Objective and Subjective Measures of Metro Map Usability: Investigating the Benefits of Breaking Design Rules

Maxwell J. Roberts,

Department of Psychology, University of Essex, UK

Elizabeth J. Newton, & Fabio D. Lagattolla

Department of Psychology, London South Bank University, UK

RUNNING HEAD: METRO MAP USABILITY

15/06/2010

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

Objective and Subjective Measures of Metro Map Usability: Investigating the Benefits of Breaking Design Rules

ABSTRACT

Around the world, schematic maps are an important component of assistance for navigating transport networks. By showing routes as simple straight lines, they reduce the cognitive load of journey planning, and by revealing the underlying structure of networks, they make their key features easier to identify and learn. However, although there are many suggestions for optimizing schematic maps so as to maximize these benefits, to date these have not been supported by published usability studies or psychological theory. In this paper, we suggest that there are circumstances in which conventional schematic maps fail to yield benefits, and we compare journey planning using the current official Paris Metro map with a new design which replaces straight lines and corners with gentle curves. This All-Curves map was superior to the official design. Subjective usability measures were only weakly related to objective usability measures. We conclude that (1) in terms of designing schematics, there is no rule-set that can be claimed to be a gold-standard, and it is important to match the design rules to the properties of the network, (2) in some circumstances, radical departures from traditional ideas can yield usability benefits, and (3) *map usability* appears to be distinct from *map engagement*, although the latter is undoubtedly important in encouraging people fully to make use of navigation aides.

KEYWORDS

Schematic Maps, Metro Maps, Journey Planning, Usability Study, Representation, Reasoning

Objective and Subjective Measures of Metro Map Usability: Investigating the Benefits of Breaking Design Rules

Throughout the world, schematic network maps have become particularly associated with self-contained urban rapid transit (or metro) systems, and are an important means of showing routes and interchanges for the purpose of journey planning. For such networks, the simplicity and frequency of services, and the importance of station sequences and inter-connections (topological information) as opposed to exact route trajectories (topographical information) makes them particularly well-suited to being depicted in this way. Maps in this style originated at least as early as the 1920s, and became widespread from the 1970s onwards. They can be seen all round the world (Ovenden, 2005) although there are notable exceptions, such as the official map of the New York Subway network [1].

Like many concepts, defining whether or not a network map can strictly be categorized as being schematic is not easy (Dow, 2005; Roberts, 2005). The key criteria are that street details are absent (although major landmarks such as parks, rivers, and seas may be shown), and that lines or routes are shown as straight lines with sharply radiused corners. The aim is broadly to simplify the information presented so that the user can identify the key elements for journey planning (routes and interchanges) and follow the line trajectories easily. Usually, a restricted number of angles is permitted, typically just four, with horizontal, vertical, and 45 degree diagonals. This is also known as *octilinearity* (Nöllenburg & Wolff, 2009; Wolff, 2007): In other words, at any point on a line, only eight different trajectories are possible. Applying these design rules almost always results in topographical distortion to the relative locations of stations, especially if the map is designed so that their names do not interrupt lines. This usually requires an expanded centre at the expense of more sparsely served outer areas. Excessive topographical distortion can lead to complaints from local users, such as when the 2007 Madrid Metro map was introduced, but to date no studies have investigated whether a high degree of distortion interferes with journey planning amongst city experts versus novices to any substantial degree. However, Berendt, Rauh, and Barkowsky (1998) have shown that users can and do make spatial inferences about station locations from schematic network maps.

In many parts of the world, cities are investing substantially in metro construction. If we define a complex metro system as exceeding a route mileage of 200 km, the historically complex networks, such as the London Underground, Moscow Metro, New York Subway, Paris Metro, and Tokyo Subway have been joined by the networks of Madrid, Mexico City and Seoul, with others set to follow. With complementary light rail and suburban rail also added to network maps, this high quantity of information comes at a potential usability cost. For example, with a printed size of just 20 cm by 20 cm, the current Paris Metro map includes fourteen Metro lines, two Metro shuttles, five RER lines, three peripheral tram routes, four suburban termini and their routes out of Paris, a river bus service, airport shuttle, a short funicular, and around 380 stations. Producing a clear usable map under such circumstances is a major design challenge, but identifying the underlying principles of schematic usability in order to assist the designer is a task for which psychologists can potentially assist. The potential scale of the problem for a complex network can be appreciated when we consider an early usability study (Bronzaft, Dobrow, & O'Hanlon, 1976) in which every single one of twenty novice participants made at least one error when using the then-current New York Subway schematic map to plan a series of journeys.

Most research into the cognitive psychology of map design has focused on topographical maps and perceptual/psychophysical issues, for example symbol discriminability and interpretation (e.g., Montello, 2002; Phillips & Noyes, 1982; Phillips et al., 1990). Another topographical theme concerns how, for example, journeys by car or on foot are planned and sketched (e.g., Tversky & Lee, 1999). The task demands of planning a metro journey from station to station, or communicating one, differs from this in many important ways: Precise directions (e.g., go straight ahead, take the second left, then turn right at the third set of traffic lights) are irrelevant, and instead a sequence of operations more akin to a computer program must be devised (e.g., take Line 1 towards La Défense, change to Line 2 at Nation, change to Line 11 at Belleville heading towards Mairie des Lilas). Spatial descriptions become mere flags, necessary to indicate only the broad direction of travel, and even when overtly spatial terms are used (e.g., *Northbound* Victoria Line) these are effectively only keywords to be identified on signage: After negotiating a maze of passages, the platforms for different directions of travel will be otherwise indiscriminable. As such, the task of creating good network schematics presents its own unique problems compared with topographical maps, and the literature on the latter offers little assistance other than very general suggestions. For example, Tversky and Morrison (2002) highlight the *apprehension principle*: “the structure and content of the external representation should be readily and accurately perceived and comprehended” (pp. 255-256) but without specific proposals for how to apply this principle to static graphics.

