Objective versus Subjective Measures of Paris Metro Map Usability: Investigating Traditional Octolinear versus All-Curves Schematic Maps

Maxwell J. Roberts,

Department of Psychology, University of Essex, UK

Elizabeth J. Newton, **Fabio D. Lagattolla**

Department of Psychology, London South Bank University, UK

Simon Hughes, & Megan C. Hasler

Department of Psychology, University of Essex, UK

RUNNING HEAD: METRO MAP USABILITY

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 versus Subjective Measures of Paris Metro Map Usability: Investigating Traditional Octolinear versus All-Curves Schematic Maps

ABSTRACT

Schematic maps are an important component of assistance for navigating transport networks worldwide. 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 directly 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 RATP Paris Metro map with an All-Curves design which replaces straight lines and corners with gentle curves. Three separate usability studies with slightly different methodologies showed that the journey planning time for the All-Curves map was better then the RATP version, with effect sizes ranging from 0.48 to 1.12. Subjective usability ratings were derived from questionnaires, and user preferences, but neither were correlated with objective usability measures. We conclude that (1) in terms of designing schematics, there is no evidence to suggest that any rule-set 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, Cognitive load, Reasoning

Objective versus Subjective Measures of Paris Metro Map Usability: Investigating Traditional Octolinear versus All-Curves Schematic Maps

1. INTRODUCTION

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, 2008a). 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 *octolinearity* (Nöllenburg & Wolff, 2011; Wolff, 2007): In other words, at any point on a line, only eight different trajectories are possible [2]. Applying these design rules almost always results in topographical distortion, especially if the map is designed so that station names do not interrupt lines. This usually requires an expanded centre at the expense of more sparsely-served outer areas, which also balances the design in terms of data density.

1.1. General usability issues of schematic 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, with ten major trunk routes or more, 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 rail 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, and identifying the underlying principles and cognitive theory of schematic usability in order to assist the designer is a task for which psychologists can potentially assist. The 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 spatial 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 or a recipe 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, Morrison and Betrancourt (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, 2011; Stott, Rodgers, Martinez-Ovando, & Walker, 2011; Roberts, 2008a, ch. 9, 2012; Wolff, 2007). Examples include: always use octolinearity; 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. 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 [3]. Often, where usability between designs is compared, this simply comprises a schematic map versus a topographically accurate one. 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 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 [4]. In contrast, Bartram (1980) found faster planning for a schematic bus map, but the topographically accurate version had considerably more detail, i.e. streets as well as bus routes. The disadvantage for the topographical map might have been owing to the considerable quantity of potentially distracting supplementary

information, irrelevant to the set tasks, rather than complex route trajectories. 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.

Another important issue concerns the effects of topographical distortion. Where this is

excessive, users express dissatisfaction with designs, 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, and

Guo (2011) suggests that map configuration can mislead people into planning inefficient

journeys under certain circumstances. Hence, designers generally distort topography only

where this yields clear benefits, such as a balanced design or simplified line trajectories.

1.2. Octolinearity as the basis of schematic maps

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 the 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 (as opposed to the Americas), that *octolinearity*, as first used in London

in 1933, constitutes some sort of design gold standard. In other words, applying this will

result in the best design possible no matter what the structure of the network (e.g., Ovenden,

2005, p. 39) [5]. In recent research to develop computer routines to automate the generation of

schematic maps, and optimize these according to usability criteria, initial attempts did not

achieve octolinear designs (e.g., Hong, Merrick, & do Nascimento, 2005) but this rapidly

became a key objective (e.g., Nöllenburg & Wolff, 2011; Wolff, 2007). Hence, Nöllenburg &

Wolff (2011) describe 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). 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 octolinear 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 findings might explain the tendency towards octolinearity amongst designers, but do not suggest that a non-octolinear 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 any case, in circumstances where octolinearity breaks down (see later), any supposed advantages implied will be lost.

The *octolinearity as a gold standard* belief has two consequences, first it discourages attempts to determine whether alternatives might result in better designs. For example, Mijksenaar and Roman (1983) proposed a hybrid map. This had a topographical centre, where there are many tourist attractions and stations within easy walking distance of each other, and octolinear schematic suburbs, where there are fewer tourist attractions and nearby stations. This concept was rejected by London Transport despite the research underlying it (Roberts, 2008a). Furthermore, there may even be circumstances in which the particular properties of a network are poorly suited to octolinearity and result in a map with more complexity in terms of numbers of kinks than might have been created had different design rules been used (see later). Second, the assumption that adopting a gold-standard will result in the best possible map deflects from the issue of whether different implementations that obey the same design 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 cognitive psychology.

1.3. Cognitive psychology and schematic map optimization

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 reduces task demands and therefore the cognitive load for 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, 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). One aspects of optimizing a schematic map, therefore, is to heighten the salience of the underlying network structure. How should this be achieved?

The effects of item appearance and logical structure on task difficulty have been particularly investigated in the domain of intelligence testing. For example, analyses of non-verbal items, such as Raven's Progressive Matrices (e.g., Raven, Raven, & Court, 1993) show 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 rules which require more processing steps to execute (e.g., Carpenter, Just, & Shell, 1990). In other words, information quantity affects performance. 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).

In terms of schematic maps, quantity of lines and stations is clearly related to cognitive load, but it should also be noted that a line trajectory with many corners contains more information than a straight line. The consequence of this additional information is more than simply making a line harder to follow. In the domain of intelligence testing, Roberts, Welfare,

Livermore and Theadom (2000) and Meo, Roberts and Marucci (2007) have shown that *element salience* is an important component of item difficulty: Problems can have the same underlying logic, but if this is concealed via complex, difficult to identify (or name) shapes and patterns, then items will be harder to solve (see Figure 1). Crucially, therefore, complexity of problem elements directly relates to people's ability to identify them, and from this to identify item structure and underlying logic. In other words, it is hard to reason if it is hard to identify what is to be reasoned about.

