www.elsevier.com/locate/ijhcs

Preference versus performance: Investigating the dissociation between objective measures and subjective ratings of usability for schematic metro maps and intuitive theories of design

Maxwell J. Roberts, Hannah Gray, Jennifer Lesnik

PII: S1071-5819(16)30075-1
DOI: http://dx.doi.org/10.1016/j.ijhcs.2016.06.003
Reference: YIJHC2045

To appear in: *Journal of Human Computer Studies*

Received date: 4 February 2016
Revised date: 26 June 2016
Accepted date: 28 June 2016

Cite this article as: Maxwell J. Roberts, Hannah Gray and Jennifer Lesnik Preference versus performance: Investigating the dissociation between objectiv measures and subjective ratings of usability for schematic metro maps an intuitive theories of design, *Journal of Human Computer Studie* http://dx.doi.org/10.1016/j.ijhcs.2016.06.003

Preference versus performance: Investigating the dissociation between objective measures and subjective ratings of usability for schematic metro maps and intuitive theories of design

Maxwell J. Roberts*, Hannah Gray, Jennifer Lesnik

Department of Psychology, University of Essex, UK

RUNNING HEAD: PREFERENCE VS. PERFORMANCE

24/06/2016

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

ABSTRACT

Three experiments are reported in which objective measures and subjective ratings of schematic metro map usability were investigated. Experiment 1 used a within-subjects design to compare octolinear and curvilinear Paris Metro maps. This replicated and extended Roberts et al. (2013); the curvilinear map was associated with faster journey planning times, and yet preference between the two was unrelated to this measure. In Experiment 2, nine matched versions of the London Underground map were rated for usability and attractiveness, and a clear octolinear bias was displayed. It was also possible to identify individuals who held a simplicity theory of effective design, versus an octolinearity theory. Experiment 3 investigated the relationship between usability ratings and journey planning times for three Berlin network maps, all optimized for simplicity of line trajectories. No differences in times were found, and yet usability ratings after experience at using the maps differed significantly, in line with the findings for the London designs in Experiment 2. Overall, the dissociation between objective measures of performance and subjective ratings of usability is robust, and appears to reflect expectations and prejudices concerning effective design. The octolinearity as a gold standard conjecture for achieving optimum usability continues to be refuted.

KEYWORDS

Schematic maps, Metro maps, Journey planning, Objective measures, Subjective ratings

1. INTRODUCTION

Worldwide, the complexity of urban transit networks increases year-by-year, with many large-scale construction projects in progress, particularly in Asian cities, and more generally, owing to increased inter-availability of different transport modes made possible by computerized ticketing systems. These trends are particularly noticeable when comparing the different editions of Ovenden (2003, 2007, 2015), which effectively provide time-lapsed snapshots of development. However, the increased flexibility and coverage comes with a usability cost, and one example of an attempt to quantify this comes from a mathematical analysis by Gallotti, Porter, and Barthelemy (2016) who suggest 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. One solution to this problem is the development of journey-planning computer software, although the disadvantage of relying on this is an inhibition of acquisition of knowledge about the network. This reliance will lead to difficulties in situations where automated assistance ceases to be available, and may prevent attainment of the deepest levels of knowledge which regular users can display, and cannot be matched by these systems (see Ishikawa et al., 2008; Willis et al., 2009, for a review of the issues associated with automated wayfinding software in general). Transport operators therefore continue to make available to passengers traditional journey planning aides such as maps and timetables. 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.

A common technique used to mitigate against complexity in the design of network maps is to create a schematized representation. This typically involves at least some of the following: (1) non-depiction of surface details irrelevant to orientation and journey planning; (2) simplification of line trajectories, often depicting these as only horizontal, vertical, and 45º diagonal straight lines linked by tightly radiused corners; and (3) expansion of congested central areas (with consequential compression of suburbs) to enable station labels to be applied without interrupting lines, and resulting in a more even station density across the design. Famously, these principles were applied by Henry Beck in creating the 1933 London Underground diagram (Garland, 1994). The use of angles chosen by Beck is today known as octolinearity (e.g., Nöllenburg & Wolff, 2011) and this is the most frequent choice by far for schematic maps in which line trajectories are depicted as linear (e.g., Ovenden 2015), although Roberts (2012, 2014a, 2014b) notes that other levels of linearity are also possible, and may be better suited to some networks. For example, implementing a hexalinear design for Central London – three angles at 60º rotations – results in simpler line trajectories, i.e. fewer changes of direction, than using octolinear angles.

1.1 Prescriptions for designing effective schematic maps

Designing a schematic map for a complex network is a difficult task, considerably more so if it is acknowledged that different design rules might suit different networks, and this leads to the problem of pre-determining fitness of purpose: which of the possible designs is most likely to best-achieve the goal of the most effective network depiction? In turn, this necessitates the identification of design criteria so that the various possibilities when creating a schematic map are prioritized appropriately. The difficulties in compiling these are compounded by disagreements concerning their importance amongst designers, amongst users, and between designers and users. For example, the criterion of maintaining relative spatial locations of stations on a schematic map is advocated by many people, and can lead to inefficient journeys under certain circumstances if neglected, but receives a lower priority than the criterion of simplified line trajectories on the London Underground map (Guo, 2011).

One of the more persistent suggested criteria for effective schematic map design is the requirement to use octolinear angles. Roberts et al. (2013) name this the octolinearity as a gold standard conjecture: applying octolinearity will result in the most effective design no matter what the structure of the network (e.g., Ovenden, 2005, p. 39). For example, Nöllenburg & Wolff (2011) described octolinearity as a Hard Constraint (i.e., it should never be broken) and suggest that "the main benefit of octilinear layouts is that they potentially consume less space and use fewer bends while still having a tidy and schematic appearance" (p. 626) and that "we believe that octilinearity, which is strictly followed by most real metro maps, is an essential ingredient for tidy and easy-to-read metro map layouts" (p. 627). The dominance of octolinearity is clear when comparing schematic maps worldwide (Ovenden 2015) but, in a series of comparisons between the official octolinear Paris Metro map and a novel curvilinear design, Roberts et al. (2013) showed that the curvilinear version was always associated with improved journey planning times. This refutes the strong version of the gold standard conjecture, although weaker versions are left intact. For example, octolinearity will usually result in the most effective design, but not in instances where this is incompatible with network structure.

More sophisticated guidance on design objectives for creating schematic maps is remarkably rare (Roberts, 2014b). Nöllenburg (2014) gives a substantial set, for example, keep line trajectories as straight as possible, space stations evenly, station labels should not occlude lines, and relative positions of stations should be preserved. Ovenden (2009, p. 151) provides a collection that is broadly compatible, but includes more subtle prescriptions that may be more likely to affect aesthetic judgement than actual usability. For example, do not bend a line twice between a pair of stations, keep station labels horizontal. Many of these prescriptions seem reasonable, and can be shown to be compatible with theories of human cognition, but there is surprisingly little empirical evidence to demonstrate their efficacy (Roberts, 2014a). They form a somewhat disparate set of principles, not necessarily compatible with each other, and Roberts (2012, 2014b) attempted to organize these into a broad framework of five categories (see Figure 1 for illustrations).

Simplicity.

The key-most requirement for a schematic map is that it converts the complex trajectories of routes into simple line trajectories on the diagram. These will be easier to discern and follow, likewise their interconnections and broad pathways, so that the overall structure of the network is in turn easier to identify and learn. Many designers appear to neglect or misunderstand this criterion, and in the process convert the complex line trajectories of reality into numerous short zig-zagging segments, despite their low utility in terms of information value. An increase in the number of angles (i.e. the linearity level of the design) might enable simpler line trajectories, but at the expense of coherence (see later). In this sense, octolinearity offers a good compromise between complexity in terms of line trajectories versus complexity in terms of number of line angles.

Coherence.

The simplicity criterion refers to individual line trajectories. The way in which these relate together to give the design overall good shape is also important, but harder to define and measure. Objectives can be specified that will contribute to coherence, such as maximizing parallel lines, symmetrical divergence, and aligning stations and termini. Coherence might also be achieved by configuring regular, easily identified shapes, such as circles, equilateral triangles, horizons (grounding the design using horizontal lines) and/or grids. A map with many angles might permit simple line trajectories (see earlier), but the design is likely to suffer from poor coherence if the incidence of parallel lines declines. Again, octolinearity has the potential for being a good compromise in this respect. For example, the relatively small number of angles available means that it is easy for a conscientious designer to ensure that lines are parallel.

Balance.

Ideally, there should be an even density of stations across a map, or at least gentle density gradients, so that congested and empty spaces are not adjacent. The natural consequence of attempting to create a balanced design for an extensive network, with a clear central focus, is that the centre will be enlarged and the suburbs compacted. Occasionally, designers go too far in this respect, and the result is a diffuse centre with wide station spacing, but compressed suburbs with narrow station intervals; the opposite of reality.

Harmony.

Roberts (2012, 2014b) suggests a placeholder category for design aspects that are likely to influence the aesthetics of a design, but are unlikely to have any measurable impact on usability. There will be individual differences in this respect (e.g., Palmer, Schloss, & Sammartino, 2013) but research does suggest that certain shapes and patterns tend to be rated as being more pleasing than others (e.g., Lindell & Mueller, 2011). In the context of maps; angles that permit equilateral triangles (e.g., hexalinear) might result in designs that are preferred over ones with tall thin isosceles triangles (e.g., resulting from decalinear angles). Similarly, angles that permit perpendicular line crossings (e.g., octolinear) might result in maps which are preferred to ones with angles that do not permit this (e.g., decalinear again). User-acceptance is an important aspect of schematic maps. An otherwise impeccable design will fail if it nonetheless suffers from widespread rejection.

Topographicity.

In order to optimize a design according to the above criteria, distortion is inevitable to at least some extent. A schematic with poor topographicity is one in which the distortion is sufficiently extreme that it adversely affects user-confidence in the design – as a result of significant conflicts with mental models of a city – and/or leads to the planning of inefficient or inappropriate journeys. The effects of topographical distortion on usability are simple to investigate but reported rarely. One noteworthy study is by Guo (2011) who looked at actual journeys taken on the London Underground, and found that for one region of the map, with poor topographicity, inefficient journeys were taken 30% of the time. There are considerable individual differences in reported tolerance of poor topographicity, but this does not seem to be an issue for many users (Forrest, 2014). Likewise, Vertesi (2008) found many users were not sensitive to distortions on the London Underground map, to the extent that their mental models of London reflected these. For a massive disorganized city such as London, a distorted mental model may be better than no mental model at all.

Once the criteria for effective design can be specified, in theory it is possible to measure success at achieving them, and score maps according to how well these have been attained, giving a prediction of usability. Quantifying simplicity of line trajectories is straightforward for linear maps, likewise the balance of a design in terms of density gradients. Overall coherence is harder to quantify, requiring higher-order measures of relatedness between lines. Topographical distortion can be scored by the size of deviation of station locations on the map from reality, but this is made complicated by the fact that topographicity is more important in some locations (the centre) than others (outlying suburbs, where deviations may be particularly large and yet of little consequence). A further problem comes from attempting to prioritize between these disparate and possibly conflicting measures, necessitating the creation of matched sets of experimental maps and large-scale usability testing.

1.2. The dissociation between objective measures and subjective ratings of usability

Until it is possible to score various designs according to objective criteria, usability testing of prototype maps will remain a valuable but time consuming exercise. For example, Roberts & Rose (2015) tested a number of London Docklands Light Railway prototype route diagrams, using a variety of station finding and journey planning tasks, with responses collected via touch-screens. The study was successful, in that designs were identified that were associated with larger numbers of errors, or else the planning of inefficient journeys, and the map that formed the basis of the production version had none of these difficulties. However, in order to identify this, it was necessary to test 240 subjects, all for a network comprising fewer than 50 stations. Under such circumstances, it might be tempting to fall back on user ratings of designs, and choices between different versions, or even present a selection of prototypes to the general public and encourage them to vote (Boston Globe, 2013). However, the problem with this, as demonstrated by Roberts et al. (2013) is that objective measures of usability, versus subjective ratings of usability, appear to be uncorrelated. For example, Roberts et al. (2013) measured map effectiveness by the mean time necessary to plan a cross-Paris journey, then asked subjects to choose between the design that they had experienced, versus a previously-unseen alternative, and compared the people who elected to remain with the original, versus those who rejected it. There was no difference in mean journey planning time between people who kept versus rejected the original, although there was a difference in terms of a questionnaire rating score. People who rejected maps had previously given them adverse scores on a statement rating task.

The dissociation between objective and subjective measures of performance is entirely in line with findings in psychology. Such metacognitive failures are commonplace (e.g., Chabris & Simons, 2010; Kruger & Dunning, 1999). One difficulty is that there are limited cues to whether performance is successful or not, so that provided this is basically competent, then subtle indicators, such as a tendency for a persistent difference between tasks of a few seconds, will be insufficiently salient, especially as effective self-monitoring also requires cognitive resources (e.g., Dierckx and Vandierendonck, 2005). Without salient performance cues, and given that most users will not be experts at visual information design, it is inevitable that maps will be evaluated according to their superficial surface properties (e.g., is there significant topographical distortion?) rather than the more subtle aspects of design that contribute more directly to usability. This tendency is analogous to findings in expert-novice physics problem solving, for example, Chi, Feltovich and Glaser (1981). Given the prevalence and familiarity of octolinearity in schematic map design, as opposed to alternatives, it is also possible that any stimulus set that contains one such map versus others with different design rules will result in a bias in preference towards the octolinear version, even if this is the less effective (see Figure 2). Hence, expectations and prejudices, whether implicit or explicit, may bias subjective evaluations.

There are many examples of familiarity leading to positive evaluations in psychology, the most well-known of these is probably the mere-exposure effect (e.g., Bornstein, 1989). For this, repeated exposure leads to increased liking for relevant stimuli but, strictly, this phenomenon applies only to briefly presented items. More generally, familiarity-liking effects have been discussed as part of a fluency framework by Alter and Oppenheimer (2009), in which factors that lead to perceived easier cognitive processing for certain items, versus others, are associated with their positive evaluation. This could be applicable to octolinear schematic maps in the context of their greater familiarity leading to perceived less effortful interpretation, although this would also be expected to manifest itself in terms of superior objective measures of usability. Of course, an easy-to-process map is the goal of every designer, and user acceptance is an important aspiration, but if a bias towards certain design rules is entirely familiarity-based, without any basis in objective usability measures, then this might prevent the adoption of unconventional designs that are nonetheless more effective than conventional versions.

2. EXPERIMENT ONE

One difficulty with Roberts et al. (2013) is that subjects only experienced one single map, which they then rated via a questionnaire, and were then requested to make a decision to retain this, or switch to a previously unseen one. In the absence of direct experience at journey planning using both designs, this is a prime opportunity for expectations and prejudices to determine choice, rather than self-observations of performance. This does not invalidate the findings of Roberts et al. (2013), which show that there is a lack of rational choice at the group level. However, this does suggest that the next step should be to investigate the relationship between objective and subjective measures of performance at the individual level, to see whether expectations and prejudices are still sufficiently powerful to dominate map choices. Hence, when all candidate designs have been experienced and an unconventional design is associated with faster planning performance, will subjective ratings of map usability and objective measures of performance now become correlated.