Specific to schematic map design for transport networks, there have been numerous suggestions for good practice (e.g., see Ovenden, 2008, p. 151 for many possibilities, and also Avelar, 2007; Nöllenburg & Wolff, 2009; Stott & Rodgers, 2005; Roberts, 2005, ch. 9; Wolff, 2007). Examples include: always use octilinearity; preserve spatial relationships between stations wherever possible; make the x-height of the lettering the same as the point-width of the lines; do not interrupt lines with station names; etc. etc. However, to date there are no published usability studies with the underlying intention of demonstrating these directly, or identifying the psychological reasons why some designs might be more usable than others [2]. Often, where usability between designs is compared, this simply comprises a schematic map versus a topographically accurate one. For example, Bartram (1980) found faster planning for a schematic bus map, but the topographically accurate map had street details included as well as bus routes. The disadvantage for the topographical map might have been due to distracting supplementary information, irrelevant to the set tasks, rather than complex route trajectories. This is in line with Everett, Anderson and Makranczy (1977), who showed that a greater level of detail on pamphlets was associated with more planning errors. One of the more well-known investigations into map design, reported by Bronzaft and Dobrow (1984), led to the demise of the New York Subway schematic map, and the adoption of a design that is clearly a forerunner of today’s broadly topographical version. But even this investigation dispensed with objective usability studies after initially failing to find any clear improvements for new designs, switching to user ratings instead [3].

The most fundamental detail of a transport schematic, what angles should be used in order to simplify the network, has received as little direct empirical or theoretical attention as other suggested principles of good practice. This lack of scrutiny may partly be due to the widespread belief amongst graphic designers, researchers, commentators and users, particularly in Europe, that *octilinearity*, as first used in London in 1933, constitutes some sort of gold standard. In other words, applying these rules will result in the best design possible no matter what the structure of the network (e.g., Ovenden, 2005, p. 39) [4]. Indeed, in many recent attempts to develop computer routines to automate the generation of schematic maps, and optimize these according to usability criteria, only octilinear maps are attempted, although limitations in computer capacity preclude a comparison of different levels of linearity (e.g., Nöllenburg & Wolff, 2009; Wolff, 2007). Empirical evidence in support of this view is very difficult to identify. In some circumstances, there is a memory bias such that diagonal lines tend to be remembered as closer to 45° than reality (Schiano & Tversky, 1992; Tversky & Schiano, 1989) but in these studies mean error never exceeded a few degrees. There is some evidence that people have an octilinear bias (but not necessarily a regular one) in organizing space, but this does not interfere with their perception of this (Klippel & Montello, 2007). Overall, these sorts of finding might explain the tendency towards octilinearity amongst designers, but do not suggest that a non-octilinear map will inevitably cause any difficulties for the user, especially as such maps will be in view while planning takes place, and remembering precise trajectories will not be necessary in order to recall a plan. In circumstances where octilinearity breaks down (see later), any supposed advantages implied will be lost in any case.

The *octilinearity as a gold* standard belief has two consequences, first it discourages attempts to determine whether breaking its rules might result in better designs. For example, Mijksenaar and Roman (1983) proposed a hybrid map. This had a topographical centre (where many stations are within easy walking distance of each other and important tourist attractions) and schematic suburbs (where there are few tourist attractions and few stations within walking distance of each other). This concept was rejected by London Transport despite the research underlying it. Furthermore, there may even be circumstances in which the particular properties of a network defeat octilinearity and result in a less usable map than might have been created had different design rules been used. Second, the assumption that adopting gold-standard rules will result in the best possible map deflects from the issue of whether different implementations that obey the same rules might be differently usable (Roberts, 2009, Newton & Roberts, 2009). Despite the lack of usability studies to guide us, we can nonetheless identify circumstances in which map usability is likely to be poor, and support these predictions theoretically by turning to the cognitive psychology of reasoning.

Why do schematic maps assist users in navigating transport networks compared with topographical maps (assuming that they do)? From this, how can we capitalize on these aspects in order to improve schematic design? A transport schematic is effectively a pre-prepared representation of the underlying structure of the network, meaning that the user does not need to identify this for him or herself (see also Freska, 1999, pp. 8-9). This therefore potentially reduces the cognitive load of the user, along with the associated risk of making errors during the planning process. The benefits of this can be identified from the literature on reasoning and intelligence. For example, all theories of deductive reasoning (e.g., Mental Models theory, Johnson-Laird & Byrne, 1991, and Deduction Rules theory, Braine & O'Brien, 1998) have, as an initial step, the need to identify key elements of the problem and represent them. Pre-represented information can improve performance (Roberts & Sykes, 2005). From the intelligence literature, analyses of non-verbal intelligence test items, such as Raven's Progressive Matrices (e.g., Raven, Raven, & Court, 1993) suggest that the hardest items, for which correct solutions indicate the highest intelligence, are those which are most demanding on working memory capacity: They have many rules and elements to identify, and have rules which require more processing steps to execute (e.g., Carpenter, Just, & Shell, 1990). Overall, the cognitive load of a reasoning task, specifically, its working memory demands, and the cognitive capacity of the individual, determine the likelihood of success (e.g., Stanovich & West, 1998a, 1998b, 2000). More specifically, on the basis that it is hard to reason if it is hard to identify what is to be reasoned about, Roberts, Welfare, Livermore and Theadom (2000) and Meo, Roberts and Marucci (2007) have shown that *element salience* is an important component of test item difficulty: Problems can have the same underlying logic, but if this is concealed via complex, difficult to identify/name shapes and patterns, then items will be harder to solve (see Figure 1). Crucially, therefore, element complexity directly relates to people's ability to identify them, and from this to identify item structure and its underlying logic.

******* INSERT FIGURE 1 ABOUT HERE *******

Roberts (2005, 2009, Newton & Roberts, 2009) argues that planning a journey is more than a simple tracking task. Taking meandering complex trajectories in real life, and converting them to straight lines, should indeed result in simpler routes that are easier to track. But a well-designed schematic, which capitalizes on this opportunity and avoids excessive topographical distortion, will also minimize the cognitive load associated with journey planning, and reveal the underlying structure of the network. Reduced cognitive load alongside increased structural salience increases the opportunity for learning, reducing the cognitive load still further in the future. We would therefore expect fast journey planning, few errors, better remembered plans, and more easily reconstructed plans in the event of a failure to remember. In comparison, a poorly designed schematic, with many unnecessary changes of direction, will not have these benefits, and may even have little to offer compared with a topographic map, other than the simplification entailed in removing street details and most other landmarks. Hence, converting a network representation from topographical to schematic is only beneficial if the meandering curves really have been straightened. If curvature is merely converted into short segments of straight lines linked by many corners, then the original trajectories have not been simplified, instead the shape of the complexity has merely been changed.

