering, $F(3,360) = 0.50$, MSe = 1.65, $p > .05$. Hence the Existing map is only preferred by those people who experienced it extensively.

Looking at the aggregate questionnaire scores, all designs were rated as reasonably usable, even the Existing map, and there were no significant differences between the rating scores of maps for either Experiment 1, $F(3,158) = 0.54$, MSe = 138.8, $p > .05$, or Experiment 2. $F(1,79) = 1.08$, MSe = 120.1, $p > .05$. However, looking at the questionnaire measures in relation to the objective measures, there does seem to be some evidence that performance differences feed through into rating differences. Combining all data for Experiments 1 and 2, correlations between response times and aggregate questionnaire scores were very small and not always significant (Task 2: $r = -.15$, $p < .05$; Task 3: $r = -.09$, $p > .05$; Task 4: $r = -.17$, $p < .01$; aggregate for all three tasks: $r = -.17$, $p < .01$) but the correlations between accuracy rates and aggregate questionnaire scores were small and always significant (Task 2: $r = .27$, $p < .01$; Task 3: $r = .31$, $p < .01$; Task 4: $r = .26$, $p < .01$; aggregate for all three tasks: $r = .35$, $p < .01$). The directions of the effects indicate that people who took longer and/or made more errors for the objective tasks had a tendency to rate their design less favourably for the questionnaire task. Hence, with the availability of finer measures of performance than previously (Roberts *et al.*, 2013) it is possible to show that users do have *some* sensitivity to their performance when forming opinions about usability, even without direct feedback.

Journey choices

A far clearer picture of map usability can be obtained by looking at the journey planning choices made by respondents. A number of categories of errors and inappropriate choices can be identified.

1) *Failed to identify a direct journey.* A direct route is available between two stations, but this is incorrectly identified as being indirect.

2) *Direct journey is asserted in error.* There is no direct route available between two stations, but this is incorrectly identified as being direct.

3) *Touched first or last*. One of the two target stations is touched in error.

4) *Major interchange error.* The chosen interchange is impossible for completing the test journey, or else the journey is so roundabout as to be absurd.

5) *Interchange error.* The chosen interchange is impossible for completing the test journey unless a very roundabout route with more than one interchange is intended.

6) *Inefficient journey.* The chosen interchange is possible for completing the test journey but the most efficient route is not the one that has been identified.

For each individual journey, an error was placed into one of the categories described. Tasks 3 and 4 were combined for this analysis. Potential touch-screen inaccuracy was accounted for by considering adjacent stations for unusual responses. For example, if an inappropriate interchange registered, but this was directly adjacent to an appropriate interchange station, this was counted as a correct response. Percentages are calculated from thirty trials, the number from which it was possible to make each type of error (thirty direct trials, and thirty indirect, giving sixty in total for the two tasks combined). Table 2 shows the percentages of responses that fell into each error category, with 4 and 5 merged.

Table 2. Percentages of planning errors and inefficient journey choices for Experiments 1 and 2, Task 3 and Task 4 have been combined. Higher scores indicate worse performance. Standard deviations are in parentheses.

	Existing map	**Major-Axis map**	**Major-Axis map (CTc)**	**Non-Axis map**
Experiment 1: Percentages				
1) Failed to identify direct journey	8.9 *(8.7)*	6.3 *(8.3)*	8.5 *(14.6)*	6.6 *(8.5)*
2) Direct journey asserted in error	16.7 *(18.6)*	8.4 *(10.2)*	10.5 *(10.0)*	8.2 *(11.3)*
3) Touched first or last	1.0 *(2.0)*	1.7 *(7.4)*	0.5 *(1.8)*	1.0 *(5.3)*
4 & 5) Interchange error	4.6 *(5.8)*	1.8 *(3.5)*	2.4 *(4.6)*	1.3 *(2.6)*
6) Inefficient journey	4.4 *(5.8)*	9.8 *(7.0)*	9.2 *(7.5)*	2.7 *(4.4)*
Experiment 2: Percentages				
1) Failed to identify direct journey	---	5.2 *(7.3)*	---	5.4 *(11.6)*
2) Direct journey asserted in error	---	8.1 *(9.3)*	---	3.7 *(7.9)*
3) Touched first or last	---	0.7 *(2.7)*	---	2.0 *(7.5)*
4 & 5) Interchange error	---	3.2 *(5.2)*	---	2.0 *(3.3)*
6) Inefficient journey	---	7.2 *(5.7)*	---	1.9 *(3.4)*