Experiment 1 repeats the basic methodology of Roberts et al. (2013). The same two Paris Metro maps were tested. The official RATP octolinear design is poorly optimized, with complex zig-zagging lines. Hence, it fails the simplicity of line trajectories criterion in the framework for effective design described earlier. The consequence of this is that the structure of the network is difficult to identify. Roberts et al. (2013) argued that the dense interconnected network of Paris lines made them particularly unsuited to octolinear schematization. The curvilinear map was devised manually by the first author, using a vector graphics package. The principles underlying this are that if a conventional octolinear schematic cannot be created with simple line trajectories, then a non-linear design may be preferable (see Figure 3). Hence, instead of numerous short zigzagging straight-line segments, smooth curves should be used. Rather than changes in direction being minimized, as on a conventional schematic, changes in curvature are minimized instead. This translates into using Bézier curves with the following optimization criteria: S-bends and other points of inflection must be avoided where possible and, for an individual Metro line, the aim should be to have the smallest number of control points necessary in order to maintain interchanges and to ensure sufficient space for station names. Also, all control points should be tangents, with no cusps permitted. For example, there are sixteen changes of direction on Line 4 of the RATP map compared with eight intermediate control points on the curvilinear version. Specific to the Paris Metro design here, it was intended to be as topographically accurate as the official RATP version but with the trajectories of all lines smoothed, and attention paid to the orbital lines (2 and 6) which together form a loop within Paris. These criteria served to simplify the design and to emphasize the underlying structure of the network, reducing the expected cognitive load of using it in comparison (see Roberts et al., 2013, for a full discussion of the cognitive factors associated with schematic map interpretation and use).

Roberts et al. (2013) used only between-subjects designs, so that each person planned Paris Metro journeys using just one map. For Experiment 1 here, each subject planned eight complex cross-Paris journeys using each of two maps, first one, then the other; four journeys for each map. The within-subjects design obviously runs the risk of item effects, and also cross-talk between maps resulting in order effects, but these can be easily analysed. A similar statement rating questionnaire to Roberts et al. (2013) was also administered after the journey planning task, this time requesting separate scores for each of the two maps. Subjects were then finally asked to make a choice between the two designs. As before, mean planning time per journey was the main objective measure of performance, with an estimation of likely journey durations as an additional measure to check that faster performance at one design was not associated with less efficient journeys. The main measures of subjective ratings were an aggregate score calculated from the statement rating task, and the final map choice. With subjects directly experiencing both designs, it is now directly possible to identify the best version for each individual, relatively, from an objective measure, and determine whether this is reflected in relative scores at the rating task, and final choice of map.

2.1. Method

2.1.1. Subjects

Eighty subjects took part in this experiment, 28 males and 52 females with a mean age of 22.0 years (SD 4.9). 48 were students at the University of Essex who received £5.00 for taking part in the study and 32 were unpaid volunteers. All were residents of Essex or adjacent counties, the majority were unfamiliar with the Paris Metro network.

2.1.2. Materials

The maps for journey planning were matched for surface area and printed to fit an A3 sheet and laminated. Two sets of journeys were assembled, taken from Roberts et al. (2013), and shown in Table 1. All were intended to be difficult to plan, with distant start and destination stations and many alternative options. Two interchanges were always required to complete each one. Each journey was shown on an individual laminated sheet with the map greyed out except for the start (arrowed) and destination stations.

A 19-item questionnaire was used, based on Roberts et al. (2013) – see Table 2. For each question, an answer or a decision was required for both of the maps. The questionnaire consisted primarily of a statement rating task (questions 1 to 15), in which seven-point response scales offered options ranging from strongly agree through to neutral and to strongly disagree for a series of statements. Questions 16 and 17 requested brief sentences for each map, i.e. qualitative responses. Question 18 requested people to compare the two maps, expressing strength of preference for them, ranging from: Curves Map strongly preferred; Curves Map preferred; Curves Map slightly preferred; Neutral: neither map preferred; Straight Line Map slightly preferred; Straight Line Map preferred; Straight Line Map strongly preferred. This decision will be referred to subsequently as the map choice rating. Question 19 gave a range of frequency options: every day; a few days every week; a few times every month; about once a month; a few times every year; once a year or less; never/not for years

2.1.3. Design

All subjects planned journeys using both maps, four journeys for one design, then four journeys for the other. Thirty-nine subjects planned journeys using the curvilinear map first, the remaining 41 received the official map first. Primarily, this was a within-subjects design with Map Type (two levels, Official versus Curvilinear) as the independent variable. Measures of map performance included the time taken to plan a journey, number of declared journeys which contained at least one error, and an estimation of the duration that the planned journeys would have taken had they been implemented. Questionnaire data provided a means of measuring people's subjective assessments of map usability.

Each subject was asked to plan journeys using one of the two sets of items for each map. For example, Set A with the official map, and Set B with the curvilinear version. Presentation of journey set and the order of presentation of maps was counterbalanced. The order in which the maps were presented was also analysed to ensure that one order permutation did not result in different patterns of performance to the other.

2.1.4. Procedure

Subjects were tested individually at a desk or table in quiet surroundings. They were informed that they would be asked to plan a series of Paris journeys using the supplied maps. They were to assume that the network was fully operational and that there were no cost considerations. They were given no guidance as to journey criteria or priorities, it was simply stated that they should devise the journey that they would choose if they were actually to undertake it in real life. They were also informed that they should only change between lines at designated interchanges shown on the map.

Subjects were given the opportunity to view the first map while the initial instructions were given. The appropriate practice journey for the item set was administered. There then followed the four test journeys presented in a random order. Each trial commenced with the experimenter placing the journey sheet indicating start and end stations above the A3 laminated map, and timing immediately commenced using a stop-watch. The subject was asked to plan the journey as requested, using a dry-wipe marker. Once satisfied with the plan, a verbal announcement was made by the subject, timing stopped, and the final chosen route was transcribed onto an A4 paper map by the experimenter. Following this, the experimenter cleaned all marks from the laminated map, or replaced it with a fresh one, and the next trial commenced. Once all four journeys were completed, the process was repeated with the second map, including the appropriate practice trial. When all journeys had been planned, subjects then completed the questionnaire with both maps in sight.

2.2. Results

2.2.1. Objective usability measures

All correlations reported throughout this paper are Pearson r values. In general, there were no significant effects or interactions that involved sex differences for any of the objective or subjective measures of performance, either in conjunction with the map type tested or the choice of map. Likewise there were no significant correlations between frequency of rail transport usage and any of the measures. The restricted age range for this experiment precluded identifying any substantial relationships between this variable and others.

For each subject, for each map, means of the four planning times were calculated to give overall journey planning times for each design. Each subject had a pair of scores for each variable, enabling relative performance between maps to be investigated. One consequence of this, despite the high variability in performance, is that there is no requirement for data transformations in order to clarify effects, unlike Roberts et al. (2013).

For each journey, its duration was estimated by allowing two minutes per station and ten minutes per interchange. This is comparable with the heuristics that passengers themselves use (e.g., Vertesi, 2008) and ignores the variable interchange quality within most metro networks, which is virtually impossible to communicate via maps. The value of ten minutes is a worst-case scenario estimation, should an unknown interchange prove to be particularly long, and/or a train is just missed (see Guo & Wilson, 2011). None of the final declared journeys contained errors, and so this is not reported. Mean performance by map is shown in Table 3.

There were no problematic order or item effects for journey planning times. A three-factor mixed design Analysis of Variance (ANOVA) was used to test for these, with Map Type (within-subjects; Official versus Curvilinear), Presentation Order (between-subjects; Curvilinear Map first versus Curvilinear Map Second), and Item Set (between-subjects; Curvilinear Map Item Set A versus Curvilinear Map Item Set B) as the factors. The only significant effects were the main effect of Map Type, F(1,76) = 38.2, MSe = 101.7, $p < .01$, and there was also a significant Map Type x Presentation Order interaction, F(1,76) = 5.96, MSe = 101.7, $p < .05$. The significant main effect indicated longer planning times for the official map than the curvilinear map, replicating the major finding of Roberts et al. (2013). The significant interaction indicated a greater advantage for the curvilinear map when it was presented second than when it was presented first: Using pairwise comparisons, F(1,39) = 7.94, MSe = 89.4, $p < .01$, and F(1,39) = 33.1, MSe = 114.3, $p < .01$, respectively. However, this interaction is merely an artefact of the ANOVA blocking, and indicates a transfer effect, virtually identical in size for both designs. For the curvilinear map presented first, mean planning time was 42.4 seconds, compared with 37.9 seconds when it was presented after the official map. For the official map presented first, mean planning time was 51.7 seconds, compared with 48.4 seconds when presented after the curvilinear map. Overall, the curvilinear map continued to outperform the official design, with 63 out of 80 subjects having mean planning times that favoured the curvilinear version. No other effects from the three factor ANOVA were significant, next greatest F(1,76) = 2.08, MSe = 101.7, $p > .05$ for the Map Type x Item Set interaction. The negligible difference in mean estimated journey duration between maps was not statistically significant, F(1,79) = 0.04, MSe = 60.3, $p > .05$. Hence, the faster planning times for the curvilinear map were not associated with poorly-planned journeys of a longer estimated duration.

2.2.2. Subjective usability measures

Subjects were given the opportunity to express opinions on the maps in a variety of different ways via the questionnaire. The most straightforward was asking them to simply to rate which of the two maps they would prefer to use in a real setting, the map choice. Overall, 38 subjects expressed a preference for the official map (8 strong, 15 medium, 15 slight) and 34 for curvilinear (10 strong, 17 medium, 7 slight). If anything, selection of the curvilinear map was slightly less frequent for the current experiment compared with Roberts et al. (2013) where, across all three experiments, 83 subjects expressed a preference for this and 69 for the official map (in a forced-choice task).

For the statement rating task, an aggregate score was created using the 11 questions that are directly relevant to usability issues (see Table 2). Scores on certain questions were reversed to reflect their polarity, such that high scores indicated a positive assessment of a map for each statement. This gave an aggregate statement rating task score for each individual for each map, in which a score of 77 indicated the highest possible rating of a design, 11 the worst possible, and 44 a neutral rating. Table 3 shows that the curvilinear map was given a slightly higher mean aggregate rating than the official design, but the difference was non-significant, using a single-factor within-subjects ANOVA, $F(1,79) = 2.66$, $MSe = 131.8$, $p > .05$.

Map choice score, and aggregate rating scores, were strongly correlated, so that if the seven-point scale for map choice is scored such that a high value indicates preference for the curvilinear map, the correlation with the aggregate rating score for the official map is –.44, and with the aggregate rating score for the curvilinear map is .64, and with the difference in aggregate rating scores (curvilinear minus official) the correlation is .72, all $p < .01$.

2.2.3. The relationship between subjective and objective measures

The curvilinear design has proven easier to use from the point of view of journey planning performance, both overall, and for 63 of the 80 individuals. However, this superior performance is not reflected in either map choice, or aggregate ratings of map usability: Both subjective measures indicate that the designs were rated roughly equally. Taking a closer look at the data, it is clear that, despite directly experiencing both designs, the primary objective measure of performance (mean journey planning time) is not influencing subjective evaluations of the maps in any way.

In general, if people are categorized by their map choice rating, ignoring strength of this, and excluding those people who were undecided (N = 8), and if a measure of performance is related to map choice in a rational way, then using a two-factor mixed design ANOVA, there should be a two-way interaction between Map Choice (between-subjects) and Test Map (within-subjects) for that variable. For example, if journey planning time is related to choice, then the people who chose the official map should have a faster journey planning time when tested with the official map compared with the curvilinear map, and the people who chose the curvilinear map should have a slower journey planning time when tested with the official map compared with the curvilinear map. Table 3 shows mean planning times, estimated journey durations, and aggregate ratings, split by map choice.

For mean journey planning times, there was no trace of any such interaction, F(1, 70) = 0.14, MSe = 109.2, p > .05. Hence, there was no association between planning time and map choice. Even if people were aware of their relative performance for the two maps, this was not influencing their map choice in any way. However, looking at mean estimated journey durations, the small interaction approached significance, F(1, 70) = 3.85, MSe = 52.8, p =

.054. The direction of the effect indicated that the map chosen tended to be the one of the pair that was associated with shorter estimated journey durations. In contrast, the interaction for the mean aggregate questionnaire scores was highly significant, F(1, 70) = 63.2, MSe = 76.0, p < .01. Map choice may not be related to objective performance in any substantial way or, indeed, any way at all, but is nonetheless not arbitrary, being strongly related to opinions that subjects have about map usability. All effects remained identical when excluding the 22 people who rated themselves as having only a slight preference for one design over the other. Other analyses also failed to identify links between objective and subjective measures of performance. For example, the correlation between aggregate questionnaire rating score for the official map, and mean planning time for the official map was non-significant, r = –.16,

p > .05, likewise for the curvilinear map, r = –.14, p > .05. The correlations between mean estimated journey duration and aggregate questionnaire rating score were likewise non-significant (r = –.17, p > .05, for the official map, r = –.07, p > .05, for the curvilinear map), although all correlations were in the expected direction; poor objective performance associated with low questionnaire rating scores.

2.3. Discussion

The findings in the current experiment replicate and extend those of Roberts et al. (2013). Hence, they continue to show that a curvilinear design Paris Metro map outperforms the conventionally designed octolinear version of the type that is used officially. From a theoretical point of view, this again irrevocably refutes the octolinearity as a gold standard conjecture, i.e. that octolinearity will always yield the most effective design of schematic metro map, compared with other design rules.

The findings also continue to show that there is little or no relationship between objective measures of performance (journey planning time, estimated journey duration) and subjective ratings of usability (map choice, aggregate questionnaire rating score). This is despite all subjects planning journeys using both designs, and this strengthens the finding, originally identified by Roberts et al., (2013), of little or no association between objective and subjective measures. This finding is now shown to apply at the individual level as well as the group level. Also replicated is the substantial interaction between Map Type and Map Choice when looking at the aggregate score of the statement rating task, so that rating score is highly associated with map choice. It is clear that people have strong opinions on the differential usability of the designs. These beliefs were either already present before testing, or developed during testing. Given the superior journey planning time of the curvilinear map, it would be surprising if such opinions concerning usability had developed during testing, but the qualitative statements requested in the questionnaires shed little light on the matter. Few of these comments directly mentioned familiarity or lack of familiarity of design rules, mainly referring to the clarity and identifiability of stations, lines, and interchanges, with considerable conflict in opinions between different subjects for the two maps. The next experiment that is reported therefore sought to identify preconceptions directly, so that subjects were asked to rate a systematically designed set of maps directly for usability and attractiveness.

3. EXPERIMENT 2

When investigating the usability of, and opinions on, different designs of schematic maps, interpretations of findings are made difficult if versions are paired that differ in design priorities as well as design rules. The Paris Metro maps in Experiment 1 are an excellent example of this. The curvilinear map was designed to be optimised in terms of simple line trajectories. The official octolinear map was not obviously designed primarily to satisfy this criterion. Hence, although Experiment 1 demonstrates a dissociation between an objective measure and subjective ratings of design effectiveness, and refutes the strong version of the octolinearity as a gold standard conjecture, it does not provide highly generalizable findings concerning the utility of, or even popularity of, octolinear versus curvilinear designs.

The reasoning behind Experiment 2 is that there exist intuitive theories of design. This concept borrows from work, for example, on intuitive or folk physics (e.g., Gelman and Legare, 2011; McCloskey, 1983). Here, people’s predictions of physical phenomena can often be shown to be mistaken, and the patterns of errors can be used to identify the supposed defective beliefs or theories that led to them. In a similar way, it is suggested that an individual's prediction about the relative usability of maps, in the absence of data, must entail a theory of effective design, whether implicit or explicit. Hence, patterns of usability ratings for different maps can be used to infer people's beliefs about the criteria necessary for usability (Roberts, 2014b). For example, universally high usability ratings for all octolinear maps (irrespective of design priorities) would indicate a belief in the strong version of the octolinearity as a gold standard conjecture, whereas merely relatively high ratings for octolinear designs would indicate a belief in a weaker version of this. Such theories are important because they will affect the procurement, production and reception of new designs. For example, an unusual, effective design that conflicts with expectations could be judged harshly and rejected, or a poorly-optimized design that conforms with expectations could be accepted with little criticism.