From the analysis above, and in the absence of evidence to suggest that octilinearity has a special status in relation to map usability, the task of the designer becomes somewhat more complex. For example, a higher linearity map (say, dodecalinear: horizontal, vertical, 30°, and 60° lines) is likely to offer the possibility of fewer changes of direction, especially for complex networks, but at the expense of the greater number of angles reducing the coherence of the design. Furthermore, the linearity need not be *regular*, in other words, angles of trajectories need not be evenly spaced. For a particular network, the actual trajectories taken by its lines may mean that its structure is better suited to a certain level of linearity and particular angles than other networks. For example, Roberts (2009) has shown that the central area of the London Underground map (inside the Circle Line) may be better suited to a *hexalinear* design than an octilinear one, because fewer changes of direction are necessary in order to show the lines in this key area (six, versus a minimum of nine for an octilinear map). It is also apparent that the octilinear official London Underground map, at the time of writing, has not been well-optimized. The current design, squeezed into dimensions of 21.5 cm by 14.5 cm approx. has no fewer than 14 kinks inside the Circle Line. Preliminary findings (Newton & Roberts, 2009) suggest a clear ordinal relationship between the number of kinks on a design and its usability.

Thus far, the main consideration for good schematic design, within a particular set of rules, appears to be the need to simplify line trajectories and minimize changes of direction (but without distorting the topographical relationships between lines and stations excessively). Breaking the standard octilinearity rules, by using irregular angle spacing, or increasing linearity, or a combination of the two, may enable this goal to be achieved more effectively, but potentially at the cost of the overall coherence of the design. However, for most simple networks, a perfectly adequate design is likely to result if octilinearity is used. There are also the expectations of the user to consider: A map that breaks traditional rules may be met with resistance. For particularly familiar designs, where the image itself has become the mental model of the city (complete with all the distortions induced by the map, see Vertesi, 2008), user-resistance to new designs may be heightened, at least initially, no matter what the potential usability benefits.

In some cases, however, there may be no adequate compromise available with a linear map, no matter what angles are adopted. Paris is an excellent example, with a dense network of interconnected lines, few of which follow anything like a straight-line trajectory in reality. A conventional octilinear design with any longevity was not produced by the RATP, the Paris transport authority, until 2001 (Ovenden, 2008). However, the requirement for a compact design which minimized distortions in spatial relationships between stations has led to one of extreme complexity. For example, Line 4, the busiest Metro line in Paris, has no fewer than sixteen changes of direction from end to end, and there is a mean of ten changes of direction per Metro line. It is far from clear that this offers any degree of simplification, and therefore reduction in cognitive load and assistance to the user, compared with a topographically accurate map. The many changes in direction make line trajectories difficult to follow, and mask what little structure the network has. The limitations of this design, the complexities of this network, and previous non-octilinear attempts to depict it, suggest that it is unlikely that a linear design is possible without a high level of cognitive load facing the user. This led to the current study, in which an alternative Paris Metro map, a departure from traditional linear rules, was investigated.

******* INSERT FIGURE 2 ABOUT HERE *******

The prime test-map in this study is an All-Curves design created by the first author (see Figure 2). The principles underlying this are that if an effective conventional schematic cannot be created, then a non-linear design may be preferable. Hence, rather than numerous short zigzagging straight-line segments, smooth curves may be preferable instead. Instead of changes in direction being minimized, as on a conventional schematic, changes in curvature are minimized instead (see Roberts, 2009). On this design, the trajectories of all lines have been smoothed, and attention has been paid to the orbital lines (2 and 6) which together form a loop within Paris, both as an attempt to simplify the design and to emphasize the underlying structure of the network, hence reducing the cognitive load associated with using it. One further map was included in the current study, an unofficial commercial map available for purchase in the USA. This has straight-line trajectories which are apparently not constrained by an angle rule (technically, *infinite linearity*). It is very compact but its key features are difficult-to-follow line trajectories, and cramped station names, many of which interrupt lines. This map was chosen because it was likely to yield very poor performance, and was included in this study as an attempt to identify the floor-effect level for these tasks that could be expected for a complicated network when a particularly poor design is used. In this way, a baseline might be established against which the benefits of good design might be evaluated.

When comparing the All-Curves map with the official schematic RATP map, it is important to be clear as to what can be concluded if there are differences in usability between the two designs in favour of the All-Curves map. These would establish that this specific All-Curves map is easier to use for journey planning than the current RATP map, but would not show that all-curves designs are superior to conventional schematics in general, or even specifically for Paris: There is always the possibility that a particularly skilled designer could create a conventional schematic that addresses usability requirements well, or that a bad designer could create an All-Curves map with many s-bends, waves, and other changes in curvature. On the other hand, such a difference in usability would conclusively disprove the belief that a competently executed conventional octilinear schematic will always yield the best possible design (e.g. Nöllenburg & Wolff, 2009; Ovenden, 2005, 2008; Wolff, 2007).

For this study, participants were asked to plan various complicated cross-Paris journeys using just one of the three designs. Usability measures included journey planning times, and planning errors (e.g. if a journey included an illegal interchange), although it was not expected that the All-Curves map, nor the RATP map would yield many errors. Journey durations were also estimated primarily as a control variable in order to ensure that, for example, a tendency for faster planning for one map was not associated with longer journeys. Hence, journey planning time was the prime variable of interest.

Many transport undertakings and independent studies seek to assess the designs of maps by obtaining subjective user ratings rather than objective usability measures (e.g., Bronzaft & Dobrow, 1984; Everett et al., 1977). A questionnaire was therefore devised whose main purpose was to investigate how usable participants considered the maps to be, so that ratings could be compared with objective measures. Finally, we investigated choices between maps in order to identify which of these the participants would prefer to use in reality.