Looking at Experiment 1, there were no significant differences between groups for proportions of errors made either for failing to identify direct journeys (1), or for failing to perform the task correctly by touching either of the goal stations rather than the necessary interchange (3), $F(3,158) = 0.65$, MSe = 107.8, $p > .05$ and $F(3,158) = 0.43$, MSe = 22.2, $p > .05$ respectively. There was a significant difference between maps for incorrect assertions of direct journeys (2), $F(3,158) = 3.80$, MSe = 169.7, $p < .05$. *Post hoc* Newman-Keuls tests indicated that the Existing map was associated with significantly more errors compared with all of the others, all $p < .05$. The other map groups did not differ significantly for this error category, all $p > .05$. There was also a significant difference between maps for interchange errors (4 & 5), $F(3,158) = 4.68$, MSe = 18.5, $p < .01$. *Post hoc* Newman-Keuls tests indicated that the Existing map was again associated with significantly more errors compared with all of the others, all $p < .05$, except versus the Non-Axis map, $p < .01$. Again, the other map groups did not differ significantly for this error category, all $p > .05$. There was also a significant difference between map groups for inefficient journeys (6), $F(3,158) = 12.30$, MSe = 39.7, $p < .01$, but this revealed a different pattern compared with the other categories. This time, *post hoc* Newman-Keuls tests indicated that the two Major-Axis maps were associated with significantly more journeys of this type than *either* of the other two designs, all $p < .01$. The two Major-Axis maps did not differ in this respect, $p > .05$, nor did the Existing map versus the Non-Axis map, $p > .05$.

Overall, for Experiment 1, where there were differences in errors, the Existing map was always worse than the other three (which were equally good by this criterion). The earlier finding of the popularity of the Existing map amongst respondents who were tested with this design is therefore not justified, given that it is associated with the worst accuracy rate for the objective measures of performance. Looking at inefficient journeys, the Major-Axis maps were much worse in this respect than the other two. The Non-Axis map was always amongst the best performers.

Where applicable, this pattern of findings was broadly replicated in Experiment 2 when comparing the Major-Axis and Non-Axis maps. The two maps did not differ for error categories (1), (3), or (4 & 5): $F(1,79) = 0.01$, MSe = 92.8, $p > .05$, $F(1,79) = 1.03$, MSe = 31.6, $p > .05$, and $F(1,79) = 1.65$, MSe = 19.0, $p > .05$, respectively. There was, however, a significant difference between maps for error category (2), incorrect assertion of direct journeys, $F(1,79) = 5.41$, MSe = 74.6, $p < .05$. However, there was no evidence for such an effect in Experiment 1, and the inconsistency across separate experiments means that caution should be exercised in interpreting this effect. The key finding of interest was replicated, with significantly more inefficient journeys proposed (6) for the Major-Axis map than the Non-Axis map, $F(1,79) = 25.9$, MSe = 22.2, $p < .01$.

Looking at the patterns of journey choices in the raw data, the inefficient options were not distributed randomly. Instead, they were particularly likely to occur for journeys between stations from Canning Town eastwards (the Beckton and Woolwich branches) and the line from Lewisham to West India Quay (plus All Saints and Langdon Park). For these, the most efficient route in reality, and implied by all maps tested here (both in terms of length of route and number of stations traversed) is to travel direct to Poplar and change trains (see Figure 4). To investigate this further, the eleven trials which matched this journey criterion were investigated separately, looking specifically at responses where an interchange at Stratford was chosen, indicating the suggestion of an inefficient route.