The use of just two maps would fail to capture the diversity of theories possible, and so subjects were provided with nine maps, designed manually by the first author, which varied according to two criteria: design rules (octolinear, multilinear, curvilinear) and design priorities (optimised for simple line trajectories, optimised for topographical accuracy, and optimised for neither – See Figure 4). This was an internet-based study in which these maps were individually rated by two criteria, their perceived usability, and their attractiveness: One hypothesis for the source of people's usability ratings is that they may simply believe that designs which they find the most attractive are the most usable. Hence, both types of rating were requested to determine whether usability and aesthetic judgements are separable.

The London Underground was chosen as the test-bed for the nine designs because of the complexity of the network and the iconic status of the official octolinear design amongst British people (BBC, 1987, 2006). International participation in this survey was anticipated, enabling cross-cultural comparisons to take place, for example investigating respondents from the USA (less familiarity with the London Underground, less frequent use of octolinear rail maps compared with the UK) and Germany (less familiarity with the London Underground, equally frequent use of octolinear rail maps compared with the UK). As well as octolinear and curvilinear designs, multilinear designs were included (see Figure 2 and Roberts, 2012). The latter allows straight lines at any angle, joined by tightly-radiused corners, thus permitting the simplest line trajectories, but potentially at the expense of the coherence of the design.

Three different design priorities were applied in order to give a matrix of nine maps: (1) geographical designs were intended to be spatially informative, created by applying a pinch distortion to a topographical map of London and attempting to preserve the resulting station locations while achieving reasonably simple line trajectories; (2) compact designs were named so as to disguise their true status: as deliberately poorly optimised designs with complex line trajectories and little attempt at topographical accuracy; and (3) stylised designs were also named to avoid revealing their true status: as carefully optimised versions with the simplest line trajectories feasible in the available space, and particular attention given to the centre of London inside the Circle Line. Because of London's somewhat chaotic rail network, it was inevitable that the geographical maps would have complex line trajectories with all three design rules. They were included to identify the extent to which people expected spatial fidelity for schematic maps: Frequent complaints are made against designs which fail to achieve this (e.g., Lloyd, 2012). The compact maps, with few redeeming features, should be rated as easy to use only by people who believe that the simplicity of line trajectories is irrelevant to usability. The stylised maps were included to identify people who were sensitive to the simplicity criterion. Furthermore, differing ratings along the design rules dimension (octolinear, curvilinear, multilinear) would indicate any expectations or prejudices concerning this aspect in relation to usability.

3.1. Method

3.1.1. Subjects

Subjects were 649 respondents recruited via a map-related newsletter and recommendations via various map-related websites. 541 were male and 108 were female. The mean age was 36.2 years, SD 14.3. Of the various countries represented, 290 were from the UK, 118 were from the USA/Canada, and 116 were from Germany. The next-most represented countries were Australia and the Netherlands (17 each). Because of the recruitment methods used, the sample cannot be said to be representative of the general population. However, the sample will nonetheless enable preferences for people from different countries to be identified, and provide a useful baseline against which future studies can be compared. Subjects were explicitly asked whether they were familiar with the work of the first author, and whether they worked in a design profession.

3.1.3. Materials

In addition to the maps, whose design rules and design priorities were discussed in Section 3, during the pre-map rating phase of the experiment subjects were asked to give their sex, age and country of residence, and the following questions were also asked:

Have you ever lived in London for more than a month? [yes/no]

Would you say that you are familiar with the London Underground? [yes/no]

Are there any other metro/subway/underground

networks that you are familiar with in the world? [yes/no]

If yes, please list them [text box for responses]

Have you ever read any books or articles on map design by Maxwell Roberts? [yes/no]

Does your profession involve graphic design,

information design, or public transport information? [yes/no]

3.1.3. Design

The two primary independent variables, each with three levels, were Design Rules (Octolinear, Multilinear, Curvilinear) and Design Priorities (Stylised, Geographical, Compact). For the Design Priorities variable, within each level of Design Rules, the stylised map had the simplest line trajectories, then the geographical map, and then the compact map. The questions in the survey enabled other variables to be investigated, including, age, sex, and country of residence. The dependent variables were the attractiveness and usability ratings for each individual design, on a three point scale: easy to use / neutral / hard to use and attractive / neutral / unpleasant.

3.1.4. Procedure

The survey was designed to be displayed on a web browser as a sequence of screens. The first screen gave a preamble concerning the nature of the task, and its likely duration. The next asked for basic demographic data along with some more detailed questions (see section 3.1.2). The third screen introduced the map usability rating task. Subjects were informed that:

You are about to see a table with nine different London Underground maps. Your task is to look at these and identify any designs that in your opinion might be particularly easy to use for planning a journey, or particularly difficult.

Please note

1. Click on each thumbnail to pop up a large version of the map in a new window so that you can view it. Click on the map again to close the window.
2. You may choose to designate as many maps as you like, in any combination, as easy or difficult to use.
3. For maps that do not appear to be particularly easy to use or difficult to use, leave them set to neutral.
4. Three of the maps are intended to be geographically accurate, you will see which ones from their names. The remaining maps are not intended to be geographically accurate.

5. There are no time limits, you can change your mind as often as you like, until you press the 'click when finished' button.
6. If you wish to stop, just press the 'quit and delete my data' button and your data will be deleted. You will still be able to try the survey again another time if you wish.

The fourth screen showed nine map thumbnails, with each map named and paired with three radio buttons corresponding to easy to use, neutral, and hard to use options. Buttons were initially set to the neutral option. Clicking on a map thumbnail expanded it for viewing, clicking on the expanded map returned to the thumbnail screen. There was also an option to view the instructions again. Once subjects were satisfied with their choices, they pressed a button to advance to the fifth screen, which warned that For Task 2, I will ask you to rate the maps in a different way, please read the next set of instructions carefully. Advancing to the sixth screen, the preamble to the attractiveness rating task was as follows:

For the next task, you will see the same nine maps again. This time, I would like you to identify whether you think that any of these designs are particularly attractive or unpleasant to look at, this is not necessarily the same as whether you think that they are easy or difficult to use, but could be.

There numbered points to note were equivalent to those for the usability rating task. Advancing to the seventh screen, the rating task was formatted similarly to before, except that the three buttons for each map offered the options of attractive / neutral / unattractive. Once decisions had been finalized, the last screen thanked subjects for taking part, and offered a contact email address for further queries.

3.2. Results

3.2.1. Preliminary analysis

Rating scores were scaled such that if all respondents rated a map as easy to use, then its mean score would be 100%, if all respondents rated a map as neutral, then its mean score would be 50%, and if all respondents rated a map as hard to use, then its mean score would be 0%. Overall ratings by map are shown in Table 4.

3.2.2. Analysis of basic usability and attractiveness ratings

Ratings of attractiveness and usability were each analysed using two-factor fully-within-subjects ANOVA, in which each factor had three levels: Design Rules (Octolinear, Multilinear, Curvilinear) and Design Priorities (stylised, geographical, compact). For usability ratings, there were main effects of Design Rules and Design Priorities, $F(2,1296) = 308.6$, $MSe = 2042.5$, $p < .01$, and $F(2,1296) = 284.3$, $MSe = 1468.3$, $p < .01$, respectively. Pairwise comparisons showed that the mean rating for the octolinear maps (64.2) significantly exceeded that for the multilinear maps (38.7), F(1,1296) = 311.2, MSe = 2042.5, p < .01, and that the mean rating for the multilinear maps significantly exceeded that for the curvilinear maps (29.5), F(1,1296) = 40.06, MSe = 2042.5, p < .01. Similarly, the mean rating for the stylised maps (60.6) significantly exceeded that for the geographical maps (39.3), F(1,1296) = 299.8, MSe = 1468.3, p < .01, and that the mean rating for the geographical maps significantly exceeded that for the compact maps (32.5), F(1,1296) = 30.7, MSe = 1468.3, p < .01. The two-way interaction was also significant, indicating a wider range of mean usability ratings for the three different stylised maps than amongst the three compact or the three geographical maps, F(4,2592) = 41.1, MSe = 742.8, p < .01.

For the attractiveness ratings, there were also main effects of Design Rules and Design Priorities, F(2,1296) = 267.4, MSe = 2177.5, p < .01, and F(2,1296) = 228.1, MSe = 1429.6,

p < .01, respectively. Pairwise comparisons showed that the mean rating for the octolinear maps (62.2) significantly exceeded that for the curvilinear maps (46.0), F(1,1296) = 117.4, MSe = 2177.5, p < .01, and that the mean rating for the curvilinear maps significantly exceeded that for the multilinear maps (27.6), F(1,1296) = 150.7, MSe = 2177.5, p < .01. Similarly, the mean rating for the stylised maps (60.1) significantly exceeded that for the geographical maps (38.9), F(1,1296) = 308.6, MSe = 1429.6, p < .01, but the mean rating for the geographical maps did not significantly exceed that for the compact maps (36.7),

F(1,1296) = 3.02, MSe = 1429.6, p > .05. The two-way interaction was also significant, indicating an out-of-step mean rating for the curvilinear compact map, slightly higher than expected in comparison with the other means, F(4,2592) = 39.7, MSe = 781.5, p < .01.

Finally, a comparison between attractiveness and usability ratings was undertaken using a three-factor fully-within-subjects ANOVA. There were two significant effects for the Rating Type factor. The Rating Type x Design Rules interaction, F(2,1296) = 169.0, MSe = 1135.0,

p < .01, indicated that for the multilinear maps, usability ratings significantly exceeded attractiveness ratings, F(1,1296) = 105.1, MSe = 1135.0, p < .01, and for the curvilinear maps, attractiveness ratings significantly exceeded usability ratings, F(1,1296) = 232.4, MSe = 1135.0, p < .01. For the octolinear maps, there was no significant difference, F(1,1296) = 3.62, MSe = 1135.0, p > .05, although the direction of the effect was the same as for the multilinear maps. The significant Rating Type x Design Priorities interaction,

F(2,1296) = 9.61, MSe = 741.5, p < .01, merely indicated that geographical and stylised maps had roughly equal usability and attractiveness ratings, but the compact map was rated as being more attractive than usable. The three-way interaction was not significant, F(4,2592) = 2.24, MSe = 584.9, p > .05.

3.2.3. Analysis of usability and attractiveness ratings in relation to other factors

In addition to the basic usability and attractiveness measures, further analyses investigated whether there were effects in conjunction with the other categorical variables, such as sex differences, or country of residence, using three-factor mixed design Analyses of Variance (Group x Design Priorities x Design Rules). For factorial designs with relatively large sample sizes, one difficulty with analyses is the prevalence of significant higher-order interactions, which may be difficult to interpret and of little theoretical interest. The most important interactions in this sense would be those in which there was a consistent/persistent reversal of overall effects, for example greater usability ratings for multilinear maps versus curvilinear maps under only certain circumstances. Looking at usability and attractiveness ratings: (1) for professionals in disciplines related to map design, versus non professionals; (2) people who had lived in London versus had not lived in London; (3) people who were familiar with London, versus unfamiliar; and (4) people who were familiar with urban rail transport in general, versus unfamiliar, there were only two significant relevant effects: a London Resident x Design Rules interaction for attractiveness ratings, F(2,1294) = 3.41, MSe = 2169.4, p < .05, and a London Familiarity x Design Rules x Design Priorities interaction for usability ratings, F(4,2588) = 4.49, MSe = 738.8, p < .01. Neither of these implied any qualification to the rank ordering of the basic effects.

For sex differences, in general females rated the maps as more usable and attractive than males rated them: For attractiveness (mean = 49.4 versus 44.4), F(1,647) = 11.2, MSe = 1783.7, p < .01, and for usability (mean = 48.2 versus 43.3), F(1,647) = 10.0, MSe = 1916.0,

p < .01. The only other significant effect was the three-way interaction for usability ratings,

F(4,2588) = 2.68, MSe = 740.8, p < .05. Again, this was not associated with a change in rank ordering for the basic effects of Design Rules and Design Priorities on usability and attractiveness ratings.

Three of the regions were sufficiently represented to see whether there were differences in ratings between them (UK, N = 290; USA/Canada, N = 118; Germany), N = 116). Despite the different design traditions of the countries concerned, there was virtually no evidence for any effects of this. For attractiveness ratings, a marginally significant Region x Design Rules interaction, F(4,1042) = 2.35, MSe = 2161.1, p = .05, in conjunction with the significant three-way interaction, F(8,2084) = 2.39, MSe = 751.3, p < .05, suggested a tendency for people from Germany to give higher attractiveness ratings to curvilinear maps than people from the

USA/Canada. However, a reverse preference for octolinear and multilinear maps by people from the USA/Canada compared with people from Germany was only clearly present for the stylised designs. The only significant effect for usability was the three-way interaction, $F(8,2084) = 2.77$, $MSe = 745.2$, $p < .01$. Again, this was not associated with any clear systematic change in rank ordering for the basic effects of Design Rules and Design Priorities.

The final factor concerns whether or not respondents were familiar with research on map design by the first author. This might be conceived as a measure of familiarity with, and interest in, issues concerned with schematic map design. Looking at the attractiveness measures, the significant three-way interaction, $F(4,2588) = 2.62$, $MSe = 779.6$, $p < .05$, together with a significant Familiarity x Design Priorities interaction, $F(2,1294) = 7.80$, $MSe = 1414.8$, $p < .01$, indicated that people familiar with design issues rated stylised maps as slightly more attractive than people who were not familiar. People who were familiar with design issues rated compact and geographical maps (octolinear and multilinear only) as less attractive than people who were not familiar (with no trend for curvilinear maps). For usability ratings, a significant Familiarity x Design Priorities interaction, $F(2,1294) = 7.09$, $MSe = 1454.7$, $p < .01$, indicated that people familiar with design issues gave more favourable ratings to stylised maps, and less favourable ratings to compact and geographical maps compared with people who were not familiar. A significant Familiarity x Design Rules interaction, $F(2,1294) = 15.0$, $MSe = 1999.4$, $p < .01$, indicated that people familiar with design issues gave more favourable ratings to curvilinear maps, and less favourable ratings to octolinear and multilinear maps, compared with people who were not familiar. Overall though, with familiarity as a factor, there was again no clear systematic change in rank ordering for the basic effects of Design Rules and Design Priorities.

3.2.4. Identifying theories of design

When looking at the patterns of usability ratings for individuals, it is possible to identify whether these are systematic enough for any of the respondents to be designated as holding a clear theory of schematic map design. For example, we can identify 61 people with a strong octolinearity theory, because every single octolinear map is given a higher usability rating than every other version. An additional 33 had a weak octolinearity theory in which the octolinear design was always given the highest rating within a design priority (e.g., received the highest rating of the three compact maps, but might have received a lower rating than any of the stylised designs). Only 57 respondents held a simplicity theory in which, within a design rule, the stylised map (with the simplest line trajectories) was always given the highest rating. Three of these respondents simultaneously held a simplicity theory and an octolinearity theory. In addition, a linearity theory could be identified in 41 respondents (9 of these also displayed an octolinearity theory) in which the curvilinear design was always given the lowest rating within a design priority. Only 5 respondents had a topographical theory in which the geographical maps received the highest rating within design rules. Overall, a clear 'theory' of design could be identified in 185 of the 649 respondents. or 29% of them.

There was little evidence for substantial systematic differences in the prevalence of theories between different respondent groups. For example, combining transport/design professionals with people who were familiar with research by the first author, 14% of them displayed an octolinearity theory, and 11% a simplicity theory. For people outside of these categories, 15% displayed an octolinearity theory, and 7% a simplicity theory.

3.3. Discussion

Overall, looking at the rating tasks across subjects, there is a clear preference for octolinear designs, which always received by far the highest usability and attractiveness ratings within each individual design priority category. It is also possible to identify clear octolinearity theories for a substantial number of individuals, from the patterns of responses. Surprisingly, there was no evidence for regional differences in this, despite the different traditions for schematic map design. Overall, this preference for octolinearity is compatible with the findings in Experiment 1, where the Paris Metro curvilinear map was associated with faster journey planning times for over 75% of individuals but was selected in preference to the octolinear design by less than 50% of them.