The All-Curves map is not a perfect match for the RATP map. It uses the same typeface and line colours, and interchanges are shown in a similar way. The maps were printed so that the surface areas were similar but, as a design intended for tourists, the All-Curves map did not include the three orbital light rail routes on or beyond the periphery of Paris. Also omitted were four suburban commuter routes terminating on the outskirts of Central Paris. These are not prominently shown on the official map and have no actual stops within the City of Paris, therefore they could not be used for journey planning. Seven out-of-station interchanges explicitly identified on the official map were also not shown. For a tourist, using these would involve leaving one station in a likely-to-be unknown locale and searching for another in the vicinity, perhaps relying on local signposting. The commercial map omitted all these details too. The test routes were chosen such that the additional features shown on the RATP map were not necessary in order to plan an effective journey, and hence these differences were not expected to have any impact on overall usability. Participants using the RATP map were free include these as components of their journeys if they wished. No participants attempted a designated out-of-station interchange for any plan, and the use of an orbital light rail route only featured for one particular journey (Porte d'Italie to Garibaldi), where eleven out of forty participants included 'T3' as part of their plan.

METHOD

Participants

120 participants took part in this experiment, forty allocated to each map. These included unpaid psychology students from London South Bank University participating via a departmental reciprocal recruitment scheme, and non-students who were also unpaid volunteers. Additional paid volunteers (£10.00 for participation) were also recruited, both students from London South Bank University and non-students. An additional participant completed the study but was replaced because of health problems leading to difficulties with motor skills. Overall, 17 males planned journeys using the RATP map, 13 using the All-Curves map, and 12 using the Commercial map. The mean age of participants was 38 years for the RATP map (*SD* = 17), 40 years for the All-Curves map (*SD* = 17), and 37 years for the

Commercial map (*SD* = 15). Virtually all participants were unfamiliar with the Paris Metro network.

Materials

Maps for journey planning were printed on A3+ sheets of paper and laminated. They were sized to be comparable in surface area in relation to network coverage: The dimensions of the RATP map were 30.4 x 30.6 cm, the All-Curves map 32.5 x 25.7 cm, and the Commercial map 43.2 x 19.8 cm. One additional clear acetate overlay sheet was also prepared for each map, which highlighted the twelve stations which were to comprise the six start/destination journey pairs. For recording the route planned, participants were supplied with smaller maps (A4 sized, paper) on which they were asked to draw the route chosen. Victor Hugo (Line 2) to Vaneau (Line 10) was given as a practice journey. The five test journeys were:

Hoches (Line 5) to Rennes (Line 12)
Glacière (Line 6) to Richelieu Drouot (Lines 8/9)
Porte d'Italie (Line 7) to Garibaldi (Line 13)
Rambuteau (Line 11) to Convention (Line 12)
Segur (Line 10) to Rue St-Maur (Line 3)

All were intended to be difficult journeys, requiring at least two changes of Metro line and either a trip across the congested centre or, for participants who preferred to avoid more complicated parts of the map, a circuitous route.

A sixteen item questionnaire was also devised in order to obtain subjective ratings of various aspects of the maps, with a seven-point rating scale (*strongly agree* to *strongly disagree* for each statement) other than for questions 15 and 16. Eleven questions were directly related to how usable each participant considered the map to be (asterisked below), the remainder sought qualitative evaluations or attempted to ascertain how the tasks were approached. The full set of question is as follows:

1* I found the journeys easy to plan using this map

2* The routes were difficult to discriminate (identify) using this map

3* The station names were easy to identify in 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* With this map design, I would rather walk or take a Taxi than use the Metro

9 I preferred to look for a direct route no matter how many interchanges were required

10 Some parts of the map looked confusing, and I planned journeys that avoided them

11 This map is intended for planning journeys, but I think that it is also probably geographically accurate

12* I found the map visually ‘disturbing’

13* I found the map cluttered

14* I would look for another design of Paris Metro map to use at the earliest opportunity

15 What aspect(s) of the map did you like the most?

16 What aspect(s) of the map did you like the least?

Design

Primarily, this was an independent groups design with Map Type (three levels, RATP, All-Curves, Commercial) as the independent variable. Measures of map performance included the time taken to plan a journey, planning errors, and an estimation of the duration that the planned journeys would have taken had they been implemented. Questionnaire scores provided a means of measuring people’s subjective assessments of map usability.

Procedure

Participants 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 Metro journeys using the supplied map. 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. They were also informed that they should only change between lines at designated interchanges shown on the map.

Participants were given the opportunity to view the map while the initial instructions were given, and the relatively easy practice journey allowed further familiarization. There then followed the five test journeys (presented in the order shown in the materials section). Each trial commenced with the experimenter placing the acetate sheet indicating all start and end stations onto the laminated map, then pointing to the particular start station and announcing its name twice, followed by the destination station. The overlay was then removed leaving just the map. Timing commenced and the participant was required to plan the journey as requested, using a dry-wipe marker to assist with planning and memory. Verbal reminders of start or end station were given by the experimenter if requested. Once satisfied with the plan, a verbal announcement was made by the participant, timing stopped, and the final chosen route was transcribed onto an A4 paper map, overseen by the experimenter to ensure accuracy. Following this, the experimenter cleaned all marks from the laminated map and the next trial commenced.

Once all journeys were planned, participants then completed the questionnaire (with the relevant map out of sight). After this, participants were shown the two maps that had not been previously used, and asked which of these new ones they would prefer if they were to repeat the experiment. Finally, they were asked to consider the same question for all three maps, hence indicating whether they would prefer to stay with the map already used, or switch to one of the new maps. For both choice tasks, they were asked to consider the actual journeys that they had planned in order to make their decisions.

RESULTS

Usability Measures

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 averages over the variable interchange quality within most metro networks, which is virtually impossible to communicate via maps. For all participants, means of the five journeys were calculated for each measure of performance. For planning times, variability was found to be high, and therefore reciprocals (i.e. journey planning *speeds*) were also calculated. After this transformation, *higher* scores indicated greater journey planning speeds and hence *better* performance.

When averaging estimated journey durations, there are difficulties where participants made errors (these tended to be illegal interchanges). Deleting these and averaging over the remainder is problematic because there is a tendency for more errors to be made with journeys that have a longer estimated duration. If errors are not randomly distributed between maps, disregarding errors would improve mean journey duration estimates for maps in which the most errors were made. As a solution to this, where a participant made an error, 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 planning error would lead to an extended journey, possibly considerably so. Mean performance by map and by journey are shown in Tables 1 and 2.