Table 3. Percentages of inefficient journeys – interchange at Stratford proposed – involving the eleven trials in which travel was required between the Lewisham branch (plus All Saints and Langdon Park) and Canning Town/eastwards. Task 3 and Task 4 have been combined. Higher scores indicate worse performance. Standard deviations are in parentheses.

	Existing map	**Major-Axis map**	**Major-Axis map (CTc)**	**Non-Axis map**
Experiment 1	11.1 *(14.9)*	23.2 *(17.1)*	19.3 *(19.5)*	6.4 *(10.3)*
Experiment 2	---	16.9 *(12.3)*	---	3.4 *(6.7)*

This analysis highlights the dominant sources of inefficient journey choices. For the worst-performing maps (Major-Axis), the roundabout route via Stratford was proposed for between 17% and 23% of trials for journeys where this was possible, but only for 3% to 6% of trials for the best map (Non-Axis). Differences between groups were significant for both Experiments 1 and 2, $F(3,158) = 9.37$, MSe = 250.9, $p < .01$, and $F(1,79) = 25.9$, MSe = 22.2, $p < .01$, respectively. For Experiment 1, *post hoc* Newman-Keuls tests again indicated that the two Major-Axis maps were significantly worse than the other two, all $p < .01$, except for the Existing map versus the Major-Axis map (CTc), $p < .05$. The two Major-Axis maps did not differ, nor did the Existing map versus the Non-Axis map, both $p > .05$. Overall, it seems that a design aspect of the Major-Axis maps is failing to discourage certain inefficient journeys.

DISCUSSION

The prime purpose of the usability experiments reported here was to enable the identification of the most effective prototype Docklands Light Railway diagram suitable for placement on trains. In this sense the findings are unequivocal. The Existing map was popular amongst those people who were tested extensively with it, but it was also subject to the most planning errors, indicating that its routes were difficult to track. The Major-Axis maps yielded fewer outright errors, but these designs were particularly associated with inefficient route selections for certain journeys. The Non-Axis map never displayed any of these weaknesses, and an adaptation of this was adopted for use on DLR trains, using it's Stratford configuration, that was associated with more efficient journey planning. The need for objective evidence-based decisions regarding information design is highlighted when we consider the subjective ratings for the different versions. At best these were uninformative, and at worst these were in contradiction with the objective data. There is some evidence that people were sensitive to their own performance when rating the designs, but the correlations were generally weak.

Why is there a disconnection between preference and performance? Users (and transport managers) possess a wide range of beliefs and expectations concerning schematic map design, and also have aesthetic preferences that will affect their judgement (Roberts, 2014b). This will inevitably influence their ratings, especially as people in general have poor self-insights into their own cognitive performance (e.g., Chabris & Simons, 2010; Kruger & Dunning, 1999). Even if people do have an awareness of their speed and accuracy when using a particular map, the absence of any opportunity for a baseline for comparison will prevent its relative effectiveness from being evaluated. The judgement of absolute effectiveness for just one single design that is only experienced by itself is virtually impossible. Hence, without direct explicit feedback concerning response times and errors, and without experience of a range of different designs, people will have little basis from which to exercise a rational judgement based on the perception of their own performance. Even so, this study did identify correlations between objective measures and aggregate questionnaire scores, with the least favourable ratings from the slowest and least accurate people. The likely source of this relationship is those individuals who had repeated difficulty with locating stations and tracking routes. Only such persistent and obvious failures would be expected to lead to adverse evaluations, although this is more likely to have reflected personal difficulties rather than any systematic design failures in the current study.