In Experiment 2, the multilinear and curvilinear designs were both rated less highly than equivalent octolinear versions, both for attractiveness and usability, but there was also an interesting double-dissociation, such that multilinear maps were rated more highly for usability than attractiveness, and curvilinear maps were rated more highly for attractiveness than usability. This suggests that there may be at least two factors present that determine individuals' responses to schematic maps.

For the Design Priority effects, there was a clear progressive effect of line trajectory simplicity, both for usability and attractiveness measures, with simpler trajectories receiving the best scores. At the very least, there is a statistical implicit belief across respondents that simpler line trajectories are associated with greater usability. There was no evidence that flagging a map (correctly) as geographically accurate boosted its ratings in any way, suggesting that from the point of view of line configuration versus displaying correct station proximities, there is a general belief that the advantages of a topographical configuration (spatial informativeness) are outweighed by the disadvantages of complex line trajectories.

Looking at the various categorization variables for the respondents, very few of these qualified the above findings in any way. There was little evidence for regional differences, other than a tendency by people from Germany to rate curvilinear maps more highly for attractiveness than people from the USA/Canada. The converse, higher ratings of multilinear and octolinear maps by people from the USA/Canada, compared with people from Germany, only applied to stylised maps, i.e. the ones with the simplest line trajectories.

The only other categorization with any clear effects was familiarity (versus lack of) with schematic map design issues, with such people more likely to give higher usability and attractiveness ratings to the maps with the simplest line trajectories, and less likely to give high ratings to octolinear maps and low ratings to curvilinear maps. Even so, such effects were generally very small in comparison with the overall effects of Design Priorities and Design Rules on usability and attractiveness ratings.

Having collected usability ratings for the nine maps, the obvious next step is to collect actual usability data in order to identify the extent to which rated usability and measured usability might be discrepant. For a number of reasons, the London network is not an ideal basis for usability studies, and Experiment 3 tests matched designs for the Berlin network. However, it should also be noted that a recent study has compared two of the London compact designs. The curvilinear version has received the lowest usability rating of all in Experiment 2, whereas the octolinear map received the fourth highest rating. Roberts and Vaeng (2016) found that the curvilinear map was the better performer for mean journey planning times.

It should also be noted that the ratings in Experiment 2 were generated without direct experience at using the maps. Looking at the Paris data for Experiment 1, the questionnaire rating for the curvilinear map, and also people's preferences, it is clear that the curvilinear map was received more favourably than the London curvilinear maps were for Experiment 2. This might be because of the poor simplicity of the octolinear design in Experiment 1, or because actual experience at the curvilinear map tempered people's initial adverse reaction to it. It is therefore clearly important to investigate evaluations in the light of experience.

4. EXPERIMENT 3

Experiment 3 was intended to commence investigation of the relationship between usability ratings and objectively measured usability. From the point of view of testing, investigating trios of maps enables an investigation of all levels of one factor (e.g., Design Rules) within one level of the other (e.g. Stylized maps from the Design Priorities factor). However, although the London Underground map is ideal for investigating subjective usability ratings due to its familiarity and status amongst British respondents, it is less well-suited to a usability study in which many journeys need to be planned. This is because the network is highly interconnected, more so than, for example, the Paris Metro or the Berlin U- and S-Bahn system, and hence it is harder to devise complex two-interchange journeys. Experiment 3 therefore investigated the usability of three Berlin U-/S-Bahn maps, created manually by the first author, and all intended to be optimised for simplified line trajectories. Octolinear, curvilinear, and multilinear versions were tested, equivalent to the stylised London Underground maps in Experiment 2 (see Figures 5, 6, 7). Previously, the stylised maps evoked the greatest range of ratings for usability and attractiveness, suggesting that these would be a good starting point for investigating correlations between ratings and objective measures of usability.

Subjects each received all three maps, one after the other, and planned six complex journeys for each one. After this, they completed a statement rating task in which they evaluated all three designs, rated them for usability and attractiveness, and finally completed a similar task to Experiment 2, rating the equivalent design priority London Underground maps for usability and attractiveness.

Creating the three Berlin maps presented few problems, irrespective of the design rules, with simple line trajectories possible for all three of them (Roberts, 2012). It is therefore unlikely that there will be substantial differences in objective usability measures between them, so that unlike Experiment 1, differences in ratings of usability and attractiveness between them will be from a flat objective baseline. By comparing ratings for London and Berlin designs in the same study, three further developments are possible. First, it can be investigated whether the ratings for the London maps in Experiment 3 are comparable with Experiment 2, where it is possible that the sample tended towards people with an interest in schematic maps. Second, the ratings for the Berlin designs in Experiment 3 can be compared with the ratings for the London designs in Experiment 3. A UK sample was tested in Experiment 3, based in SE England, who would be expected to be familiar with the London network but not the Berlin network, and so, for example, ratings of octolinear maps for high versus low familiarity cities can be compared. Finally, it is possible to investigate whether it is possible that actual experience at using maps is able to temper the very low ratings given to curvilinear and multilinear maps in Experiment 2. These low ratings should be contrasted with those for the octolinear versus curvilinear maps in Experiment 1, where the curvilinear map was not adversely rated compared with the octolinear version.

4.1. Method

4.1.1. Subjects

Seventy-two subjects took part in this experiment, 27 males and 45 females. They were acquaintances of the experimenters and were unpaid volunteers. Ages ranged from 22 to 62 with a mean of 35.9 years (SD 13.0). The majority were unfamiliar with Berlin, 58 of them had never visited the city, and only four had visited more than once.

4.1.2. Materials

The three maps for journey planning (curvilinear, octolinear, multilinear) were printed to fit an A3 sheet and laminated. Three sets of journeys were assembled, shown in Table 5. An attempt was made to represent all parts of the map and all lines equally in the stimulus set. Journeys were intended to be difficult to plan, with distant start and destination stations and many alternative options. Two interchanges were always required, either to complete a journey at all, or to avoid a roundabout journey. Each journey was shown on an individual laminated sheet with the map greyed out except for the start (arrowed) and destination stations.

The 24-item questionnaire was similar to Experiment 1, but owing to the large number of ratings to be made, five-point scales were used, ranging from strongly agree to neutral to strongly disagree for each statement. For each question, an answer or a decision was required for all three maps. Questions were modified to name Berlin but the first seventeen were otherwise identical to those used in Experiment 1.

Question 18 requested people to choose a preferred design from the three maps, as a forced choice rather than an expression of strength of preference.

18) Of the three designs you have used, which one do you think you would prefer for everyday use?

Question 19 gave a range of frequency options: every day; a few days every week; a few times every month; about once a month; a few times every year; once a year or less; never/not for years
19) Roughly how often do you travel by rail to make a journey in a town or city?

Question 20 gave a range of frequency options: never; once; twice; three times; more than three times
20) Roughly how often have you visited Berlin?

Questions 21 and 22 were similar rating tasks to Experiment 2, and people were requested to rate each of the three Berlin maps on the same three point scale for usability (question 21) and attractiveness (questions 22).

Then, for questions 23 and 24, the three equivalent London Underground maps – the stylized versions – were shown for the first time, and similar usability and attractiveness ratings requested.

4.1.3. Design and procedure

All subjects planned journeys using all three maps, six journeys for each design. Order of map testing and journey allocation to maps were both fully counterbalanced. Primarily, this was a within-subjects design with Map Type (three levels, Octolinear, Multilinear, and Curvilinear) as the independent variable. Measures of map performance included the time taken to plan a journey, the number of declared journeys which contained at least one error, and an estimation of the duration that the planned journeys would have taken had they been implemented. Questionnaire data provided a means of measuring people's subjective assessments of map usability.

The procedure was identical to that used for Experiment 1, with the exception that completed journeys on the laminated map were photographed rather than transcribed onto paper. Because of the large number of trials, no practice was given. Instead, subjects were permitted to view the first experimental map for 30 seconds before commencing the first trial, and subsequent maps for 15 seconds each.

4.2. Results

4.2.1. Calculating overall performance measures and ratings

For this experiment, seven final declared journeys contained at least one error, in which subjects proposed to make interchanges at points not designated as such on the map. When calculating estimated journey durations, deleting errors and averaging over the remainder is problematic because, if the planning of invalid routes is not randomly distributed between journeys and maps, disregarding these could improve mean journey duration estimates for maps in which the most invalid routes are proposed. As a solution to this, where a subject proposed an invalid route, the 90th percentile value for that journey for that map was substituted (i.e., the value of the upper tail from a box plot). This seems a reasonable compromise given that, in reality, any invalid route planned would lead to an extended journey, possibly considerably so, and is identical to the method used by Roberts et al. (2013). All other objective measures of performance were calculated as per Experiments 1 and 2, and means and standard deviations are given in Table 6.

Three-factor fully between-subjects analyses of variance were used to investigate the possibility of interactions between Map Type (Curvilinear, Octolinear, Multilinear) Order (map presented First, Second, or Third) or Item (map with Set A, Set B, or Set C) for journey planning time, estimated journey duration, and statement rating task aggregate. None of the relevant interactions was significant. The greatest F value was for the Map Type x Item Order interaction for journey planning time, F(4,189) = 2.38, MSe = 191.8, p > .05. Inspecting the interaction plot indicated that the main source of this potential interaction was the curvilinear map being associated with the fastest planning times of all when tested third. With no other coherent effects discernible, order and item effects seem to contribute no notable caveats to overall interpretations. There was a significant main effect of Item Set for the journey duration measure of performance, F(2,189) = 69.2, MSe = 24.2, p < .01, but all this indicated was that it is possible to plan journeys with a faster estimated duration for Set 3, then Set 2, then Set 1.

There were no significant interactions that involved sex differences in conjunction with the map type tested, either for journey planning time, estimated journey duration, or the statement rating task aggregate. There were also no significant correlations between these dependent variables, and age, frequency of visits to Berlin, or frequency of rail transport usage.

4.2.2. Basic usability measures and ratings

There were no significant differences between the maps for planning time or estimated journey duration, using a single-factor within-subjects ANOVA, F(2,142) = 0.245, MSe = 66.9, p > .05, and F(2,142) = 1.72, MSe = 37.8, p > .05. Hence, the three maps were almost perfectly matched for the objective measures of usability. Despite this, the maps differed significantly for the mean statement rating task aggregate score, F(2,142) = 6.08, MSe = 98.3, p < .01. The means for the curvilinear and octolinear maps did not differ significantly, F(1,142) = 1.57, MSe = 98.3, p > .05, but both significantly exceeded the mean for the multilinear map: octolinear versus multilinear; F(1,142) = 11.9, MSe = 98.3, p < .01, curvilinear versus multilinear; F(1,142) = 4.81, MSe = 98.3, p < .05.

For the Berlin maps, the usability and attractiveness rating means were comparable to those from Experiment 2, showing the same broad patterns, with the octolinear map receiving high ratings, and the multilinear map low ratings. Hence, looking at rated usability for the Berlin maps, using a single-factor within-subjects ANOVA, the differences in means were significant, F(2,142) = 5.17, MSe = 1531.7, p < .01. The mean for the octolinear map significantly exceeded that for the multilinear map, F(1,142) = 10.2, MSe = 1531.7, p < .01, with the curvilinear map between the two and not significantly differing from either: octolinear versus curvilinear; F(1,142) = 1.63, MSe = 1531.7, p > .05, curvilinear versus multilinear; F(1,142) = 3.67, MSe = 1531.7, p > .05. Looking at rated attractiveness for the Berlin maps, again the differences in means were significant, F(2,142) = 18.7, MSe = 1533.0, p < .01. The mean for the curvilinear map significantly exceeded that for the multilinear map, F(1,142) = 33.0, MSe = 1533.0, p < .01, with the octolinear map between the two and not significantly differing from the curvilinear map, F(1,142) = 1.13, MSe = 1533.0, p > .05, but differing from the multilinear map, F(1,142) = 21.9, MSe = 1533.0, p < .01.

The effects were considerably stronger for the London maps, indicating that familiarity influences them. For rated usability, the octolinear map was clearly rated as more usable than the other two, with a significant difference amongst the means, F(2,142) = 27.5, MSe = 1488.5, p < .01. Its mean rating exceeded the multilinear map, F(1,142) = 42.0, MSe = 1488.5, p < .01, and the curvilinear map, F(1,142) = 40.6, MSe = 1488.5, p < .01, which in turn did not differ from each other, F(1,142) = 0.012, MSe = 1488.5, p > .05. The same was the case for attractiveness, the octolinear map was again clearly rated as more attractive than the other two, with a significant difference amongst the means, F(2,142) = 41.1, MSe = 1509.5, p < .01. Its mean rating exceeded the multilinear map, F(1,142) = 81.2, MSe = 1509.5, p < .01, and the curvilinear map, F(1,142) = 13.3, MSe = 1509.5, p < .01, but this time the curvilinear map exceeded the multilinear map, F(1,142) = 28.8, MSe = 1509.5, p < .01. The pattern of the curvilinear map being rated as more attractive than usable, and the multilinear map rated as more usable than attractive, was virtually identical to Experiment 2, and confirmed by a significant Design Rules x Rating Task interaction, F(1,142) = 9.81, MSe = 582.6, p < .01.

Berlin and London usability ratings by map type were combined into a single two-factor within-subjects ANOVA. The Map Type x City interaction was significant, F(2,142) = 12.7, MSe = 777.0, p < .01. Pairwise comparisons showed that for the octolinear maps, on average the London version was rated as significantly more usable then the Berlin version, F(1,142) = 8.07, MSe = 777.0, p < .01, but for the curvilinear maps, on average the Berlin version was rated as significantly more usable than the London version, F(1,142) = 17.5, MSe = 777.0, p < .01. The means of the multilinear maps did not differ, F(1,142) = 2.70, MSe = 777.0, p > .05. Hence, multilinear maps received universally low usability ratings, octolinear maps received the highest ratings, but curvilinear maps received higher ratings for Berlin than London, and octolinear maps received higher ratings for London than Berlin.

Do usability and attractiveness ratings indicate general tendencies in preference that transcend cities, encompassing even unfamiliar ones? Looking at correlations in equivalent rating scores for equivalent pairs (e.g., usability rating for Berlin octolinear and London octolinear maps), all but one were significant, all $p < .01$. For usability: curvilinear maps, $r = .55$; multilinear maps, $r = .47$. For attractiveness: octolinear maps, $r = .36$; curvilinear maps, $r = .60$; multilinear maps, $r = .48$. Only the correlation for usability for the octolinear maps was non-significant; $r = .17$, $p > .05$. Here, the ceiling effect for the London octolinear map depressed the correlation, with the highest usability rating awarded by 57 out of 72 people.

Even with the ceiling effect, it is possible to investigate map choices in relation to usability ratings. As for Experiment 2, certain people's ratings were sufficiently consistent for them to be classified as holding an intuitive theory of design. Hence, eight people always gave the highest usability ratings to both London and Berlin octolinear maps, which were never equalled by ratings for any other design, and these people always chose the Berlin octolinear map from the three options. Although not a direct replication of Experiment 2, the proportion of subjects categorized in this way is similar to the proportion of people identified as holding a strong octolinearity theory in Experiment 2, and it is plausible to suggest that their theories of design are equivalent. Experiment 3, in which case, validates the identification of intuitive theories of design, so that that holding the strong octolinearity theory seems to have considerable consequences for map choice. Overall, as the rating advantage for octolinear maps declined, so did likelihood that these would be selected. Table 7 shows a clear relationship, χ^2 (df = 3) = 27.3, $p < .01$.