******* INSERT TABLE 1 ABOUT HERE *******

******* INSERT TABLE 2 ABOUT HERE *******

The effects of Map Type on the usability measures were analysed using single factor between-subjects Analyses of Variance. The effect of Map Type on planning time approached significance, $F(2,117) = 2.84$, MSe = 780, $p = .06$, but there was a clear significant effect when planning speed was analysed, $F(2,117) = 4.75$, MSe = 0.530, $p = .01$. Post-hoc

Newman–Keuls tests showed that planning for the All-Curves map was significantly faster than for the other two, $p < .05$. For estimated journey duration (with 90th percentile substitutions for errors), there was also a significant effect, $F(2,117) = 5.94$, MSe = 16.6, $p < .01$. Post-hoc Newman–Keuls tests showed that significantly (albeit slightly) faster journeys were planned with both the RATP map and the All-Curves map compared with the Commercial map ($p < .01$ and $p < .05$ respectively). Finally, the differences in errors were also significant, $F(2,117) = 3.74$, MSe = 172, $p < .05$, with the Newman–Keuls test showing that the All-Curves map yielded significantly fewer errors than the Commercial map, $p < .05$. Across all maps, there was no correlation between age and any of the usability measures (greatest $r = .16$ for planning time, $p > .05$), nor within maps (greatest $r = .24$ for RATP map planning time, $p > .05$). There was also no evidence for sex differences: Using 2x3 fully-between ANOVAs, all F values < 1 for the main effects of Sex and also the Sex x Map Type interactions.

Overall, the All-Curves map was faster for planning (around 30% better than both alternatives), and journeys were relatively error-free compared with the Commercial map. The RATP map was slower for planning. The Commercial map was also slower for planning (equal to the RATP map), the worst for planning errors (compared with the All-Curves map), and yielded journeys that were, on average, the longest duration compared with both alternatives. The overall rank ordering was confirmed using pairwise MANOVAs with percentage error, planning speed, and estimated journey duration (with 90th percentile substitutions for errors) as the dependent variables. For the All-Curves map versus the RATP map, Wilks' Lambda = .862, $F(3,76) = 4.05$, $p = .01$. For the All-Curves map versus the Commercial map, Wilks' Lambda = .796, $F(3,76) = 6.48$, $p < .01$. For the RATP map versus Commercial map, Wilks' Lambda = .872, $F(3,76) = 3.73$, $p < .05$.

Questionnaire Measures

Most of the questions were answered using a seven-point scale. Initially, these ratings were analysed on an individual question by question basis. Wherever there were significant differences of Map Type, these were almost always because of adverse ratings for the

Commercial map. The only question where the RATP map differed from the All-Curves map concerned geographical accuracy, where the All-Curves map (mean rating 4.5/7, SD 1.6) was rated as being significantly more likely to be geographically accurate than the RATP map (mean rating 3.6/7, SD 1.7), $F(1,78) = 5.93$, MSe = 2.58, $p < .05$. No other pairwise comparison of ratings between the two maps approached significance, no *F* value exceeded 1.

Questions 9 and 10 were intended to gauge whether differences in the maps had led to rated differences in planning behavior, either by avoiding interchanges, or by avoiding certain parts of a map altogether. There were no significant differences comparing the three maps on either measure, $F < 1$ in each case.

As stated earlier, eleven of the questions could be grouped together because they directly requested people to rate their map on design aspects that were directly associated with usability. The scores for these questions were combined (utilizing the seven point ratings, with answers reversed as necessary to ensure consistency of direction) in order to create a composite usability scale with totals ranging from 11 to 77: High scores reflect a positive evaluation of the map, and a score of 44 corresponds to the mid point. Cronbach's Alpha yielded a value of .91, which could not be increased by deleting any of the items, indicating a very high degree of internal consistency/coherence amongst them. Similarly, a Principle Components analysis yielded a first factor which accounted for 70% of the variance, with high loadings from all individual questions, .52 or greater.

The three maps significantly differed in their aggregate usability ratings (see Table 1), $F(2,117)$, = 14.5, MSe = 151, $p < .01$. Newman–Keuls tests showed that both the RATP and All-Curves maps had ratings that significantly exceeded that of the Commercial map, $p < .01$. However, these subjective usability ratings did not correlate with any of the objective usability measures. The greatest correlation was between aggregate ratings and planning errors, $r = -.10$, $p > .05$. Hence, although subjective usability is a coherent measure, it is not related to objective usability in any obvious way.

Map Choice

For the final part of the experiment, participants were shown the alternatives to the map that they had used for planning, and were asked whether they would like to switch from the one that had just been used to an alternative. In general, the Commercial map was unpopular: It was *always* rejected by people who had used it, and *never* chosen in preference to one that had been experienced. However, if we focus on the RATP and All-Curves maps, there is no clear pattern. 24/40 RATP map users rejected their map in favour of the All-Curves alternative, and 17/40 All-Curves map users rejected their map, Chi-square = 2.45, $p > .05$.

We can investigate whether any objective measures of usability underlie these decisions. In theory, as a group, if people are behaving rationally, then it would be expected that those people who reject a map are those who have relative difficulty using it. Hence, people who switch from a map to its alternative should have worse performance at a usability measure than people who elect to keep the same map. In other words, there should be a significant main effect of Map Choice and, ideally, no interaction between Map Type and Map Choice. This hypothesis was investigated using a series 2x2 fully between ANOVAs (Map Type: RATP versus All-Curves, and Map Choice: Switch versus Keep). The dependent variables analysed were: mean planning time; mean planning speed; mean estimated journey duration (with 90th percentile substitutions for errors); percentage error; and aggregate usability rating (see Table 3).

******* INSERT TABLE 3 ABOUT HERE *******

There were no significant interactions between Map Type and Map Choice, all $F < 1$. There were also no significant main effects of Map Choice (all $F < 1$) with one exception: There was a significant main effect for aggregate usability rating, $F(1,76)$, = 10.5, MSe = 150, $p < .01$. People who elected to keep a map had previously tended to give it a higher usability rating (mean = 60.6, *SD* = 10.8) than people who elected to switch from the map (mean = 52.0, *SD* = 13.4).