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

Roberts (2008a, 2009, 2012, Newton & Roberts, 2009) argues that the simplification of line trajectories is a crucial aspect of optimizing a schematic map. Taking meandering complex trajectories in real life, and converting them to simple straight lines will 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, setting up a virtuous circle and reducing the cognitive load still further in the future. For a well-optimized map, 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. Any supposed benefits for an octolinear layout, as discussed earlier, such as trackability, memorability and reproducibility, will be considerably reduced if the configuration comprises complex sequences of zigzags.

1.4. Departing from octolinearity to improve optimization

From the analyses above, it is proposed that to maximize the benefits of a schematic map, this should be optimized by simplifying line trajectories in order to minimize changes of direction, but without distorting the topographical relationships between lines and stations excessively. In the absence of evidence to suggest that octolinearity has a special status in relation to map usability, then depending on the structure of the network, breaking the octolinearity rules may permit better optimization. However, for most simple networks, a perfectly adequate schematic is likely to result if octolinearity is used. There are also the expectations of the user to consider: A map that breaks with tradition 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. Even so, at the time of writing, the octolinear official London Underground map, may not be well-optimized. The current design, squeezed into dimensions of 21.5 cm by 14.5 cm approx. has no fewer than 12 kinks inside the Circle Line. Findings suggest a clear ordinal relationship between the number of kinks on a design and its usability (Newton & Roberts, 2009).

If departures from octolinearity are permissible, the task of the designer becomes somewhat more complex. Alternative approaches include higher linearity maps (e.g., dodecalinear: horizontal, vertical, 30º, and 60º lines) which almost always offer the possibility of fewer changes of direction, especially for complex networks, but at the expense of the greater number of angles increasing the complexity and reducing the coherence of the design (Roberts, 2012). Alternatively, linearity need not be *regular*, in other words, angles 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 others. A *hexalinear* map has just three angles at 60º intervals. Roberts (2009, 2012) has shown that the central area of the London Underground map (inside the Circle Line) is better suited to this than an octolinear design, 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 octolinear map).

In some cases there may be no adequate compromise available for a linear map, no matter what angles are adopted. Paris is an excellent example, with a dense network of highly interconnected lines, few of which follow anything like a straight-line trajectory in reality. A conventional octolinear 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. Other official straight-line maps of Paris have also been attempted using higher levels of linearity (see Ovenden, 2008) but even these have failed to simplify the network adequately: The use of additional angles adds a new source of complexity and reduces visual coherence, diluting the supposed advantages of using straight lines. The limitations of the designs, both octolinear and higher-order linear, and the complexities of this network, suggest that it is unlikely that a linear design is possible without a high level of cognitive load for 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 *****

1.5. All-curves map of the Paris Metro: design and rationale

For the experiments reported here, an All-Curves design was created manually by the first author using a vector graphics package (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, 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 (see Roberts, 2009, 2012). This translates into using Bezier curves with the following optimization criteria: S-bends and other points of inflexion 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. Specific to the Paris Metro design here, it was intended to be as topographically accurate as the official RATP version (which is by no means perfect in this respect) 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 in an attempt to reduce the cognitive load associated with using it. In terms of information complexity and cognitive load, the suggestion is that a bezier curve, with a minimum number of control points (and hence changes in curvature) comprises less information than short straight-line segments, interrupted by frequent changes of direction. For example, the sixteen changes of direction of Line 4 on the RATP map compared with eight intermediate control points on the All Curves design.

2. GENERAL METHODOLOGY AND EXPERIMENT ONE

For all experiments, participants were asked to plan various complicated (i.e., minimum two interchanges required) cross-Paris journeys using one single Paris Metro map. For a dense network such as this, many journeys will have multiple options, and the two-interchange criterion maximizes competing potential alternative routes, thus increasing the complexity of the planning task in an attempt to maximize effect size. The Paris Metro network is very dense but compact (for example, compared with New York, London, or Berlin). Hence, even cross-Paris journeys by metro are relatively short in terms of distance. Other usability-evaluation methodologies have also been applied, for example intermediate station counting tasks (Stott, Rodgers, Martinez-Ovando, & Walker, 2011) but these are only straightforward to implement and score when there is a small number of route options possible, which

generally confines such methodology to simpler networks. In turn, the relatively easy journey planning that such networks entail means that usability differences between different designs will be harder to identify.

Participants were asked to plan journeys that they would implement themselves if they were required to undertake them in reality, but were not given any explicit optimization instructions, for example to plan the fastest journeys possible. There are three reasons for this. First, information regarding the fine details of navigating a metro network (inter-station journey times, interchange quality, train timing points) are too complex to present on schematic maps without defeating their prime objective of network structure simplification (Roberts, 2008a, Guo, 2011). Without the availability of such information, a requirement to plan the fastest possible journey would be an impossible task, especially as a complex network such as Paris will often have multiple difficult-to-distinguish options for completing the same journey. Second, assuming that there are individual differences in journey preference in real life, it is not appropriate to attempt to eliminate this variability in a study of map usability. For example, as a novice in an unfamiliar city, a person may prefer simply to arrive safely at the destination via a reasonable route, rather than the best route possible. Third, given that the Paris Metro has regions of particular complexity, we wished to give the opportunity for participants to avoid these areas in planning their journeys. For example, on a map, it might appear that a short route is available by changing trains at the complex interchange of Châtelet-Les Halles. An instruction to plan the most efficient journey might lead to the use of this route, but an open specification might result in its avoidance. This would then leave open the possibility of identifying systematic biases in journey choices induced by the different designs. The most obvious possibility in this respect would be participants avoiding central interchanges, preferring roundabout routes by using the orbital Lines; 2 and 6 instead.