4.2.3. The relationship between subjective and objective measures

With three different maps to choose between, the question of whether map choice was related to performance at individual maps was investigated by comparing the values for the chosen maps versus the mean of the two non-chosen values. Hence, if an individual selected a curvilinear map, then the curvilinear planning time would be compared with the mean of the octolinear/multilinear map values. If an individual had an awareness of the planning times for each design, and this was influencing map choice, then the curvilinear planning times should be faster than the mean of the two non-chosen versions, giving an equivalent analysis to Experiment 1. Table 8 shows the means for chosen versus non-chosen maps for planning time, estimated journey duration, and aggregate statement rating score.

Similarly to Experiment 1, there was no association between map choice and performance for the objective measures. Using two-factor mixed design ANOVA, with Map Choice and Chosen/Non-chosen maps as the factors, neither of the F values associated with map choice exceeded 1. However, as before, map choice was highly associated with statement rating task aggregate scores, with a significant main effect of Map Choice, $F(1,69) = 109.7$, $MSe = 31.0$, $p < .01$, but no significant interaction. $F(2,69) = 0.25$, $MSe = 31.0$, $p > .05$. Hence, the questionnaire statement rating scores for the maps that were chosen considerably exceeded those for the maps that were not chosen.

4.3. Discussion

For Experiment 3, there were no differences in objective usability measures between maps, despite the different design rules. Even so, there were clear differences in preferences and ratings, with the multilinear map performing particularly poorly in this respect. There continued to be a dissociation between objective and subjective measures of performance, so that map choice was unrelated to planning time: the chosen design was not necessarily the most effective for the individual in that respect, although there was a strong tendency for the chosen design to receive high ratings for the aggregate score of the statement rating task.

In terms of usability and attractiveness ratings. The pattern for the London designs was virtually identical to Experiment 2. The pattern for the Berlin designs was broadly in line these, with a clear octolinear bias, and likewise a dissociation between usability and attractiveness, such that the multilinear map was again rated as more usable than it was attractive, and the curvilinear map was rated as being more attractive than it was usable. The main differences between Experiments 2 and 3 was that the Berlin curvilinear map received more favourable usability ratings than the London equivalent.

Like Experiment 2, it is possible to identify a subset of respondents who displayed a strong octolinearity theory, and the validity of this classification was demonstrated with the selection by these people of the Berlin octolinear design without exception.

Overall, Experiment 3 replicated and extended the key effects of Experiments 1 and 2, and the broader implications of these findings will be identified in the general discussion.

5. GENERAL DISCUSSION

There are a number of consistent patterns of findings across the three experiments. From a methodological point of view, the practicality of repeated measures designs has been demonstrated. It is viable to administer experiments in which all subjects plan journeys using all designs of map – up to three in Experiment 3 – with little evidence for crosstalk influencing performance between designs in any important way. Obviously this is a potential problem that researchers should always be alert to, but there is a clear indication that the benefits of the increase in statistical power will usually outweigh the potential costs. There was also never evidence for age or sex effects on preference or performance.

Experiments 1 and 3 also extend the major finding of Roberts et al. (2013); that there is little or no relationship between objective measures of performance (journey planning time, estimated journey duration) and subjective ratings of usability (map choice, aggregate questionnaire rating). This is despite all subjects planning journeys using all designs, and this strengthens the original finding considerably, which is now shown to apply at the individual as well as the group level. In both Experiments 1 and 3, the ultimate selection of a map was related to its mean aggregate statement rating task score. Hence, map choices are not arbitrary, but are unrelated to performance in terms of journey planning time.

As discussed in the introduction, failures to attain self-awareness of objective performance measures should not be surprising, because differences in mean times of a few seconds across several trials would not be particularly salient, and therefore would be difficult to detect. Future research could address this by, for example, providing subjects with their response times, correctly or incorrectly, via visible timers or announcements, or else request that they estimate their own journey planning times. However, an advance request (versus a retrospective one) for such an estimate, for high versus low load cognitive tasks, (for example, a poorly-designed map versus an effectively designed one) can result in biases in opposite directions for the differences between actual versus estimated times (Block, Hancock, & Zakay, 2010). It may even be the case that planning time are not a prime element in map selection, even if this information is salient. An alternative is hinted at in Experiment 1, with the marginally significant interaction between Map Type and Map Choice for estimated journey duration. This potentially suggests that subjects may be aware of some aspect of route directness, and this may be influencing their evaluations of map utility. A similar interpretation is also suggested by Roberts, Newton, & Canals (2016) in which a novel design for the Berlin U-Bahn/S-Bahn network, based upon concentric circles, received particularly adverse ratings from subjects, along with comments which indicated a belief that the design was forcing them to choose roundabout journeys. It would be premature to draw conclusions on the basis of these findings, but they do indicate that although subjects may have little self-awareness of their performance in terms of planning time, there may be at least some sensitivity to other objectively measurable aspects of performance, which could be investigated in future research. This would be an easy aspect of map design to test, for example by asking subjects to provide their own estimated journey durations, and to manipulate this aspect between different maps by the allocation of items for journey planning.

In both Experiments 1 and 3, the selection rates and aggregate statement ratings of the curvilinear designs, in relation to the octolinear versions, were very similar, even though the Paris curvilinear map was superior in terms of journey planning time to the octolinear version in Experiment 1, whereas in Experiment 3, the Berlin curvilinear design was identical to the octolinear version. This does suggest that the reasonably high selection rates of the curvilinear maps is more a function of their acceptability than their relative usability. In contrast, in Experiments 2 and 3, the London curvilinear map (stylised) received considerably lower usability ratings than the equivalent octolinear design. Obviously, this must partially reflect the uniquely high octolinear bias displayed for the London maps but, even so, the Berlin curvilinear map was rated as more usable than the London equivalent. This ratings boost might be because of a lower octolinear bias for unfamiliar cities versus familiar cities, with preferences being redistributed accordingly, or else that, as a result of direct experience, there is a discovery that the curvilinear map is basically usable and this improves people’s overall rating of it. This boost in rating was not present for the Berlin multilinear map, but this particular design also received a low attractiveness rating. Overall, it is possible that the less adverse usability expectations for maps of unfamiliar cities using unfamiliar design rules allowed attractiveness ratings to determine preferences more strongly. A third possibility is that the stylised London curvilinear map is simply less usable than the octolinear equivalent, although a lack of user-sensitivity to this in the experiments thus far thus far would indicate a lack of plausibility for this suggestion. Furthermore, in Experiment 1, despite over 75% of subjects having faster journey planning times for the Paris curvilinear map, the patterns of scores at the statement rating task, and the proportion of choices for that design are almost identical to Roberts et al. (2013), suggesting that direct experience of this map does not boost its overall acceptability and rating. Hence, where observed in these experiments, the higher acceptability of curvilinear maps for Berlin and Paris versus London would appear to be owing to subjects being less familiar with the networks and maps of these cities, resulting in a lower expectation for the existence of octolinear designs, and hence decreasing the rating bias in their favour and raising ratings for other designs that they consider to be attractive.

Experiments 1 and 3 also replicate and extend those of Roberts et al. (2013) by continuing to refute the strong version of the octolinearity as a gold standard conjecture, i.e. that octolinearity will always yield the most effective design of schematic metro map, compared with other design rules but now demonstrating this finding for a second city; Berlin. Hence, the octolinear design was the worst performer for Paris, and failed to be the best performer for Berlin. However, weaker versions of the conjecture remain intact, for example, octolinearity will usually yield the most effective design of schematic metro map, except for networks whose structures are incompatible with these design rules. Even so, refuting the strong conjecture leads to the desirability of experimenting with other design rules for any network, attempting to maximize simplicity and coherence whilst minimizing topographical distortion. This finding likewise continues to feed forward to, for example, researchers who are investigating the automation of schematic map production (e.g., Nöllenburg, 2014; Nöllenburg & Wolff, 2011; Stott et al. 2011) where it suggests that layout algorithms need to be devised such that maps with different levels of linearity are created, each version is optimized in its own right for usability criteria, and then higher order measures of effectiveness are used to select the map whose design rules best enable simplicity, coherence, harmony, balance, and topographicity.

In terms of people's expectations and beliefs about usability, these appear to be strongly held within individuals, derived as a result of past exposure, and brought with them to the experimental situation, rather than developed during testing. These beliefs are, not surprisingly, subject to considerable individual differences, as evidenced by the correlations in usability ratings of same-design-rules maps in Experiment 3 (e.g. Berlin curvilinear/London curvilinear). There is a clear pattern across the experiments, such that there is an octolinear bias in usability ratings compared with other design rules (Experiments 2 and 3) and this is particularly strong for the city with the oldest and most famous octolinear design, London (Experiments 2 and 3). The bias is weaker but still present for less familiar cities (Berlin, Experiment 3) and persists to the extent that even a poorly optimised design that is outperformed by an alternative is still chosen by half of the subjects (Paris, Experiment 1). The similarity in the patterns of ratings for the London maps in Experiments 2 and 3 indicates that the sample of subjects in Experiment 2 was not unrepresentative in their beliefs or interests about effective design. Experiment 2 also demonstrated a smaller, but still clear simplicity preference so that, within a set of design rules, the version with the simplest line trajectories tended to be rated as the most usable. This shows, in conjunction with the low attractiveness ratings for multilinear maps, that there is at least some sensitivity by some users to the components of the framework for effective design outlined earlier.

The sources of the beliefs, or intuitive theories, that lead to usability ratings are not simply related to the perceived attractiveness of designs. Hence there was a clear dissociation between attractiveness and usability ratings for the curvilinear versus the multilinear maps for London (Experiments 2 and 3) and also Berlin (Experiment 3). For all applicable cities/experiments, curvilinear maps were rated as being more attractive than usable, whereas multilinear maps were rated as being more usable than attractive. The bias in usability ratings against curvilinear designs manifested itself sufficiently consistently in Experiment 2 for 41 subjects to be identified as holding a linearity theory, i.e. that a schematic map based upon any combination of straight lines will always be more usable than an equivalent curvilinear version matched for design priorities. In Experiment 3, lack of familiarity of a city appears to lead to improved usability ratings for the curvilinear design, but not for the equivalent multilinear version. This suggests that for London, subjects expect octolinearity to result in the most usable design, but for less familiar cities, this expectation is not present to such a strong degree, and attractiveness drives usability judgements more strongly. The low attractiveness ratings for multilinear designs possibly result from the difficulty in achieving coherence, possibly to the extent that many users find such designs disconcerting. The high attractiveness ratings for curvilinear maps might reflect an intrinsic preference for curved as opposed to awkwardly angled shapes (Silvia & Barona, 2009). This suggests that, should a decision be made for a particular network that octolinearity is not be appropriate, a curvilinear design is likely to result in more user acceptance than a multilinear one, and therefore should be implemented in preference.

In Experiment 2, looking at the various categorization variables for the respondents, very few qualified the usability ratings in any substantial way. There was little evidence for differences between the UK, USA-Canada, and Germany, which might be surprising given the different cartographic traditions. The lack of clear effects might imply that octolinear designs are so well known internationally that their familiarity is high, even when they are not adopted locally. Alternatively, that the visual properties of such designs may indeed give them an intrinsic advantage in terms of ratings, even if actual usability advantages cannot be detected. This returns us to consideration of the fluency framework (Alter & Oppenheimer, 2009) and the suggestion that perceived ease-of-processing is a factor that contributes towards the positive assessment of stimuli. If we take a starting point that the favourable ratings of octolinear maps might be related to some sort of cognitive fluency advantage compared with other designs, this raises the question of what aspects of octolinear angles leads to this, and what sources of converging evidence might confirm this. Although such designs are undoubtedly familiar, the curvilinear Paris Metro map in Experiment 1 was that one that was associated with faster planning times. Clearly, although the fluency framework is potentially useful, it does lack specification, but the use of metro maps, their usability data, and their subjective ratings might yield an interesting way to explore the framework further, especially in conjunction with the various elements of the framework for effective design.

Looking more closely at the usability ratings in Experiment 2, almost 1/3 of responses were sufficiently consistent and coherent for it to be possible for subjects to be classified as having a clear intuitive theory of design. Of particular interest is that almost twice as many people demonstrated an octolinearity theory than a simplicity theory, and virtually no people demonstrated both. The hypothesized consequences of holding a strong octolinearity theory were partially validated in Experiment 3, where people identified as holding an equivalent of this always selected the Berlin octolinear map as opposed to the curvilinear and multilinear designs. Also of interest is the finding in Experiment 2, that the prevalence and proportions of octolinearity versus simplicity theories was essentially identical when comparing people who might be expected to have expertise in, or sensitivity to, issues relevant to schematic map design, versus people with no basis for being expected to have such expertise. Hence the responses of ‘experts’ were no more coherent than ‘novices’, and displayed no identifiably greater level of theoretical sophistication, for example greater prevalence of any sort of intuitive theory of design. This obviously has implications concerning the end-products of procurement and design procedures for schematic maps at the hands of transport and design professionals. The generalisability of such findings is limited in the current context, but interesting future research is suggested. For example, will the possession, or not, of an octolinearity theory affect the choice of design rules in mapping a network? Alternatively, will designers who hold a simplicity theory implement a map differently within any given set design rules – with a preference for simpler line trajectories – compared with designers who do not? The lack of obvious evidence of sophistication of evaluation skills by supposed experts also calls into question studies whose conclusions, in part, rely solely on expert judgements of design rather than empirical evidence, with Field and Cartwright (2014) as a notable example.

6. CONCLUSIONS

The experiments here, and previously conducted (e.g., Roberts et al., 2013; Roberts, Newton, & Canals, 2016; Roberts & Vaeng, 2016) all show that objectively measured usability differences between different designs of schematic map can be identified, and that these can be shown to be consistent with the framework for effective design, and also that there are underlying psychological principles from which they can be shown to be compatible (Roberts, 2016; Roberts et al., 2013). In conjunction with Roberts et al. (2013) and Roberts and Vaeng (2016) the findings continue to refute the strong version of the octolinearity as a gold standard conjecture. Hence, schematic maps should be designed such that they maximise simplicity, coherence, harmony, balance, and topographicity, and if a particular network has properties that make octolinear layouts struggle to achieve these criteria, then alternative design rules should be used instead.

These experiments, and those previously cited, also all show that objectively measured schematic map usability differences are unrelated to: subjective ratings of the different versions; intuitive theories of design; and preferences between different designs. These subjective aspects, measured in different ways, are correlated, coherent (often strongly so) and drive map acceptability. They appear to be in the possession of subjects before usability testing commences, and hence are not modified by their experience at using the maps to plan journeys during a test session. The manifestation of the subjective aspects detected so far appears to be owing to a combination of (1) expectations concerning appropriate design rules for especially familiar cities and their urban rail network maps, and (2) impressions of attractiveness, which are more likely to impinge on usability judgements for less familiar cities/networks. These subjective aspects strongly drive design acceptability in the absence of any objective basis to make usability judgments, but are also subject to considerable individual differences.

Other subjective aspects of map evaluation should also be considered in future studies, for example, factors that contribute to perceived complexity and fluency of designs and their use. Gaze-tracking might be an appropriate methodology (e.g., Burch, Kurzhals, Raschke, Blascheck, & Weiskopf, 2014) potentially showing, for example, that complex maps are associated with more frequent and less systematic saccades. Also potentially important are self-perceptions of performance and the factors that influence them, for example subjective estimations of journey planning times and journey duration. Investigating metacognitive aspects of performance may be able to provide a bridge between purely subjective versus purely objective measures, and enable potential dissociations between acceptability versus usability of effective but unconventional designs to be narrowed in the future

Of course, it is possible that hitherto unmeasured objective aspects of usability might be usable as a basis to distinguish between the effectiveness of different designs and be more closely correlated with subjective judgements. One possibility might be the ease with which the most efficient, direct journey (where it exists) can be identified from a map, rather than the time taken to plan any journey at all. However, until such measures are identified and refined, or until theories of effective schematic map design are identified that make strong usability predictions between designs, there is no short cut to identifying usability differences between prototypes, whether on the basis of focus groups, questionnaire ratings, or public vote. Only extensive objective usability tests, such as those conducted by Roberts and Rose (2015) can hope to identify the designs most suitable for official adoption.