In line with previous patterns of data, there were significant main effects of Map Type for mean planning time ($F(1,76)$, = 4.96, MSe = 837, $p < .05$) and mean planning speed ($F(1,76)$, = 5.44 MSe = 0.678, $p < .05$), but not for estimated journey duration ($F(1,76)$, = 2.18, MSe = 13.2, $p > .05$), percentage error ($F(1,76)$, = 3.78, MSe = 107, $p > .05$) or aggregate usability rating ($F(1,76)$, = 0.50, MSe = 150, $p > .05$). Overall, there was little evidence for Map Choice rationality in terms of objective usability: People's preferences were not a function of whether a map was actually relatively easy or difficult for them to use. However, despite a lack of relationship between subjective aggregate usability rating and objective measures, the subjective rating does relate to people's behavior in at least one important way, whether or not they are prepared to persist with the use of a map.

DISCUSSION

Looking at the usability of the three maps, there are clear differences. The All-Curves design performed the best of the three, with around 30% faster journey planning compared with the rest, as well as the fewest planning errors. The official RATP map was the next best, although surprisingly close in terms of user performance to the Commercial map. The main difference between the RATP map and the Commercial map was that the latter was associated with longer estimated journey durations, an average of two minutes per journey. Although this is a reliable difference, it is certainly not substantial, but it is a reasonable hypothesis that different maps of the same network may encourage different journey patterns, either by the visual appearance of the directness of routes, or by making certain parts of the map look more or less formidable to navigate.

It should be noted that journey planning took place under ideal conditions. Although participants knew that they were being timed, there was no pressure to act quickly because of an imminently arriving train, no distractions, and no stress (for example owing to an urgent appointment, service problems, or both). It is also reasonable to suggest that under more realistic planning conditions, proportional differences between maps might be amplified, particularly for errors of planning.

What can be concluded from these findings in terms of schematic map design? It is clear from the MANOVA findings that one particular map here outperformed the rest, but in terms of more general findings about how maps should be designed, we can say that there is no evidence in support of the view that octilinear schematic rules (which the RATP map conforms to) form some sort of gold standard, guaranteeing the best possible design. Indeed, in the case of a complex interconnected network such as Paris, adherence to octilinearity may prevent an effective map from being created due to its inherent constraints leading to numerous changes of direction and preventing a genuine simplification of the network depiction.

Overall, the suggestion here is that no set of rules could enjoy a universal gold-standard status. Instead, the task of the designer is to identify those that best match the properties of the network, such that the map can be optimized: Kinks should be minimized, but not at the expense of excessive topographical distortion, or loss of coherence because the level of linearity adopted is too high. In this sense, although the rules for the map must be a good match for the network, as long as there are few kinks, a low degree of geographical distortion, and a coherent design is produced overall, it does not matter what rules are adopted provided that these goals satisfied.

A departure from linearity via an All-Curves design is clearly more controversial, although this approach is by no means only suited to Paris (Roberts, 2009). Focusing on Paris, ever since the opening of the first RER lines in the 1970s, the network has proved extremely difficult to map, with considerable design instability (Ovenden, 2008). Its properties certainly lend themselves to the All-Curves approach, and although it is possible to find octilinear commercial Paris Metro maps which have fewer changes of direction than the official version, these often have considerable topographical distortion, or have errors, for example in the configuration of interchange stations. This, of course, cannot rule out the production of an outstanding octilinear design in the future. Likewise, the All-Curves approach is open to abuse in the hands of a designer who does not understand how to optimize a map or why (Roberts, 2009). Overall though, the All-Curves design is a considerable improvement over the current RATP design, and on average its user ratings were not adverse. It certainly never should be taken for granted that an official map is the best possible, even when issued by an organization with a reputation for design quality.

Turning to subjective usability ratings, overall, these were at first sight very weakly related to objective measures of planning, which is surprising but not unprecedented in psychology. For example, people are often poor at rating their own performance, or the relative effectiveness of different strategies (Dierckx & Vandierendonck, 2005; Roberts, Taylor, & Newton, 2007). This certainly should sound a note of caution to people attempting to assess usability without conducting usability studies, but there are three interesting caveats to this.

First, we should note that when participants were choosing between maps, they had only directly experienced one of the options. It is therefore important to compare subjective and objective measures via repeated measures designs in the future.

Second, we should note that the subjective measures were not completely divorced from objective usability. All participants rejected the Commercial map here, preferring one of the alternatives offered. This was the worst map in terms of planning performance, and it received the lowest mean rating along with the lowest acceptance rate by far. However, we should exercise some caution here. Although this map was objectively the worst, the gap between this and the second-worst map in terms of usability was smaller than the gap between the second and the best maps. It is possible that participants rejected this map on the basis of some sort of aesthetic preference, rather than an awareness that it was difficult to use.

Finally, even if the measure of subjective usability is actually a measure of *map engagement*; how aesthetic the design appears in the absence of any consideration of its actual usability, this is still an important variable. Subjective usability was related to whether participants wished to *persist* with using a map, an important consideration when designing information provision: If it is rejected by customers, then it has failed no matter how easy it is to use. However, other than rejecting the Commercial map, it is difficult to see how these issues could be addressed comprehensively. There were no clear patterns of preference between the

RATP and All-Curves maps, and some of the participants' comments were illuminating in this respect, for example explicitly rejecting the All-Curves design because it differed from the rules that they were used to, or how they expected a well-designed map to look. Clearly, not everyone held these views (comparing maps, the mean aggregate rating scores were almost identical), which suggests that to cater properly for map engagement, it would be necessary to provide for individual differences. However, unlike alternatives to the QWERTY keyboard design, familiarity with traditional schematic maps did not preclude the use of the All-Curves design, and so there need be no barriers to the introduction of radical solutions provided that these really are supported by positive objective usability data. Overall, another useful next step for this research would be to conduct the study on Paris experts versus Paris novices, and also expert users of public transport in general versus novice users (probably not possible within the same study) so that the relationships between specific knowledge, general preconceptions, map ratings, and map usability can be fully investigated, and to ensure that unusual designs that improve novice performance do not have an adverse effect on expert planning.

FOOTNOTES

1. Of course, in order to entice people to use public transport, and enable them to do so effectively, maps are just one component of the wayfinding package. However, maps do have unique features: their complexity; their ability to be taken home and used for planning; and their general publicity value. This justifies their separate treatment from direction signage etc. when investigating usability.