For Experiment 1, journey planning was compared using the All-Curves design discussed earlier, the official RATP map, and one further version: an unofficial commercial design of the Paris Metro available for purchase in the USA. This has straight-line trajectories which are

apparently not constrained by any angle rule (technically, *infinite linearity*). It is intended to be 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 likely 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 could be evaluated. For all experiments, participants planned journeys using just one design, preventing the risk of crosstalk between different maps.

When comparing the All-Curves map with the octolinear 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 favor 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 always 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 (as is the case for Madrid, see Roberts, 2012). On the other hand, such a difference in usability would conclusively disprove the belief that a conventional octolinear schematic will always yield the best possible design (e.g., Nöllenburg & Wolff, 2011; Ovenden, 2005, 2008; Wolff, 2007).

The usability measures that were investigated included journey planning times and the percentage of invalid routes (i.e., journeys that included an attempt to change between lines at a point where no interchange was shown) although it was not expected that the All-Curves map, nor the RATP map would yield many invalid routes. Journey durations were also estimated, but primarily as a control variable. Because participants were not requested to plan the most efficient journeys possible, the journey duration rating is not a precise measure of how successfully each map functioned in terms of guiding people efficiently. In any case, this information cannot easily be ascertained without surveying the real journeys. Instead, this is a

control measure that ensures that no map induces biases such that, in comparison with the rest, its effectiveness in terms of journey planning time would need to be qualified. Hence, if any map had a significantly raised overall journey duration measure, this would indicate that the configuration of the design was systematically causing people to travel via inefficient routes. In the absence of journey duration differences, so that a tendency for faster planning for one map was *not* associated with longer journey durations, planning time differences can be interpreted with confidence. 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 purpose was to investigate how usable participants considered the maps to be, so that their ratings could be compared with objective performance measures. Finally, we investigated choices between maps in order to identify which of these the participants would prefer to use in reality.

Comparing the All-Curves map with the RATP map, both use 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. 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 RATP map and have no actual stops within the City of Paris, therefore they could not be used for journey planning. Seven interchanges on the RATP map were explicitly identified as requiring a transfer between separate stations. These were not shown on the All-Curves map. 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

substantive impact on overall usability. Participants using the RATP map were free to 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 using the RATP map. In Experiment 3, the additional features were deleted, with no detectable influence on overall findings.

2.1. Method

2.1.1. Participants

120 participants took part in this experiment, forty randomly allocated to each map. Each map had equal numbers of (1) unpaid psychology students from London South Bank University participating via a departmental reciprocal recruitment scheme, (2) non-students who were also unpaid volunteers, and (3) paid volunteers (£10.00 for participation), either students from London South Bank University or 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 38.1 years for the RATP map ($SD =$ 16.8); 40.1 years for the All-Curves map ($SD = 16.6$); and 37.2 years for the Commercial map ($SD = 15.4$). All participants were residents of London or adjacent counties, the majority were unfamiliar with the Paris Metro network.

2.1.2. Materials

Maps for journey planning were printed on photo-quality 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)

Ségur (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 was 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?

2.1.3. 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, percentage of invalid routes proposed, 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.

2.1.4. 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, using a stop-watch, 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. No feedback was given regarding performance at the tasks in

terms of the route of the journey planned, unless invalid, nor the planning time.

When all journeys were planned, participants then completed the questionnaire (with the

relevant map out of sight). After this, they 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, participants 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 versions. For both of the choice tasks, they were asked to consider

the actual journeys that they had planned in order to make their decisions.

2.2. Results

2.2.1. 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 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). When averaging estimated

journey durations, there are difficulties where participants proposed invalid routes (owing to

illegal interchanges). Deleting these 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 participant 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. On the rating questionnaire, questions 9 and 10 invited participants to indicate their agreement with two journey planning statements. However, there was little evidence to suggest that either of these were related to the journey quality measure across or within maps. The only significant correlation was for the All-Curves map: people who tended to agree with the statement about avoiding confusing parts of the map also tended to have longer estimated-duration journeys, $r = .32$, $p = .043$.

For all participants, means of the five journeys were calculated for each measure of performance. For planning times, the data were positively skewed, and log transformations were found to be the most effective at reducing this (along with variability) across all three experiments. All analyses involving times were therefore performed on group means calculated from log transformed datapoints (natural logarithms were used, i.e. base e). For clarity, group means after transformation but expressed in original units (i.e., geometric means) are also reported. The three usability measures subject to statistical analyses in this and subsequent studies are therefore: (1) log planning time; (2) journey duration with 90th percentile substitutions for invalid routes; and (3) percentage invalid routes. Mean performance by map and individual journeys by map are shown in Tables 1 and 2 respectively.

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

The usability measures were first analysed using two-factor mixed design Analyses of Variance. Map Type (independent groups, three levels) enabled the designs to be compared for overall performance. Journey (repeated measures, five levels, one for each journey) was

included as a factor so that Map Type x Journey interactions could indicate the consistency of the Map Type main effects. The main effects of Journey simply indicate that these differ in their planning time and duration, and so are not reported for simplicity.

The effect of Map Type on log planning time was significant, $F(2,117) = 3.46$, MSe $= 1.06$, $p = .035$, partial eta-squared $= 0.06$. Post-hoc Newman–Keuls tests identified only a significant difference between the All-Curves map and the Commercial map, $p < .05$. No other differences were significant, $p > .05$. With negligible differences between the RATP and Commercial maps, and with a linear contrast comparing the All-Curves mean versus the other two maps combined that yielded a significant effect, $F(1,117) = 6.91$, MSe $= 1.06$, $p = .010$, it is reasonable to conclude that the All-Curves map was the best performer in this respect. For estimated journey duration, there was also a significant effect, $F(2,117) = 5.94$, MSe $= 82.9$, $p = .004$, partial eta-squared $= 0.09$. 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). The All-Curves and RATP maps did not differ, $p > .05$. Finally, the differences in percentage invalid routes was analysed as a single-factor ANOVA, owing to the small number of interchange errors for by individual journeys. Again, the effect of Map Type was significant, $F(2,117) = 3.74$, MSe $= 172$, $p = .027$, partial eta-squared $= 0.06$. A Newman–Keuls test showed that the All-Curves map yielded significantly fewer invalid routes than the Commercial map, $p < .05$. No other differences were significant, $p > .05$.