Overall, the experiments here demonstrate the potential value of objective usability studies in selecting between prototype schematic maps. Differences in effectiveness and efficiency between options can be identified even for versions depicting simple networks created by expert designers. However, it is important to note that numerous trials, and a fine-grained analysis of patterns of performance may be necessary to identify potential subtle weaknesses. The inefficiency of the Major-Axis designs here was not apparent from the basic usability measures in Table 2. Less extensive studies may appear to be more economical to implement but would be expected to yield usability differences only if there are gross design inadequacies for maps of complex networks (Roberts *et al.*, 2013).

Given the lack of relationships between subjective preference and objective performance, it is safe to conclude that a selection process based upon, say, a public vote, should be treated with caution (*Boston Globe*, 2013). Obviously, there will always be a role for soliciting public opinions in marketing, for example, when identifying flavour preference when introducing new food products. But, to continue this analogy, to ask non-experts to determine nutritional value from these new products simply by tasting them would not lead to useful or coherent data. The key here is to be able to identify when user preference by itself is likely to be an inappropriate measure of design effectiveness. Information design should focus on user performance if speed and ease of use is the priority.

One potential criticism of this study is that it was conducted under laboratory conditions. The respondents were not planning real journeys that they intended to implement. Instead, they were identifying numerous fictitious journeys sequentially. It is possible that the substantial bias towards inefficient journeys for the Major-Axis maps in certain instances might not manifest itself in reality. In response to this criticism we should note that the actual planning errors for the Existing map were made under ideal conditions in terms of lighting and lack of distraction, and were presumably due to difficulties in reading the design. The inefficiencies of the Major-Axis maps were substantial and replicable. Having identified potential weaknesses for these designs, as well as identifying an alternative that displayed no evidence of them at all, a very clear picture was obtained from the point of view of prototype selection and development. Might there be hidden weaknesses with the Non-Axis map that can only be revealed under real-life conditions? Specific hypotheses relating the structure of information to performance can always be empirically tested, but for a general unspecified criticism such as this, the *empirical research skeptic* would need to explain how real life conditions might extinguish biases for certain designs – presumably, increased conscientiousness due to real-time investment and consequences of mistakes – but create new biases for other designs. Overall, the utility of the methodology used in this study for identifying potential usability difficulties and journey planning biases is clear. In general, it would be a very brave transportation authority, or operator, indeed, that chose to disregarded any such warning signals identified in this way.

Past research (Guo, 2011) has shown that information configuration on maps can have real consequences for actual journeys made, and so the way forward here is to attempt to understand and explain the differential map bias observed here, in relation to inefficient journey choices, and integrate this into what is already known about journey planning behaviour, then determine whether this is likely to have a bearing on travel in reality, and from this make predictions for future research.

The planning inefficiency observed is that journeys for the Lewisham branch to and from eastward destinations – for example Canary Wharf to London City Airport – were often routed via Stratford, rather than the more direct route changing at Poplar. Not only is the inefficient option explicitly shown on the Major-Axis map as a lengthy detour (around 50% farther) but eight additional stations must be traversed compared with the more efficient route. For the Major-Axis maps, which were particularly susceptible to inefficient journeys towards eastern destinations, the roundabout route from Poplar to Stratford likewise has an eastwards component to it. For the Non-Axis map, which is much less susceptible to this inefficiency, this same section of line heads north-west, away from eastern target stations.

This pattern of findings suggests that respondents were navigating the maps by using a simple, relatively unsophisticated *hill-climbing strategy* (e.g., Anderson, 1993; Bassok & Novick, 2012). For this, people evaluate different options at each choice point, choosing the one with the most perceived benefits and the least costs in terms of achieving the goal. In this situation, the costs of the not-quite direct route on the Major-Axis map are set against the greater perceived costs of making an interchange, leading to the general behaviour pattern of continuing to travel for as long as a route remains sufficiently direct, until an interchange is necessitated. For the Non-Axis map, the route to Stratford, apparently in the opposite direction to the same goal, carries a greater perceived cost than the interchange at Poplar, leading to the selection of the more efficient route in preference.