2. The focus of the current manuscript will be on the configurations of line trajectories. Issues concerning how best to group services and colour-code them, and how to show interchanges and other supplementary information, add further complexity to the problem of designing good schematic maps and will not be addressed here. For examples of the bewildering range of solutions from around the world, see Avelar and Hurni (2006), Ovenden (2005), and -- just for Paris -- Ovenden (2008). Suffice it to say that on many official maps, such information is all too often unclear or ambiguous (for example, see Roberts, 2008).

3. The New York Subway has many unique features that make the creation of clear maps a challenge for the designer, and successful navigation of the network a challenge for the user. These include very different service patterns at peak hours, off-peak, weekends and nights; express and local trains, so that many stations may be bypassed; and (because of the convention of naming stations after intersecting streets) identically named stations on different lines in different locations. For example, there are five separate “23rd Street” stations in Manhattan. In trying to create an effective map, the designer should always remember that *a good map can’t fix a bad network.*

4. At the time of writing, of the eight complex networks listed earlier, the only exceptions to octilinearity are the New York Subway and Mexico City Metro. Other notable exceptions in America include the Chicago Transit Authority network, Washington DC Metro, and the Bay Area Rapid Transit system (BART). In Europe, the Barcelona Metro is one of the few sizeable and mature networks not to adhere to octilinearity; the network is shown as straight line routes following a topographical street plan, inevitably resulting in irregular linearity. Away

from these networks, the illustrations in Ovenden (2005) indicate that, historically, networks tend to commence with topographical maps, and then develop octilinear schematics as they grow and mature, sometimes with irregular linearity as an intermediate step. The overwhelming majority of network maps are therefore either octilinear, or topographical.

REFERENCES

Avelar, S. (2007). Convergence analysis and quality criteria for an iterative schematization of networks. *Geoinformatica*, *11*, 497–513.

Avelar, S., & Hurni, L. (2006). On the design of schematic transport maps. *Cartographica*, *41*, 217–228.

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

Berendt, B., Rauh, R., & Barkowsky, T. (1998). Spatial thinking with geographic maps: An empirical study. In H. Czap, H. P. Ohly, & S. Pribbenow (Eds.), *Herausforderungen an die Wissensorganisation: Visualisierung, multimediale Dokumente, Internetstrukturen* (pp. 63-73). Würzburg: Ergon.

Braine, M. D. S., & O'Brien, D. P. (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

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

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

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

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.

Dow, A. (2005). *Telling the passenger where to get off*. Harrow Weald, UK: Capital Transport Publishing.

Everett, P. B., Anderson, V. B., & Makranczy, U. (1977). Transit route pamphlets: Do they work? *Transit Journal*, 59–70.

Freska, C. (1999). Spatial aspects of task-specific wayfinding maps: A representation-theoretic perspective. In J. S. Gero and B. Tversky (Eds.), *Visual and spatial reasoning in design* (pp. 15–32). Sydney: Key Centre of Design Computing and Cognition, University of Sydney.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove: Psychology Press.

Klippel, A., & Montello, D. R. (2007). Linguistic and nonlinguistic turn direction concepts. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds.), *Spatial information theory* (pp. 354-372). Berlin: Springer.

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

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

Montello, D. R. (2002). Cognitive map-design research in the twentieth century: Theoretical and empirical approaches. *Cartography and Geographic Information Science*, *29*, 283–304.

Newton, E. J., & Roberts, M. J. (2009). Applying cognitive psychology principles to Underground map design. Poster presented at London South Bank University research day, July.

Nöllenburg, M., & Wolff, A. (2009). Drawing and labeling high-quality metro maps by mixed-integer programming. Unpublished manuscript under review.

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

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

Phillips, R. J., Coe, B., Kono, E., Knapp, J., Barrett, S., Wiseman, G. And Eveleigh, P. (1990). An experimental approach to the design of cartographic symbols. *Applied Cognitive Psychology*, *4*, 485–497.

Phillips, R. J., & Noyes, L. (1982). An investigation of visual clutter in the topographic base of a geological map. *The Cartographic Journal*, *19*, 122–132.

Raven, J., Raven, J. C., & Court, J. H. (1993). *Manual for Raven's Progressive Matrices and Mill Hill Vocabulary Scales*. Oxford: Oxford Psychologists Press.

Roberts, M. J. (2005). *Underground maps after Beck*. Harrow Weald, UK: Capital Transport Publishing.

Roberts, M. J., (2008). Information pollution on the Underground Map. www.tubemapcentral.com/information_pollution/ip.html

Roberts, M. J. (2009). Henry Beck rules, not OK? Breaking the Rules of Diagrammatic Map design. Unpublished manuscript under review.

Roberts, M. J., & Sykes, E. D. A. (2005). Categorical reasoning from multiple diagrams. *Quarterly Journal of Experimental Psychology*, *58A*, 333–376.

Roberts, M. J., Taylor, R. J., & Newton, E. J. (2007). Explaining inappropriate strategy selection in a simple reasoning task. *British Journal of Psychology*, *98*, 627–644.

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

Schiano, D., & Tversky, B. (1992). Structure and strategy in viewing simple graphs. *Memory and Cognition*, *20*, 12-20.

Stanovich, K. E., & West, R. F. (1998a). Cognitive ability and variation in selection task performance. *Thinking and Reasoning*, *4*, 193–231.

Stanovich, K. E., & West, R. F. (1998b). Individual differences in rational thought. *Journal of Experimental Psychology*: *General*, *127*, 161–188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–665.

Stott, J. M., & Rodgers, P. (2005). Automatic metro map design techniques. XXII International Cartographic Conference, A Coruña, July.

Tversky, B., & Lee, P. U. (1999). Pictorial and verbal tools for conveying routes. In C. Freksa & D. M. Mark (Eds.), *Spatial information theory: Cognitive and computational foundations of geographic information science* (pp. 51-64). Berlin: Springer.

Tversky, B., Morrison, J. B. & Betrancourt, M (2002). Animation: Can it facilitate? International Journal of Human Computer Studies. *International Journal of Human Computer Studies*, *57*, 247-262.

Tversky, B., & Schiano, D. (1989). Perceptual and conceptual factors in distortions in memory for maps and graphs. *Journal of Experimental Psychology: General*, *118*, 387-398.

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

Wolff, A. (2007). Drawing subway maps: A survey. *Informatik*, *22*, 23–44.

Table 1. Mean usability measures by Map Type, also mean usability ratings.