In terms of consistency of Map Type differences across all five journeys, there was a significant Map Type x Journey interaction for log planning time, $F(8,468) = 5.48$, MSe $= 0.140$, $p < .001$, partial eta-squared $= .08$. Table two shows that the All-Curves map had the best planning time, on average, for four of the five journeys, and the RATP map for the other one. For estimated journey duration, the interaction was also significant, $F(8,468) = 2.32$, MSe $= 60.7$, $p = .019$, partial eta-squared $= 0.04$. Table two shows that the All-Curves map had the best mean journey duration for two of the five journeys, and the RATP map for three of them.

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

Across maps, there was little evidence for correlation between age and the usability measures, the only marginally significant relationship was for log planning time, $r = .18$, $p = .050$, indicating that older people tended to take slightly longer to plan. There was also no evidence for male/female differences: Using 2x3 independent groups ANOVAs, all F values < 1 for the main effects of Male/Female and also the Male/Female x Map Type interactions.

Overall, the All-Curves map was faster for planning than the alternatives, this was associated with a neither significant nor substantial detriment in estimated journey duration. Relatively few invalid routes were proposed 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 invalid routes (compared with the All-Curves map), and yielded journeys that were, on average, the longest duration compared with both alternatives. Aggregating measures, a clear overall rank ordering (AC > RATP > Commercial) was shown using pairwise MANOVAs with the three usability measures as the dependent variables. For the All-Curves map versus the RATP map, Wilks' Lambda $= .858$, $F(3,76) = 4.18$, $p = .009$. For the All-Curves map versus the Commercial map, Wilks' Lambda $= .810$, $F(3,76) = 5.95$, $p = .001$. For the RATP map versus Commercial map, Wilks' Lambda $= .875$, $F(3,76) = 3.62$, $p = .017$. Calculating the effect size of log planning time for the RATP versus the All-Curves map, Cohen's $d = 0.48$, a medium-sized effect.

2.2.2. Questionnaire measures

Most of the questions were answered using a seven-point scale. Initially, 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 $= 2.58$, $p = .017$. No other pairwise

comparison of ratings between these two maps approached significance, no F value exceeded 1. Individual question means and standard deviations by Map Type are given in the Appendix.

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.

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. This created 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 questionnaire ratings (see Table 1), F (2,117), = 14.5, MSe = 151, $p < .001$, partial eta-squared = 0.20. Newman–Keuls tests showed that both the RATP and All-Curves maps had ratings that significantly exceeded that of the Commercial map, $p < .01$. The All-Curves and RATP maps did not differ, $p > .05$. However, these ratings did not correlate with any of the objective usability measures. The greatest correlation was between aggregate ratings and percentage of invalid routes, $r = -.10, p = .281$. Overall, although the aggregate subjective usability rating obtained via questionnaire is a coherent and internally consistent measure, it seems not to be related to objective usability in any prevalent, systematic or useful way.

2.2.3. 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 used to their preferred 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 favor of the All-Curves alternative, and 17/40 All-Curves map users rejected their map, Chi-square = 2.45, *p* = .117.

We can investigate whether the objective measures of usability underlie these decisions. In theory, as a group, if people are behaving rationally, then it would be expected that the 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 Map Type x Map Choice interactions. This hypothesis was investigated using a series of 2x2 independent groups ANOVAs (Map Type: RATP versus All-Curves, and Map Choice: Switch versus Keep) testing the three objective usability measures plus aggregate questionnaire rating (see Table 3).

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

There were no significant Map Type x Map Choice interactions and, with one exception, no significant main effects of Map Choice; all but one *F* < 1. The only significant Map Choice main effect was for aggregate questionnaire rating, *F*(1,76), = 10.5, MSe = 150, *p* = .002, partial eta-squared = 0.12. People who elected to keep a map had previously tended to give it a higher rating (mean = 60.6, *SD* = 10.8) than people who elected to switch from the map (mean = 52.0, *SD* = 13.4). 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 aggregate questionnaire rating and objective measures, the aggregate 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.

2.3. Discussion

Looking at the objective usability measures of the three maps, there are clear differences, particularly taking the MANOVAs into account. The All-Curves design performed the best of the three, with faster journey planning compared with the rest, as well as the fewest invalid routes. The RATP map was the next best, although surprisingly close in terms of 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. The other important finding in Experiment 1 is the lack of association between objective measures of performance (planning time, journey duration, invalid routes) and subjective measures (aggregate questionnaire rating, preference). The lack of evidence for rational preference on the basis of objective usability was particularly striking, but the relationship between preference and questionnaire rating pointed towards the importance of producing maps that people at least *believe* are easy to use. Before interpreting these findings in detail, two further experiments are reported whose purpose was to replicate the key effects with variants on the basic methodology.

3. EXPERIMENT TWO

For Experiment 2, only the RATP and All-Curves maps were tested with the same journeys and materials as before, but using a methodology where a person plans a journey *without* drawing on a map during the process. Instead, participants planned journeys mentally and, when each was completed, this was announced to the experimenter, who then supplied a paper map on which the plan could be transcribed. Such a methodology might be expected to raise the cognitive load of the task, and also separates planning ability from drawing ability. It

carries the risk that some participants may announce completion prematurely, having only partially planned a journey. To some extent, this behavior can be spotted by the experimenter and discouraged, but by recording drawing time as well as planning time, a general systematic strategic trade-off can be identified: partial planning should be associated with rapid planning times and longer drawing times, owing to the need to complete the plan while drawing.