Hill-climbing can easily lead to inefficient routes, but in theory people can use alternative strategies. *Means ends analysis* applied here would involve the task being decomposed into subcomponents. If no direct route is available, potential intermediate points would be identified as individual sub-goals, with the full task broken down into a set of sequential problem-solving tasks, each one aiming for a new sub-goal until the final destination is reached. Such a strategy should eliminate the efficient route via Stratford as it would not be deemed an appropriate intermediate point. However, more sophisticated strategies that result in more efficient behaviour impose a greater cognitive load on the user, and if cognitive resources are insufficient, then less demanding strategies such as hill-climbing may be resorted to. In the case of the maps here, it is suggested that the different designs result in different planning behaviour, not because they present the user with different levels of cognitive load, resulting in the application of different search strategies, but because their configurations trigger different selections within the same strategy.

The Docklands Light Railway network is relatively simple, and so its navigation is unlikely to be associated with high levels of cognitive load, so what might be triggering the use of a hill-climbing strategy? One possibility is the lack of colour-coding of individual routes, which makes the map more difficult to navigate. For example, Garland, Haynes, and Grubb (1979) found that for a complicated map, removal of colour coding reduced performance, reduced confidence, and raised frustration when planning journeys. With colour-coding of routes not a permissible alternative for the prototype DLR maps, a choice of design that minimizes the potential for journey inefficiency is therefore the next best option.

There are other configural aspects of design that might also impinge on the sophistication of journey planning strategies. This includes the complexity of the line trajectories themselves, as this will have a direct impact on the cognitive load of using a design (e.g., Roberts, 2012, 2014b). For a particularly complex system, with poorly optimized line trajectories, the underlying structure of the network may be so difficult to discern that, not only will sophisticated strategies be less likely to be used, but when they are applied, the difficulty in identifying appropriate subgoals may make strategies such as means-ends analysis function less effectively. We should note that even without colour-coding, the DLR network is relatively simple. The potential for inefficient planning and resulting roundabout routes will surely be compounded for more complicated networks, with multiple options for many journeys, the need to change trains more than once, and new coding issues that arise once the number of lines exceeds the number of easily differentiable colours that are available.

The analysis here is compatible with earlier suggestions for the factors that influence journey choice decisions when planning by using schematic maps, but does indicate that important revisions need to be made to these, in particular taking account of line configurations and the cognitive load associated with using a particular map. Although the current suggestions are easily testable in laboratory-based usability studies, possibly using fictitious networks expressly designed to test these issues, transport undertakings may be reluctant to allow them to be tested in the field by altering configurations of official network maps. On the other hand, if journey data are available, then it could be possible to identify official maps with configurations that potentially lead to inefficient journeys, and identify the extent to which passengers actually choose these.

To put these findings into perspective, we need to return to the specific example of Paddington to Bond Street identified by Guo (2011). For this, and in line with his finding that apparent map distance is an important determinant of journey choice, of the two apparently equidistant options depicted on the map, the one that was less efficient in reality attracted 30% of real journeys (presumably, this figure would approach 50% amongst inexperienced Underground users). In the DLR map usability experiments described here, on certain designs, the less efficient option attracted around 20% of suggested journeys even though this was clearly depicted as sub-optimal. Hence, people may well use apparent map distance in order to choose journeys, and may well be able to use this information to select the most efficient routes if reality is correctly depicted, but if the sophistication of journey planning strategies becomes insufficient for whatever reason, then the ability of people to make full use of configural information may be limited. This returns us to the original points raised in the introduction. For many complex networks, the design priority of simple line trajectories will often be incompatible with the design priority of maintaining topographical accuracy. If one purpose of the latter is to assist users in their ability to choose efficient journeys where there is a range of options available, then an unfortunate side effect of topographical accuracy could be the inevitable consequences of (1) an increase in complexity of line trajectories; and (2) a reduction in salience of network structure. Together, these could well mean that users are unable to capitalize on topographical assistance, owing to their resulting inability to apply the most sophisticated journey planning strategies, so that people become *less* able to evaluate and choose between different options despite this supposed assistance.