ACKNOWLEDGEMENTS

Data for Experiment 3 were collected by Sarah Wendland as part of a final year dissertation, and Paul Twyman.

REFERENCES

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. Personality and Social Psychology Review, 13, 219-235.

BBC Television (1987). Design classics. First broadcast 29th June.

BBC Television (2006). The culture show. First broadcast 16th March.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. Acta Psychologica, 134, 330-343.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. Psychological Bulletin, 106, 265-289.

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]

Burch, M., Kurzhals, K., Raschke, M., Blascheck, T., & Weiskopf, D. (2014). How do people read metro maps? An eye tracking study. Schematic Mapping Workshop 2014, University of Essex, April. https://sites.google.com/site/schematicmapping/Burch-EyeTracking.pdf [accessed 20/11/2015]

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

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.

Dierckx, V., & Vandierendonck, A. (2005). Adaptive strategy application in linear reasoning. In M. J. Roberts, & E. J. Newton (Eds.), Methods of thought: Individual differences in reasoning strategies (pp. 107–127). Hove: Psychology Press.

Field, K., & Cartwright, W. (2014). Becksploitation: The over-use of a cartographic icon. The Cartographic Journal, 51, 343-359.

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. (2016). Lost in transportation: Information measures and cognitive limits in multilayer navigation. Science Advances, 2, e1500445.

Garland, K. (1994). Mr Beck's Underground map. Harrow Weald, UK: Capital Transport Publishing.

Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. Annual Review of Anthropology, 40, 379–388.

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

Guo, Z., & Wilson, N. H. M. (2011). Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. Transportation Research Part A, 45, 91-104.

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. Journal of Environmental Psychology, 28, 74-82.

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

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

Lloyd, P. B. (2012). Vignelli: Transit maps. Rochester, NY: RIT Cary Graphic Arts Press.

McCloskey, M. (1983). Intuitive physics. Scientific American, 248, 122-130.

Nöllenburg, M. (2014). A survey on automated metro map layout methods. Schematic Mapping Workshop 2014, University of Essex, April. https://sites.google.com/site/schematicmapping/Nöllenburg_survey.pdf [accessed 29/05/2015]

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.

Ovenden, M. (2005). Metro maps of the world (2nd ed). Harrow Weald, UK: Capital Transport Publishing.

Ovenden, M. (2007). Transit maps of the world. New York: Penguin Books.

Ovenden, M. (2009). Paris underground: The maps, stations, and design of the Métro. New York: Penguin Books.

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

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. Annual Review of Psychology, 64, 77-107.

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. http://www.tubemapcentral.com/articles/roberts_empirical_2014.pdf [accessed 23/06/2016]

Roberts, M. J. (2014b). What's your theory of effective schematic map design? Schematic Mapping Workshop 2014, University of Essex, April. http://www.tubemapcentral.com/articles/roberts_theoretical_2014.pdf [accessed 23/06/2016]

Roberts, M. J. (2016). From reasoning and intelligence research to information design: Understanding and optimising the usability and acceptability of schematic transit maps. In N.D. Galbraith, E. Lucas & D. Over (eds.), The thinking mind: A festschrift for Ken Manktelow. Hove, UK: Psychology Press.

Roberts, M. J., Newton, E. J., & Canals, M. (2016). Radi(c)al departures: Comparing conventional octolinear versus concentric circles schematic maps for the Berlin U-Bahn/S-Bahn networks using objective and subjective measures of effectiveness. Information Design Journal.

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., & Rose, D. (2015). Map-Induced Journey-Planning Biases for a Simple Network: A Docklands Light Railway Study. Unpublished Manuscript.

Roberts, M. J., & Vaeng, I. C. N. (2016). Expectations and prejudices usurp judgements of schematic map effectiveness. Design Research Society 50th Anniversary

Conference, Brighton, June. http://www.tubemapcentral.com/articles/Roberts_Vaeng_DRS2016.pdf [accessed 23/06/2016]

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.

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.

Vertesi, J. (2008). Mind the Gap: The London Underground map and users' representations of urban space. Social Studies of Science 38, 7–33.

Willis, K. S., Hölscher, C., Wilbertz, G., & Li, C. (2009). A comparison of spatial knowledge acquisition with maps and mobile maps. Computers, Environment and Urban Systems, 33, 100-110.

Figure 1. Sections of maps showing the line trajectories of an area of central London around Paddington. (1) represents topographical reality, as shown in maps immediately prior to Henry Beck. (2) shows Henry Beck's schematic depiction of reality. (3) demonstrates poor simplicity, with many corners, (4) has poor coherence, the trajectories are simple, but there are many different angles, and no pairs of lines are parallel. (5) has few angles and straight, parallel lines, but these do not cross at 90º and hence the harmony is poor. Such a design might be easy to use, but rejected by users. (6) has uneven station distribution and therefore poor balance, and (7) distorts topography considerably, hence poor topographicity. Image © Maxwell J. Roberts, 2014, reproduced with permission, all rights retained.

Figure 2. Three configurations of the same section of the London Underground network, illustrating the approaches to schematic mapping investigated in this paper. The octolinear diagram (top) represents a conventional approach, with four angles permitted: horizontal, vertical, and 45º diagonals. The disadvantage of such this is that where the network configuration does not quite match the available angles, complex line trajectories may be necessary. A possible solution, a multilinear design (centre) permits any angle of line to be used. This can result in simpler, straighter line trajectories, although the risk in using many angles is that the coherence of the design is undermined. Even with the multilinear design, there still may be harsh changes in direction, and a curvilinear approach seeks to improve the overall flow of the diagram by depicting the lines as smooth curves with minimal changes in curvature and avoiding points of inflection wherever possible. Image © Maxwell J. Roberts, 2014, reproduced with permission, all rights retained.

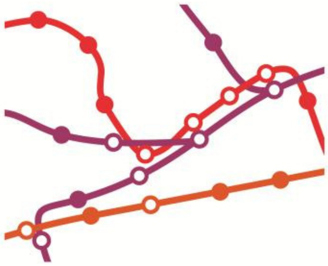
Figure 3. Curvilinear Paris Metro Map used for Experiment 1, © Maxwell J. Roberts, 2007. Reproduced with permission, all rights retained. The version of the official map used in this experiment is included in a package of maps that is available at http://www.tubemapcentral.com/Map_Package_Roberts.zip (accessed 23/06/2012)

Figure 4. The matrix of nine London Underground maps presented in Experiment 2. All designs © Maxwell J. Roberts, 2007-2014. Reproduced with permission, all rights retained.

Figure 5. Curvilinear Berlin U-/S-Bahn Map used for Experiment 3, © Maxwell J. Roberts, 2012. Reproduced with permission, all rights retained.

Figure 6. Octolinear Berlin U-/S-Bahn Map used for Experiment 3, © Maxwell J. Roberts, 2012. Reproduced with permission, all rights retained.

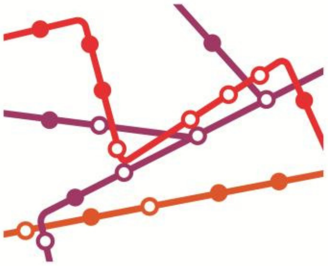
Figure 7. Multilinear Berlin U-/S-Bahn Map used for Experiment 3, © Maxwell J. Roberts, 2012. Reproduced with permission, all rights retained.