3.1. Method

40 participants took part in this experiment, nineteen planned journeys using the All-Curves map (seven males), and 21 the RATP map (ten males). All were unpaid volunteers; those participants who were students were from the University of Essex. The mean age was 20.8 years for the RATP map ($SD = 1.3$), and 22.2 years for the All-Curves map ($SD = 7.8$). Materials, journeys and questionnaire were identical to the previous experiment. This was now an independent groups design with two levels for Map Type (RATP versus All-Curves). Measures of map performance were identical to previously, except that a planning time and a drawing time were recorded for each journey for each participant.

The procedural differences to the previous experiment were: (1) the order of test journeys was randomized for each individual, but there was no practice trial; (2) each of the journeys was shown to participants using an A4 laminated map, greyed out except for the the start station (indicated with an arrow) and the end station; and (3) participants were asked to plan each journey in their heads. The journey indication map was available throughout the trial, although actual journey planning was via an A3+ laminated map as before. Grey was applied to the journey indication map such that the two stations of interest were effectively circled and highlighted in white. Participants were instructed that once they had planned each journey to their satisfaction, they should announce this to the experimenter, who would hand them an A4 map on which the journey would be drawn. Time was recorded for this announcement, plus the completion of drawing. Participants were requested not to announce prematurely that planning had been completed. After all of the trials had been completed, the questionnaire was

administered as before, and then presentation of the test-map versus the previously unseen map, along with request to choose the preferred one, RATP or All-Curves.

3.2. Results

3.2.1. Usability measures

Data were analysed identically to Experiment 1, the only difference was the availability of separate timing measures for the two phases; planning time and drawing time. Log transformed times were calculated as before, using these values when interpreting effects. Likewise, where invalid routes were proposed, 90th percentile substitutions were performed in order to estimate journey duration. Overall performance by Map Type is shown in Table 1, and Map Type by Journey in Table 4. There was no evidence for any planning–drawing trade-offs. All correlations for planning time against drawing time, overall, and within map type, were positive. For the All-Curves map, $r = .05$, $p = .834$, and for the RATP map, $r = .40$, $p = .073$. Overall, $r = .41$, $p = .009$, indicating a tendency for longer planning time to be associated with longer drawing time, perhaps particularly so for the RATP map. Because of the lack of evidence for trade-offs, only log planning time is analysed here, along with the other two objective performance measures, as for Experiment 1. Identical patterns of results are obtained when drawing time is added to planning time so that total time is analysed. Looking at the estimated journey duration measure, there was again little to suggest strong associations with participants' ratings of journey planning statements (questions 9 and 10). No correlations were significant, although the greatest value was again for the All-Curves map concerning the avoidance of confusing parts of the design, $r = .18$, $p = .720$.

As for Experiment 1, the usability measures were analysed using two-factor mixed design Analyses of Variance. The factors were Map Type (independent groups, two levels) and Journey (repeated measures, five levels, one for each journey). For percentage invalid routes, only the Map Type factor was analysed, as before. The main effect of effect of Map Type on log planning time was significant, $F(1,38) = 12.6$, $MSe = 1.57$, $p = .001$, partial eta-squared =

0.25. For estimated journey duration, there was no significant effect, $F(1,38) = 0.30$, MSe = 177.6, $p = .585$. The difference in percentage invalid routes was also non-significant, $F(1,38) = 0.877$, MSe = 37.0, $p = .355$.

In terms of consistency of Map Type differences across all five journeys, unlike Experiment 1, there was a no significant Map Type x Journey interaction for log planning time, $F(4,152) = 0.947$, MSe = 0.324, $p = .439$. The All-Curves map had a planning time advantage for all five journeys although, interestingly, the journey which had the greatest advantage in Experiment 2, likewise had the greatest advantage in Experiment 1 (Porte d'Italie to Garibaldi) and the journey with the smallest advantage in Experiment 2 (Rambuteau to Convention) was the one where the RATP map had an advantage in Experiment 1. For estimated journey duration, the Map Type x Journey interaction was significant, $F(4,152) = 3.33$, MSe = 68.4, $p = .012$, partial eta-squared = 0.08. This time, four of the five journeys had a shorter estimated duration with the RATP map, the exception (Hoches to Rennes), plus the journey with the smallest advantage for the RATP map (Glacière to Richelieu Drouot) are the same as the two journeys in Experiment 1 where the All-Curves map had an estimated journey duration advantage.

***** **INSERT TABLE 4 ABOUT HERE** *****

Across all maps, there was no correlation between age and any of the usability measures, all $r < .1$. There was again little evidence for male/female differences: Using 2x2 independent groups ANOVAs, all F values < 1 for the main effect of Male/Female and one of the two-way interactions. Two of the Male/Female x Map Type interactions approached significance. For log planning time, $F(1,36) = 3.70$, MSe = 0.295, $p = 0.062$, indicating that the planning time advantage for the All-Curves map over the RATP map was greater for females than males. For estimated journey duration, $F(1,36) = 3.39$, MSe = 33.9, $p = 0.074$, indicating that females tended to plan longer journeys for the All-Curves map than the RATP map, and the reverse for males. Together these interactions imply a strategic trade-off unique to this dataset rather than generalizable effects.

The All-Curves map continued to be faster for journey planning compared with the RATP map, replicating the finding from Experiment 1, with a somewhat more substantial effect size: Cohen's $d = 1.12$. The overall superiority of the All-Curves map was confirmed using MANOVA with the three usability measures as the dependent variables, Wilks' Lambda = .708, $F(3,36) = 4.96$, $p = .006$. However, there appear to be some potentially interesting individual journey effects, which will be discussed after Experiment 3.