		RATP	All-Curves	Commercial
Planning time (seconds per journey)	Mean	66.3	52.4	63.9
	SD	28.5	28.9	26.3
Planning speed (journeys per minute)	Mean	1.27	1.69	1.25
	SD	0.55	1.01	0.51
Estimated journey duration, errors deleted (minutes)	Mean	57.4	59.9	60.1
	SD	3.4	4.2	5.1
Estimated journey duration, errors substituted (minutes)	Mean	58.9	60.2	62.1
	SD	3.4	3.8	4.9
Errors (percent)	Mean	6.5	2.0	10.0
	SD	13.1	6.1	17.5
Aggregate usability rating (11 to 77, high scores better)	Mean	56.4	56.0	43.4
	SD	12.9	13.0	10.9

Table 2. Mean usability measures by journey.

		Hoches Rennes	Glacière Richelieu Drouot	Porte d'Italie Garibaldi	Rambuteau Convention	Segur Rue St-Maur
Planning time (seconds per journey)	Mean	78.3	58.2	63.1	50.4	56.1
	SD	43.6	36.1	42.1	29.2	30.6
Planning speed (journeys per minute)	Mean	1.02	1.46	1.48	1.62	1.45
	SD	0.58	0.92	1.19	0.96	0.93
Estimated journey duration, errors deleted (minutes)	Mean	69.2	51.5	67.8	51.2	58.3
	SD	10.3	6.2	8.0	5.5	8.7
Estimated journey duration, errors substituted (minutes)	Mean	71.7	51.9	68.4	51.2	58.8
	SD	10.9	6.2	8.2	5.5	9.1
Errors (percent)		16.7	6.7	5.0	0	2.5

Table 3. Usability measures as a function of Map Type and Map Choice.

		Switch	Switch	Keep	Keep
		RATP	All-Curves	RATP	All-Curves
		(*N* = 24)	(*N* = 17)	(*N* = 16)	(*N* = 23)
Planning time (seconds per journey)	Mean	63.0	53.2	71.3	51.8
	SD	27.7	31.2	29.8	27.9
Planning speed (journeys per minute)	Mean	1.34	1.67	1.17	1.71
	SD	0.55	0.87	0.55	1.13
Estimated journey duration, errors substituted (minutes)	Mean	59.2	59.7	58.6	60.5
	SD	3.8	3.3	2.9	4.1
Errors (percent)	Mean	6.7	1.2	6.3	2.6
	SD	15.2	4.9	9.6	6.9
Aggregate usability rating (11 to 77, high scores better)	Mean	52.1	51.8	62.8	59.1
	SD	13.7	13.4	8.6	12.0

Figure 1. Examples of matrix items with complex, difficult to name/identify shapes/patterns (upper) and simple, easy to identify/name shapes/patterns (lower). The underlying logic for these is the same, each combining one *constant in a row* rule with two *distribution of three values* rules (Carpenter, Just, & Shell, 1990). Overlapped elements increase complexity, and therefore difficulty, still further.

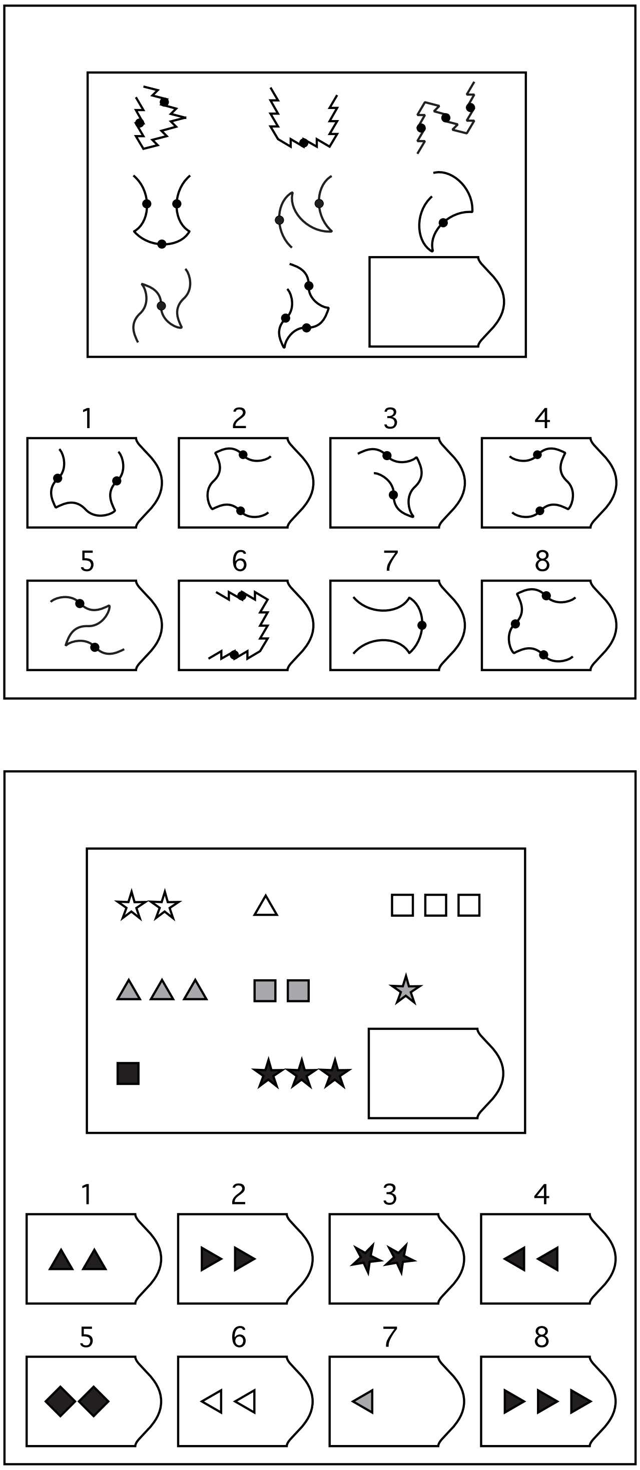

Figure 2. Monochrome image of the All-Curves map used in this study. © Maxwell J. Roberts, 2007. Reproduced with permission. The official RATP Paris Metro map can be downloaded from www.ratp.fr and a section of the Commercial map is detailed at http://www.streetwisemaps.com/metro-map/paris-metro-map.html