1: geographical reality

2: Henry Beck's 1933 configuration

3: poor simplicity

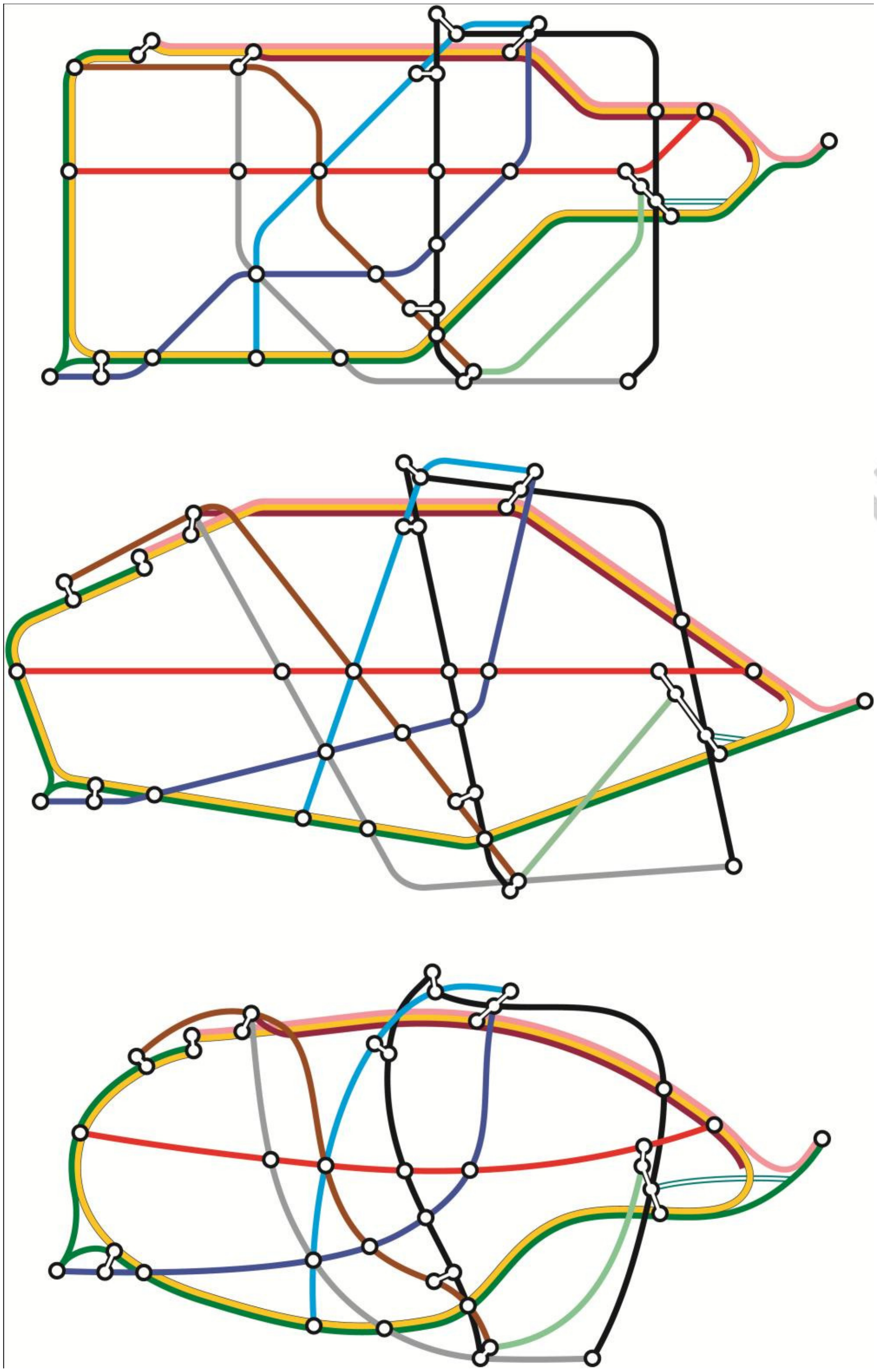


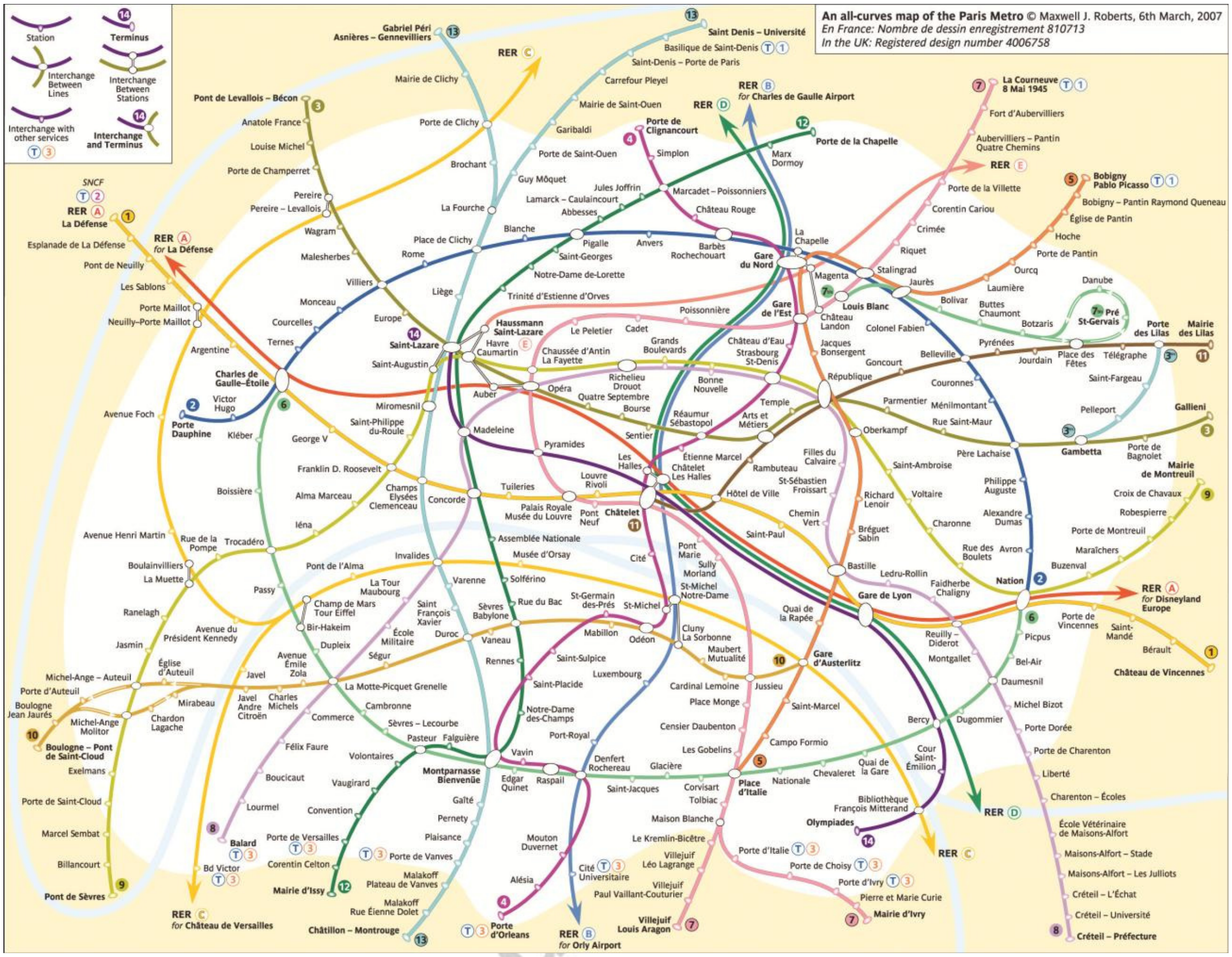
4: poor coherence

5: poor harmony

6: poor balance

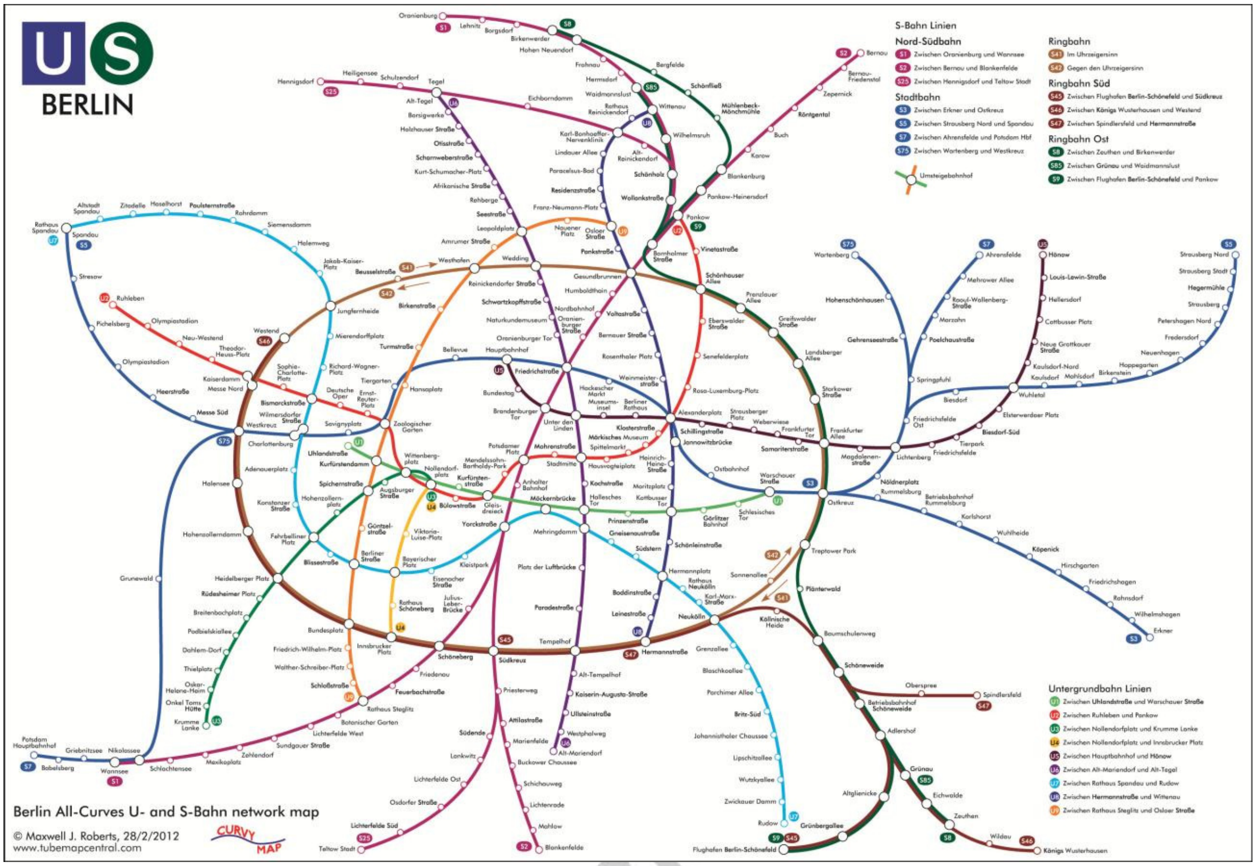
7: poor topographicity

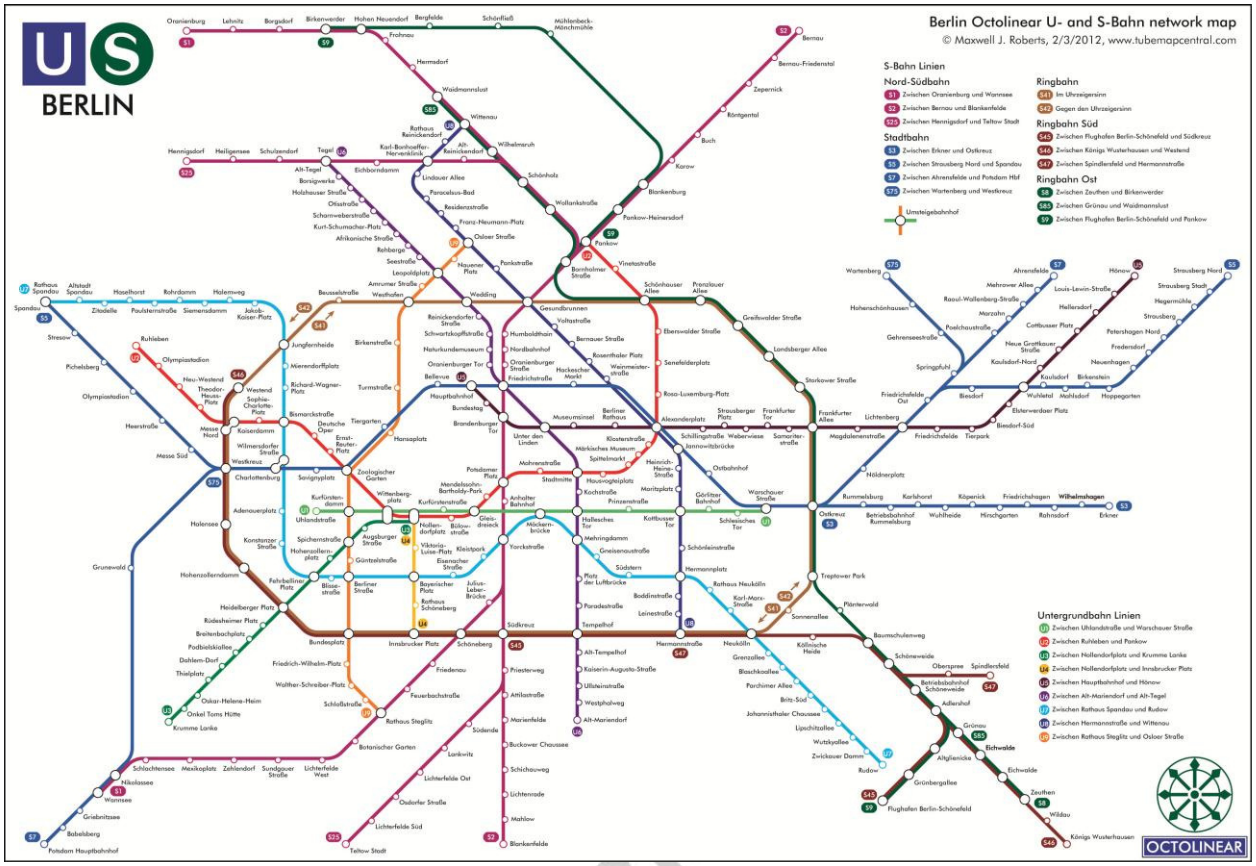




An all-curves map of the Paris Metro © Maxwell J. Roberts, 6th March, 2007
En France: Nombre de dessin enregistrement 810713
In the UK: Registered design number 4006758


GEOGRAPHICAL MULTILINEAR
COMPACT MULTILINEAR
STYLISED MULTILINEAR
GEOGRAPHICAL OCTOLINEAR
COMPACT OCTOLINEAR
STYLISED OCTOLINEAR
GEOGRAPHICAL CURVILINEAR
COMPACT CURVILINEAR
STYLISED CURVILINEAR


U S
BERLIN
S-Bahn Linien
Nord-Südbahn
S1 Zwischen Oranienburg und Wannsee
S2 Zwischen Bernau und Blankenfelde
S25 Zwischen Hennigsdorf und Teltow Stadt
Stadtbahn
S3 Zwischen Erkner und Ostkreuz
S5 Zwischen Strausberg Nord und Spandau
S7 Zwischen Ahrensfelde und Potsdam Hbf
S75 Zwischen Wartenberg und Westkreuz
Umsteigebahnhof
Ringbahn
S41 Im Uhrzeigersinn
S42 Gegen den Uhrzeigersinn
Ringbahn Süd
S45 Zwischen Flughafen Berlin-Schönefeld und Südkreuz
S46 Zwischen Königs Wusterhausen und Westend
S47 Zwischen Spindlersfeld und Hermannstraße
Ringbahn Ost
S8 Zwischen Zeuthen und Birkenwerder
S85 Zwischen Grünau und Waidmannslust
S9 Zwischen Flughafen Berlin-Schönefeld und Pankow
Untergrundbahn Linien
U1 Zwischen Uhlandstraße und Warschauer Straße
U2 Zwischen Ruhleben und Pankow
U3 Zwischen Nollendorfplatz und Krumme Lanke
U4 Zwischen Nollendorfplatz und Innsbrucker Platz
U5 Zwischen Hauptbahnhof und Hönow
U6 Zwischen Alt-Mariendorf und Alt-Tegel
U7 Zwischen Rathaus Spandau und Rudow
U8 Zwischen Hermannstraße und Wittenau
U9 Zwischen Rathaus Steglitz und Osloer Straße
Berlin All-Curves U- and S-Bahn network map
© Maxwell J. Roberts, 28/2/2012
www.tubemapcentral.com
CURVY MAP


U S
BERLIN
Berlin Octolinear U- and S-Bahn network map
© Maxwell J. Roberts, 2/3/2012, www.tubemapcentral.com
S-Bahn Linien
Nord-Südbahn
S1 Zwischen Oranienburg und Wannsee
S2 Zwischen Bernau und Blankenfelde
S25 Zwischen Hennigsdorf und Teltow Stadt
Stadtbahn
S3 Zwischen Erkner und Ostkreuz
S5 Zwischen Strausberg Nord und Spandau
S7 Zwischen Ahrensfelde und Potsdam Hbf
S75 Zwischen Wartenberg und Westkreuz
Umsteigebahnhof
Ringbahn
S41 Im Uhrzeigersinn
S42 Gegen den Uhrzeigersinn
Ringbahn Süd
S45 Zwischen Flughafen Berlin-Schönefeld und Südkreuz
S46 Zwischen Königs Wusterhausen und Westend
S47 Zwischen Spindlersfeld und Hermannstraße
Ringbahn Ost
S8 Zwischen Zeuthen und Birkenwerder
S85 Zwischen Grünau und Waidmannslust
S9 Zwischen Flughafen Berlin-Schönefeld und Pankow
Untergrundbahn Linien
U1 Zwischen Uhlandstraße und Warschauer Straße
U2 Zwischen Ruhleben und Pankow
U3 Zwischen Nollendorfplatz und Krumme Lanke
U4 Zwischen Nollendorfplatz und Innsbrucker Platz
U5 Zwischen Hauptbahnhof und Hönow
U6 Zwischen Alt-Mariendorf und Alt-Tegel
U7 Zwischen Rathaus Spandau und Rudow
U8 Zwischen Hermannstraße und Wittenau
U9 Zwischen Rathaus Steglitz und Osloer Straße
OCTOLINEAR