3.2.2. Questionnaire measures

The questionnaire was the same as used in Experiment 1 and was scored identically. There were only two individual questions where the maps differed significantly. For question 1 [*I found the journeys easy to plan using this map*] the All-Curves map received significantly better ratings (mean 5.6/7, *SD* 1.1) than the RATP map (mean rating 4.9/7, *SD* 1.1), $F(1,38) = 4.18$, MSe = 1.24, $p = .048$. For question 9 [*I preferred to look for a direct route no matter how many interchanges were required*] the RATP map received significantly stronger ratings in agreement (mean 5.3/7, *SD* 1.3) than the All-Curves map (mean rating 4.2/7, *SD* 1.6), $F(1,38) = 5.38$, MSe = 2.14, $p = .026$. No other pairwise comparison of ratings between the two maps was significant, next greatest F value = 3.17, $p = .083$. Overall, the maps did not differ in their aggregate questionnaire ratings (see Table 1), $F(1,38)$, = 1.86, MSe = 121, $p = .181$. For this study, the question on perceived geographical accuracy yielded no significant difference (All Curves mean rating 4.2/7, *SD* 1.5; RATP mean rating 4.1/7, *SD* 1.5), $F(1,38) = 0.057$, MSe = 2.34, $p = .813$. There was some evidence for a relationship between aggregate ratings and objective usability measures, with a correlation of $r = -.32$, $p = .044$, against log planning time across maps. People who rated maps more favourably tended to be faster at planning. For percentage invalid routes, $r = .22$, $p = .180$, and for estimated journey duration, $r = -.16$, $p = .340$.

3.2.3. Map choice

For aggregate questionnaire score, the All-Curves map received slightly higher ratings overall than the RATP design, but this was not reflected in people's choices. 7/21 RATP map users

rejected their map in favor of the All-Curves alternative, and 11/19 All-Curves users rejected their map, Chi-square = 2.43, p = .119. Looking at whether people's choices were related to objective measures, again there were no significant Map Type x Map Choice interactions and no significant main effects of Map Choice, greatest $F(1,36)$, = 1.75, p = .194 (see Table 5). For aggregate questionnaire rating, however, there was a significant main effect, $F(1,36)$, = 11.1, MSe = 97.4, p = .002, partial eta-squared = .24, and no significant interaction, $F(1,36)$, = 0.041, MSe = 97.4, p = .841 People who elected to keep a map had previously tended to give it a higher rating (mean = 58.9, SD = 10.3) than people who elected to switch from the map (mean = 49.9, SD = 10.3). As for Experiment 1, there was little evidence for Map Choice rationality in terms of objective usability, but this does seem to be related to people's subjective evaluations.

***** INSERT TABLE 5 ABOUT HERE *****

3.3. Discussion

Comparing the RATP and All-Curves maps, Experiment 2 replicated all of the major effects of Experiment 1, for both objective and subjective measures of performance. Indeed, for log planning time, the effect size was greater than previously. Some recurring patterns in differential efficacy between maps were also identified, but these will be discussed after the next experiment.

4. EXPERIMENT THREE

For Experiments 1 and 2, participants were timed for planning their journeys, but other than this there was no particular requirement for rapid planning. For the final experiment, we sought to include an element of time pressure, attempting to mimic a situation in which a person knows from a platform indicator than a train is due soon, and must determine the journey before deciding whether or not to get on the train. Based on previous results, a

deadline of 40 seconds per trial was decided upon. This duration was shorter than the overall mean planning time for both maps in Experiment 1 (see Table 1) and was chosen with the intention of speeding people, but not putting the majority of them in difficulty for every trial. To counter the likely tendency towards partial planning that this methodology could induce, the same drawing/planning methodology as Experiment 1 was adopted, as opposed to separating these components in Experiment 2. In order to encourage participants to attempt to plan quickly, they were informed that a final trial would constitute a 'time-bank'. Initially with zero seconds allocated to it, unused seconds from previous trials would be added to it, but if any previous trials overran, their excess time would be subtracted from the time-bank. To enhance the generalizability of findings, three new test journeys were devised, but to ensure some continuity, two test journeys were retained from Experiments 1 and 2. In terms of the planning time difference between All-Curves and RATP maps, Rambuteau to Convention had always been the worst journey for the All-Curves map, and the estimated journey duration difference was adverse for the All-Curves map in both experiments (the worst in Experiment 1, and the second worst in Experiment 2). Hoches to Rennes had consistently been the best journey for the All-Curves map in terms of estimated journey duration difference, but for planning time difference, it had never been the best journey for the All-Curves map (third best in Experiment 1, second best in Experiment 2). Hence, the journeys carried forward to Experiment 3 were not biased in favor of the All-Curves map.

4.1. Method

4.1.1. Participants

32 participants took part in this experiment, sixteen allocated to each map. All were unpaid volunteers; those participants who were students were from the University of Essex. Overall, 9 males planned journeys using the RATP map, and 7 using the All-Curves map. The mean age was 34.9 years for the RATP map (SD = 16.0), and 35.6 years for the All-Curves map (SD = 15.0).

4.1.2. Materials and design

General materials format was identical to the Experiment 1, but to familiarize participants with the deadline procedure, three practice trials were administered (in the following order): Chemin Vert (Line 8) to Charonne (Line 9); Argentine (Line 1) to Vaneau (Line 10); and Olympiades (Line 14) to Colonel Fabien (Line 2). For the experimental 'deadline' trials, two were retained from the previous experiments (Hoches to Rennes and Rambuteau to Convention). Three new journeys were chosen: Avenue du President Kennedy (RER Line C) to Danube (Line 7b); Pelleport (Line 3b) to Charles Michel (Line 10); and Richard Lenoir (Line 5) to St Francois Xavier (Line 13). The final, 'time-bank' trial comprised Jourdain (Line 11) to Passy (Line 6). All of the new experimental trials were minimum two-interchange cross-Paris journeys, as before. Some modifications were made to the maps in order to align their general qualities more closely. For the RATP map, the following were deleted: the cream background to the central area; the peripheral light rail routes (replaced with interchange flags as appropriate); RER stations outside of Paris; suburban railways and termini; and out-of-station interchanges. For the All-Curves map, line thickness was thinned slightly to bring it in line with the RATP version.