ACKNOWLEDGEMENTS

This project was commissioned and financed by the *Docklands Light Railway*, part of the *Transport for London* rail network. The prototype designs were created on their behalf by Doug Rose at FWT. We are grateful to Samuel Waterson, Stephen Southworth, Jane Roberts, Ayesha Merali and Priscila Buschinelli for their invaluable assistance in data collection.

REFERENCES

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, *48*, 35-44.

Bailenson, J. N., Shum, M. S., Uttal, D. H. (2000). The initial segment strategy: a heuristic for route selection. *Memory and Cognition*, *28*, 306–318.

Bassok, M., & Novick, L. R. (2012). Problem solving. In K. J. Holyoak & R. G. Morrison (eds.) *Oxford handbook of thinking and reasoning* (pp 413-432). Oxford: Oxford University Press.

Boston Globe (2013). MBTA map-making contest garners 17,000 votes. http://www.bostonglobe.com/metro/2013/09/21/map-making-contest-garners-votes/egeCO7x7Q8rbfciGVHaSyI/story.html [accessed 01/06/2015]

Carpenter, P. A. Just, M. A., and Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, *97*, 404–431.

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

Christenfeld, N. (1995). Choices from identical options. *Psychological Science*, *6*, 50-55.

Forrest, D. (2014). Causes and consequences of scale change in schematic maps: are users aware and do they care? Schematic Mapping Workshop 2014, University of Essex, April. https://sites.google.com/site/schematicmapping/Forrest_ScaleChange.pdf [accessed 29/05/2015]

Gallotti, R., Porter, M. A., & Barthelemy, M. (2015). Information Measures and Cognitive Limits in Multilayer Navigation. arXiv:1506.01978

Garland, H.C., Haynes, J. J., & Grubb, G. C. (1979). Transit map color coding and street detail effects on trip planning performance. *Environment and Behavior*, *11*, 162–184.

Guo, Z. (2011). Mind the map! The impact of transit maps on travel decisions in public transit. *Transportation Research Part A*, *45*, 625-639.

Hochmair, H. (2009). The Influence of map design on route choice from public transportation maps in urban areas. *The Cartographic Journal*, *46*, 242–256.

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

Meo, M., Roberts, M. J. and Marucci, F. S. (2007). Element salience as a predictor of item difficulty for Raven's Progressive Matrices. *Intelligence*, *35*, 359–368.

Mijksenaar, P., & Vroman, R. V. (1983). London Transport map: A Delft project. Typos, 6, 36–40.

Ovenden, M. (2003). *Metro maps of the world* (1st ed.). Harrow Weald, UK: Capital Transport Publishing.

Ovenden, M. (2009). *Transit maps of the world*. New York: Penguin.

Ovenden, M. (2015). *Transit maps of the world* (2nd ed). New York: Penguin.

Raveau, S., Muñoz, J. C., de Grange, L. (2011). A topological route choice model for metro. *Transportation Research, Part A*, *45*, 138–147.

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

Roberts, M.J. (2014a). Schematic maps in the laboratory. Schematic Mapping Workshop 2014, University of Essex, April. https://sites.google.com/site/schematicmapping/Roberts-Empirical.pdf [accessed 29/05/2015]

Roberts, M.J. (2014b). What's your theory of effective schematic map design? Schematic Mapping Workshop 2014, University of Essex, April. https://sites.google.com/site/schematicmapping/Roberts-Theoretical.pdf [accessed 29/05/2015]

Roberts, M.J., Newton, E.J., Lagattolla, F.D., Hughes, S., & Hasler, M.C. (2013). Objective versus subjective measures of Paris Metro map usability: Investigating traditional octolinear versus all-curves schematic maps. *International Journal of Human Computer Studies*, *71*, 363-386.

Roberts, M. J., Welfare, H., Livermore, D. P. and Theadom, A. M. (2000). Context, visual salience, and inductive reasoning, *Thinking and Reasoning*, *6*, 349–374.