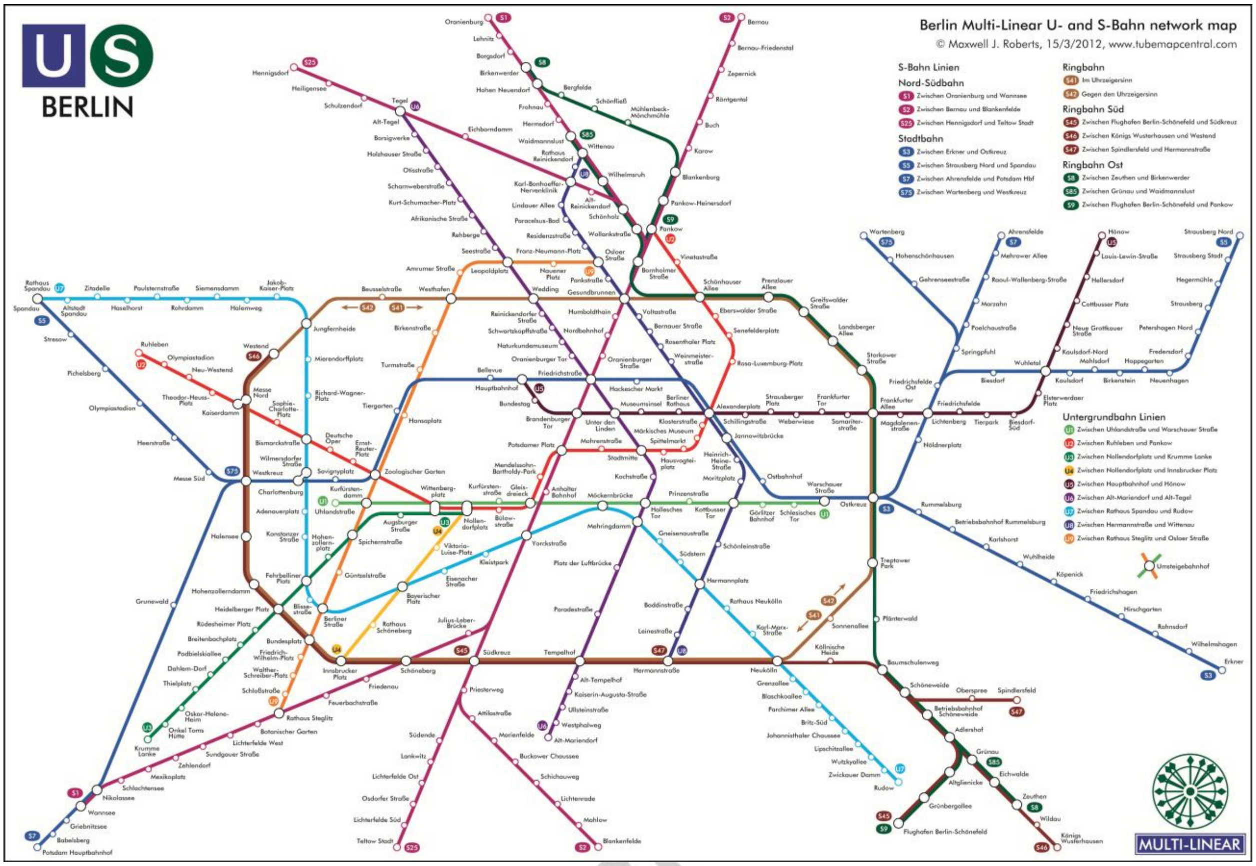

U S
BERLIN
Berlin Multi-Linear U- and S-Bahn network map
© Maxwell J. Roberts, 15/3/2012, www.tubemapcentral.com
S-Bahn Linien
Nord-Südbahn
S1 Zwischen Oranienburg und Wannsee
S2 Zwischen Bernau und Blankenfelde
S25 Zwischen Hennigsdorf und Teltow Stadt
Stadtbahn
S3 Zwischen Erkner und Ostkreuz
S5 Zwischen Strausberg Nord und Spandau
S7 Zwischen Ahrensfelde und Potsdam Hbf
S75 Zwischen Wartenberg und Westkreuz
Ringbahn
S41 Im Uhrzeigersinn
S42 Gegen den Uhrzeigersinn
Ringbahn Süd
S45 Zwischen Flughafen Berlin-Schönefeld und Südkreuz
S46 Zwischen Königs Wusterhausen und Westend
S47 Zwischen Spindlersfeld und Hermannstraße
Ringbahn Ost
S8 Zwischen Zeuthen und Birkenwerder
S85 Zwischen Grünau und Waidmannslust
S9 Zwischen Flughafen Berlin-Schönefeld und Pankow
Untergrundbahn Linien
U1 Zwischen Uhlandstraße und Warschauer Straße
U2 Zwischen Ruhleben und Pankow
U3 Zwischen Nollendorfplatz und Krumme Lanke
U4 Zwischen Nollendorfplatz und Innsbrucker Platz
U5 Zwischen Hauptbahnhof und Hönow
U6 Zwischen Alt-Mariendorf und Alt-Tegel
U7 Zwischen Rathaus Spandau und Rudow
U8 Zwischen Hermannstraße und Wittenau
U9 Zwischen Rathaus Steglitz und Osloer Straße
Umsteigebahnhof
Oranienburg
Lehnitz
Borgsdorf
Birkenwerder
Hohen Neuendorf
Bergfelde
Schönfließ
Mühlenbeck-Mönchmühle
Frohnau
Hermsdorf
Waidmannslust
Wittenau
Rathaus Reinickendorf
Wilhelmsruh
Karl-Bonhoeffer-Nervenklinik
Alt-Reinickendorf
Lindauer Allee
Paracelsus-Bad
Schönholz
Residenzstraße
Wollankstraße
Franz-Neumann-Platz
Osloer Straße
Pankow
Vinetastraße
Bernau
Bernau-Friedenstal
Zepernick
Röntgental
Buch
Karow
Blankenburg
Pankow-Heinersdorf
Hennigsdorf
Heiligensee
Schulzendorf
Tegel
Alt-Tegel
Eichborndamm
Borsigwerke
Holzhauser Straße
Otisstraße
Scharnweberstraße
Kurt-Schumacher-Platz
Afrikanische Straße
Rehberge
Seestraße
Amrumer Straße
Leopoldplatz
Nauener Platz
Pankstraße
Wedding
Gesundbrunnen
Bornholmer Straße
Schönhauser Allee
Prenzlauer Allee
Greifswalder Straße
Landsberger Allee
Storkower Straße
Frankfurter Allee
Rathaus Spandau
Zitadelle
Paulsternstraße
Siemensdamm
Jakob-Kaiser-Platz
Altstadt Spandau
Haselhorst
Rohrdamm
Halemweg
Spandau
Stresow
Pichelsberg
Olympiastadion
Beusselstraße
Westhafen
Jungfernheide
Mierendorffplatz
Richard-Wagner-Platz
Ruhleben
Neu-Westend
Theodor-Heuss-Platz
Kaiserdamm
Westend
Messe Nord
Sophie-Charlotte-Platz
Bismarckstraße
Deutsche Oper
Ernst-Reuter-Platz
Wilmersdorfer Straße
Heerstraße
Messe Süd
Westkreuz
Charlottenburg
Savignyplatz
Zoologischer Garten
Birkenstraße
Turmstraße
Bellevue
Tiergarten
Hansaplatz
Hauptbahnhof
Bundestag
Reinickendorfer Straße
Schwartzkopffstraße
Naturkundemuseum
Oranienburger Tor
Humboldthain
Nordbahnhof
Voltastraße
Bernauer Straße
Rosenthaler Platz
Weinmeisterstraße
Oranienburger Straße
Friedrichstraße
Hackescher Markt
Eberswalder Straße
Senefelderplatz
Rosa-Luxemburg-Platz
Alexanderplatz
Brandenburger Tor
Unter den Linden
Museumsinsel
Berliner Rathaus
Klosterstraße
Märkisches Museum
Strausberger Platz
Weberwiese
Frankfurter Tor
Samariterstraße
Schillingstraße
Jannowitzbrücke
Heinrich-Heine-Straße
Potsdamer Platz
Mohrenstraße
Stadtmitte
Spittelmarkt
Hausvogteiplatz
Kochstraße
Moritzplatz
Ostbahnhof
Warschauer Straße
Ostkreuz
Kurfürstendamm
Wittenbergplatz
Kurfürstenstraße
Gleisdreieck
Mendelssohn-Bartholdy-Park
Anhalter Bahnhof
Möckernbrücke
Hallesches Tor
Prinzenstraße
Kottbusser Tor
Görlitzer Bahnhof
Schlesisches Tor
Uhlandstraße
Adenauerplatz
Augsburger Straße
Nollendorfplatz
Bülowstraße
Konstanzer Straße
Hohenzollernplatz
Spichernstraße
Viktoria-Luise-Platz
Kleistpark
Yorckstraße
Mehringdamm
Gneisenaustraße
Südstern
Schönleinstraße
Hermannplatz
Rathaus Neukölln
Karl-Marx-Straße
Neukölln
Halensee
Fehrbelliner Platz
Güntzelstraße
Eisenacher Straße
Bayerischer Platz
Platz der Luftbrücke
Hohenzollerndamm
Heidelberger Platz
Blissestraße
Berliner Straße
Rathaus Schöneberg
Julius-Leber-Brücke
Paradestraße
Boddinstraße
Leinestraße
Grunewald
Rüdesheimer Platz
Breitenbachplatz
Podbielskiallee
Dahlem-Dorf
Thielplatz
Oskar-Helene-Heim
Onkel Toms Hütte
Krumme Lanke
Bundesplatz
Friedrich-Wilhelm-Platz
Walther-Schreiber-Platz
Schloßstraße
Rathaus Steglitz
Innsbrucker Platz
Schöneberg
Südkreuz
Tempelhof
Hermannstraße
Friedenau
Feuerbachstraße
Botanischer Garten
Lichterfelde West
Sundgauer Straße
Zehlendorf
Mexikoplatz
Schlachtensee
Nikolassee
Wannsee
Griebnitzsee
Babelsberg
Potsdam Hauptbahnhof
Priesterweg
Attilastraße
Marienfelde
Buckower Chaussee
Schichauweg
Lichtenrade
Mahlow
Blankenfelde
Südende
Lankwitz
Lichterfelde Ost
Osdorfer Straße
Lichterfelde Süd
Teltow Stadt
Alt-Tempelhof
Kaiserin-Augusta-Straße
Ullsteinstraße
Westphalweg
Alt-Mariendorf
Grenzallee
Blaschkoallee
Parchimer Allee
Britz-Süd
Johannisthaler Chaussee
Lipschitzallee
Wutzkyallee
Zwickauer Damm
Rudow
Sonnenallee
Köllnische Heide
Treptower Park
Plänterwald
Baumschulenweg
Schöneweide
Oberspree
Spindlersfeld
Betriebsbahnhof Schöneweide
Adlershof
Grünau
Altglienicke
Eichwalde
Grünbergallee
Zeuthen
Wildau
Königs Wusterhausen
Flughafen Berlin-Schönefeld
Wartenberg
Hohenschönhausen
Gehrenseestraße
Ahrensfelde
Mehrower Allee
Raoul-Wallenberg-Straße
Marzahn
Poelchaustraße
Springpfuhl
Friedrichsfelde Ost
Friedrichsfelde
Magdalenenstraße
Lichtenberg
Tierpark
Biesdorf-Süd
Nöldnerplatz
Biesdorf
Wuhletal
Hönow
Louis-Lewin-Straße
Hellersdorf
Cottbusser Platz
Neue Grottkauer Straße
Kaulsdorf-Nord
Kaulsdorf
Elsterwerdaer Platz
Mahlsdorf
Birkenstein
Hoppegarten
Neuenhagen
Fredersdorf
Petershagen Nord
Strausberg
Hegermühle
Strausberg Stadt
Strausberg Nord
Rummelsburg
Betriebsbahnhof Rummelsburg
Karlshorst
Wuhlheide
Köpenick
Friedrichshagen
Hirschgarten
Rahnsdorf
Wilhelmshagen
Erkner
MULTI-LINEAR

Table 1. Set A and Set B Paris Metro journeys used for Experiment 1

Set A

Practice trial: Dugommier (Line 6) to Goncourt (Line 11)

Avenue du President Kennedy (RER Line C) to Danube (Line 7b)

Glacière (Line 6) to Richelieu Drouot (Lines 8/9)

Richard Lenoir (Line 5) to St Francois Xavier (Line 13)

Ségur (Line 10) to Rue St-Maur (Line 3)

Set B

Practice trial: Victor Hugo (Line 2) to Vaneau (Line 10)

Porte d’Italie (Line 7) to Garibaldi (Line 13)

Rambuteau (Line 11) to Convention (Line 12)

Pelleport (Line 3b) to Charles Michel (Line 10)

Hoches (Line 5) to Rennes (Line 12)

Table 2. Questionnaire items used for Experiment 1 and adapted for Experiment 3. Asterisks denote statement rating task questions directly relevant to usability.

*	1)	I found journeys easy to plan using this map
*	2)	Routes were difficult to discriminate (identify) using this map
*	3)	Station names were easy to identify using this map
*	4)	Station interchanges were difficult to negotiate using this map
*	5)	Line trajectories were easy to follow using this map
*	6)	I found this map disorientating to use
*	7)	I would be happy to use this map to plan real-life journeys around Paris
	8)	I preferred a direct-looking route, no matter how many interchanges required
	9)	Some parts of the map looked complicated, and I planned journeys to avoid them
	10)	This map is intended for planning journeys but I think it is also geographically accurate
*	11)	With this map, I would rather walk or take a taxi than use the Paris Metro
	12)	The best routes for me had the fewest station stops along the way
*	13)	I found the map visually disturbing
*	14)	I found the map clean and uncluttered
*	15)	I would look for another design of Paris Metro map to use at the earliest opportunity
	16)	Briefly, what, if any, aspect of this map did you like the most?
	17)	Briefly, what, if any, aspect of this map did you like the least?
	18)	Of the two Paris Metro map designs, which one do you think you would prefer for everyday use?
	19)	Roughly how often do you travel by rail to make a journey in a town or city

Table 3. Mean usability measures by Map Type, and aggregate questionnaire ratings, and performance by map presentation order and map preference for Experiment 1.

		Official	Curvilinear
Planning time	Mean	50.0	40.2
(seconds per journey)	SD	23.0	17.5
Planning time (seconds per journey)	Mean	51.7	37.9
Official map presented first (N = 40)	SD	21.0	15.9
Planning time (seconds per journey)	Mean	48.4	42.4
Curvilinear map presented first (N = 40)	SD	24.9	19.0
Estimated journey duration	Mean	57.9	58.2
(minutes)	SD	6.5	6.6
Aggregate statement rating task score	Mean	50.3	53.3
(11 to 77, high scores better)	SD	10.0	11.4
Planning time (seconds per journey)	Mean	52.8	43.1
Official map is chosen (N = 38)	SD	25.5	18.4
Planning time (seconds per journey)	Mean	46.3	37.8
Curvilinear map is chosen (N = 34)	SD	20.3	17.8
Estimated journey duration (minutes)	Mean	56.4	59.8
Official map is chosen (N = 38)	SD	6.1	6.6
Estimated journey duration (minutes)	Mean	58.9	57.6
Curvilinear map is chosen (N = 34)	SD	5.9	6.3
Aggregate statement rating task score	Mean	54.9	47.1
Official map is chosen (N = 38)	SD	8.9	11.5
Aggregate statement rating task score	Mean	45.4	60.8
Curvilinear map is chosen (N = 34)	SD	9.4	6.8

Table 4. Mean usability and attractiveness ratings by Design Rules, and Design Priorities, for Experiment 2.

		Compact	Geographical	Stylised
Usability Ratings				
Octolinear	Mean	48.1	55.4	89.2
	SD	40.5	38.8	23.2
Multilinear	Mean	28.3	34.1	53.6
	SD	35.8	38.9	38.1
Curvilinear	Mean	21.2	28.4	38.9
	SD	33.4	36.3	39.7
Attractiveness Ratings				
Octolinear	Mean	48.4	52.5	85.6
	SD	39.8	37.8	26.0
Multilinear	Mean	19.1	23.2	40.5
	SD	31.0	32.6	37.6
Curvilinear	Mean	42.8	40.8	54.3
	SD	42.3	38.8	41.5

Table 5. Set A, Set B, and Set C Berlin U-Bahn journeys used for Experiment 3

Set A

S8 -- U4 Mühlenbeck-Mönchmühle to Rathaus Schöneberg

U7 -- S5 Eisenacher Straße to Birkenstein

U1 -- S2 Görlitzerbahnhof to Mahlow

U5 -- U3 Kaulsdorf-Nord to Hohenzollernplatz

U8 -- U2 Paracelsus-Bad to Olympiastadion

U9 -- S25 Güntzelstraße to Eichborndamm

Set B

S85 -- U9 Plänterwald to Hansaplatz

U6 -- U8 Westphalweg to Rosenthaler Platz

U3 -- S1 Thielplatz to Borgsdorf

S8 -- U7 Bergfelde to Südstern

S2 -- U2 Zepernick to Neu-Westend

S46 -- S5 Wildau to Strausberg

Set C

U4 -- S5 Viktoria-Luise-Platz to Neuenhagen

U9 -- U5 Walther-Schreiber-Platz to Louis-Lewin-Straße

U7 -- S1 Haselhorst to Frohnau

U8 -- U3 Rathaus Reinickendorf to Onkel Toms Hütte

U1 -- S47 Kurfürstenstraße to Oberspree

S7 -- S8 Babelsberg to Schönfließ

Table 6. Mean usability measures by Map Type, and aggregate questionnaire ratings, and attractiveness/usability ratings for Experiment 3.

		Curvilinear	Octolinear	Multilinear
Planning time	Mean	31.5	30.5	31.2
(seconds per journey)	SD	15.2	14.6	12.0
Estimated journey duration	Mean	63.0	64.7	64.5
(minutes)	SD	5.7	6.7	6.6
Aggregate statement rating task score	Mean	39.0	41.1	35.4
(11 to 55, high scores better)	SD	9.7	8.6	9.3
Mean usability rating	Mean	66.7	75.0	54.2
(0 to 100, high scores better)	SD	40.2	30.3	38.2
Mean attractiveness rating	Mean	73.6	66.7	36.1
(0 to 100, high scores better)	SD	40.2	31.4	34.8
London Maps				
Mean usability rating (Expt 3)	Mean	47.2	88.2	46.5
(0 to 100, high scores better)	SD	45.1	24.5	37.8
Mean usability rating (Expt 2)	Mean	38.9	89.2	53.6
(0 to 100, high scores better)	SD	39.7	23.2	38.1
Mean attractiveness rating (Expt 3)	Mean	59.7	83.3	25.0
(0 to 100, high scores better)	SD	47.2	25.2	33.6
Mean attractiveness rating (Expt 2)	Mean	54.3	85.6	40.5
(0 to 100, high scores better)	SD	41.5	26.0	37.6

Table 7. Berlin map choice by usability ratings of Berlin and London maps combined, for Experiment 3

	Chose Berlin Octolinear Map	Chose Berlin Curvilinear or Multilinear Map
Both octolinear maps received the highest usability rating, no other map received the highest usability rating: strong octolinearity	8	0
Both octolinear maps received the highest usability rating, at least one other map received the highest usability rating	17	13
At least one non-octolinear map received a higher usability rating than at least one octolinear map	4	19
Neither octolinear map received the highest rating, at least one other map exceeded them both	0	8

Table 8. Mean performance by map preference for Experiment 3.

		Chosen Maps	Non-Chosen Maps
Planning time: Curvilinear map chosen, N = 31	Mean	28.6	30.0
(seconds per journey)	SD	14.4	11.1
Planning time: Octolinear map chosen, N = 29	Mean	30.5	30.4
(seconds per journey)	SD	18.4	11.2
Planning time: Multilinear map chosen, N = 12	Mean	35.6	36.7
(seconds per journey)	SD	10.1	15.6
Estimated journey duration:	Mean	63.9	64.8
Curvilinear map chosen, N = 31 (minutes)	SD	5.1	4.8
Estimated journey duration:	Mean	65.2	64.0
Octolinear map chosen, N = 29 (minutes)	SD	6.7	4.6
Estimated journey duration:	Mean	63.6	61.2
Multilinear map chosen, N = 12 (minutes)	SD	4.5	5.8
Aggregate statement rating task score:	Mean	46.4	34.9
Curvilinear map chosen, N = 31 (11 to 55)	SD	5.3	6.0
Aggregate statement rating task score:	Mean	45.5	34.8
Octolinear map chosen, N = 29 (11 to 55)	SD	6.7	5.1
Aggregate statement rating task score:	Mean	44.7	35.0
Multilinear map chosen, N = 12 (11 to 55)	SD	5.5	6.6

Highlights

- Dissociation between objective and subjective measures of usability is demonstrated
- Octolinearity as a gold standard conjecture is refuted
- Usability and attractiveness ratings are collected in a large scale usability study
- Intuitive theories of design are identified