A minor change was made to the questionnaire in order to balance positive and negative statements. Compared with Experiment 1, the following alterations were made:

I found the map clean and uncluttered [previously: I found the map cluttered]
The best routes for me had the fewest station stops along the way [new question]

An additional question asked about frequency of urban rail travel in general, with the following options: every day; a few days every week; a few times every month; about once a month; a few times a year; once a year or less; and never/not for years.

As for Experiment 2, this was now primarily an independent groups design with two levels for Map Type (RATP versus All-Curves). Measures of map performance were identical to Experiment 1.

4.1.3. Procedure

There were three practice trials of increasing complexity, administered in a fixed order (whose performance did not affect the time-bank trial), followed by five experimental deadline trials, randomly ordered for each participant. The time-bank trial was always administered last, and was always the same journey. Each journey was indicated identically to Experiment 2, using a map that was greyed except for start (indicated by an arrow) and end stations. This could remain in view throughout its trial. The basic planning procedure was identical to Experiment 1: Each journey was planned by drawing it on an A3+ laminated map. Once the journey was fully drawn, timing stopped and the journey was transcribed onto an A4 paper map, with the experimenter ensuring that this was accurate. The laminated map was cleaned after each trial. The main difference for Experiment 3 was the use of a deadline methodology. Participants were informed that 40 seconds would be permitted for each journey. If a journey was planned within this, the unused time would be added to that available to plan a final journey (the time-bank trial). If not, then the excess would be subtracted from the time available. Zero time was initially allocated to the time-bank trial, so that sufficient would only be available for it if participants completed at least some of the test trials within the 40 second target. Nonetheless, participants were asked to attempt and complete the time-bank trial even if little or no time was available. To remind participants of the deadline, a recorded verbal countdown commenced at 40 seconds, stating the time remaining at five second intervals, and every second from 20 seconds downwards. Journey planning and timing continued even if the countdown reached zero. For the time-bank trial, participants were informed of how much time was available, and the countdown was modified accordingly. No feedback was given regarding performance at the tasks in terms of the route of the journey planned, unless this was invalid, but the nature of the deadline task meant that this would now yield feedback regarding planning time. After all trials were completed, the questionnaire and the choice task were administered as before.

4.2. Results

4.2.1. Usability measures

Data were processed and analysed identically to Experiment 1, so that mean planning times, estimated journey durations, and percentages of invalid routes were computed, but only for the five deadline journeys. The data for the time-bank journey, Jourdain to Passy, were not included in these means because of the variable time available to participants in order to plan the route. Overall performance by Map Type is shown in Table 1, and Map Type by Journey in Table 6. For the standard trials, 16.3% of journeys for the RATP map exceeded the time available for the deadline trials, and just 1.3% of journeys for the All-Curves map. Four of the 32 participants were unable to complete the time-bank trial because they had insufficient time remaining (all using the RATP map), including two negative times and a positive time of 3 seconds. Their data for the previous five deadline trials are analysed as normal.

The measure of urban rail experience was not correlated with any of the performance measures, all $r < .1$. Looking at the journey duration measure, there was again little to suggest strong associations with participants' ratings of journey planning statements (including the new question for this experiment). No correlations were significant, although the greatest value was again for the All-Curves map, but this time for the question concerning direct routes even if many interchanges were involved: People who agreed with this statement tended to plan journeys with a shorter estimated duration, $r = -.28$, $p = .293$.

As before, the usability measures were analysed using two-factor mixed design Analyses of Variance with Map Type (independent groups, two levels) and Journey (repeated measures, five levels, one for each journey) factors except for percentage invalid routes. The effect of Map Type on log planning time was significant, $F(1,30) = 6.14$, MSe $= 0.571$, $p = .019$, partial eta-squared $= .17$. For estimated journey duration, there was no significant effect, $F(1,30) = 1.83$, MSe $= 61.3$, $p = .186$. The difference in percentage invalid routes was also non-significant, $F(1,30) = 1.11$, MSe $= 45.0$, $p = .300$.

***** INSERT TABLE 6 ABOUT HERE *****

In terms of consistency of Map Type differences across all five journeys, there was a significant Map Type x Journey interaction for log planning time, $F(4,120) = 2.89$, MSe = 0.067, $p = .025$, partial eta-squared = .09. The All-Curves map had a planning time advantage for all five journeys, but the advantages for Pelleport to Charles Michel and Richard Lenoir to St Francois Xavier were particularly strong. Rambuteau to Convention continued to be the journey where the All-Curves map was least advantageous. For estimated journey duration, the Map Type x Journey interaction was not significant, $F(4,120) = 0.95$, MSe = 75.1, $p = .436$. Here, there were no clear patterns, and differences between maps for individual journeys were generally small. The All-Curves map had a slight advantage for four of the five journeys. Hoches to Rennes continued to be the most advantageous journey for the All-Curves map in this respect.

Across all maps, there was no correlation between age and any of the usability measures (all $r < .17$). Looking at male/female differences via 2x2 independent groups ANOVAs, no main effects of Male/Female, nor Male/Female x Map Type interactions approached significance, greatest $F(1,28) = 2.48$, $p = .126$.

The All-Curves map continued to be faster for planning compared with the RATP map, replicating the finding from Experiments 1 and 2, with a substantial effect size: Cohen's $d = 0.86$. However, for this experiment, the overall superiority of the All-Curves map could not be confirmed using MANOVA, Wilks' Lambda = .789, $F(3,28) = 2.49$, $p = .081$.

4.2.2. Questionnaire measures

The modifications to the questionnaire did not affect the overall scoring, with a maximum of 77, a minimum of 11, and a neutral point of 44. The only individual question where the maps differed significantly concerned geography. As for Experiment 1, the All-Curves map (mean 4.8/7, SD 1.4) was rated as being more likely to be geographically accurate than the RATP map (mean rating 3.8/7, SD 0.9), $F(1,30) = 5.87$, MSe = 1.36, $p = .022$. No other pairwise comparison of ratings between the two maps was significant, next greatest F value = 2.55, $p = .121$. Overall, the maps did not differ in their aggregate questionnaire ratings (see Table 1),

$F(1,30), = 0.094$, MSe = 108, p = .761. There were no significant correlations between the aggregate rating and any of the objective performance measures, greatest $r = -.17$. Across the three experiments, there has been no evidence to suggest that subjective measures of usability are related to objective ones in any substantial or consistent way.

4.2.3. Map choice

10/16 RATP map users rejected their map in favor of the All-Curves alternative, and 5/16 All-Curves users rejected their map, Chi-square = 3.14, p = .077. Looking at whether people's choices were related to the objective measures (see Table 7) there were again no significant Map Type x Map Choice interactions and no significant main effects of Map Choice, greatest $F(1,28) = 1.68$, p = .206. Looking at aggregate rating score, the pattern was different to Experiments 1 and 2. The main effect of map choice was not significant, $F(1,28), = 1.41$, MSe = 97.2, p = .246. However, the interaction approached significance, $F(1,28), = 4.04$, MSe = 97.2, p = .054, partial eta-squared = .13. Pairwise comparisons showed that the same effect as previously was clearly present for the All-Curves design, $F(1,28), = 4.90$, MSe = 97.2, p = .035, partial eta-squared = .15, and clearly absent for the RATP map, $F(1,28), = 0.36$, MSe = 97.2, p = .556. Hence, the preference to retain a map that had previously been well-rated, or to reject a map that had previously been poorly rated, only applied for the All-Curves design. The non-significant difference for the RATP map was in the opposite direction.

***** INSERT TABLE 7 ABOUT HERE *****

5. GENERAL DISCUSSION

Across the three experiments, the All-Curves map outperformed the RATP map in terms of mean planning time, with effect sizes ranging from moderate to strong. For two of the three experiments, the All-Curves map was superior taking into account all three measures of performance using MANOVA. Even for Experiment 3, with no significant MANOVA, means for all three objective measures of performance favoured the All-Curves map. There was

some evidence for a slight overall ability effect: faster journey planning was associated with shorter duration journeys, but the correlation was small: looking at mean performance across all three studies, excluding the Commercial map, $r = -.19$, $p = .019$. It is interesting to note that for Experiment 3, mean planning times are considerably shorter in their duration than for Experiment 1. Other than the deadline methodology, both involved a similar planning procedure. This faster planning for Experiment 3 is not associated with any overall increase in invalid routes, and suggests that for tasks without a deadline procedure, times may be raised unnecessarily, perhaps due to participants double-checking their plans or considering more alternatives. Even so, this did not affect the relative usability of the maps, and for the common journeys between Experiments 1 and 3, there is little to suggest that the additional planning time for Experiment 1 resulted in better journeys. However, the time deadline was exceeded for considerably more journeys for the RATP map then the All-Curves map in Experiment 3. Looking across the experiments at other potential variables of interest (excluding the Commercial map), there was a correlation between age and log planning time, $r = .32$, $p < .001$, so that older participants tended to take longer to plan, but there were no male/female differences for any of the planning measures or aggregate rating score, greatest $F(1,150) = 1.9$, $p = .166$.

5.1. Journey patterns

Looking at individual journeys, there are some repeating patterns across studies. In Experiments 1 and 2, Ségur to Rue St Maur and, for all three experiments, Rambuteau to Convention were the least good journeys for the All-Curves map. They had a smaller planning time advantage (a disadvantage in one instance) and also had mean estimated journey durations that favored the RATP map. For these journeys, standard deviations for mean estimated journey duration were always higher for the All-Curves map than the RATP map, indicating that a wider variety of options was being taken, rather than a systematic tendency to explore particular indirect routes. This could also account for the less advantageous planning times, owing to exploring more options and equivocation. For Ségur to Rue St Maur,

it appears that the many kinks along Line 10 on the RATP map discouraged participants from exploring possibilities that looked more direct and reasonable on the All-Curves map, but ultimately yielded longer estimated journey durations. In the case of Rambuteau to Convention, the relative elongation of Line 4 on the All-Curves Map may have discouraged participants from using this as part of their route, which would have yielded the best estimated journey durations by far. In the light of the findings of Guo (2011), it is a reasonable hypothesis that if map configurations differ, then alternative versions of the same network may encourage different journey patterns, either by the visual appearance of the directness and distance of routes, or by making certain parts of the map look more or less formidable to navigate. This may particularly be exacerbated if the map is believed to very accurate topographically, a possibly by-product of an All-Curves design: For Experiments 1 and 3 it was rated by users as being likely to be more accurate on this measure compared with the RATP map. It could be argued that for the two journeys described here, the increase in estimated journey duration for the All-Curves map has wiped out any planning time advantage. However, there are numerous responses to this argument.

First, to confirm these speculations, it would be necessary to survey the actual routes in order to confirm the estimated journey durations, and identify their likely level of variability. Second, the differences in the durations were small in relation to the total durations, and all routes cost the same fare. Tourists would be unlikely to be inconvenienced unduly. Third, if an obvious, single, best-appearing route is difficult to identify from a map, usage can be spread between lines, thus avoiding overcrowding. Returning to the arguments in the previous paragraph, Line 4 is amongst the busiest in Paris – the All-Curves map may have discouraged its use – and Line 10 amongst the least busy – the All-Curves map may have encouraged its use. Fourth, if group variability indicates individual variability, the greater variety of routes attempted for the All-Curves map will encourage a more general network learning and understanding. Fifth, when planning a journey, calculating the many different options and estimating their likely durations would be a very difficult task for a network such as Paris, hence the need to satisfice and find a reasonable route quickly. The only straightforward fail-safe way to communicate true distances, and therefore probable durations, would be via a

genuine topographical map, with all the inconvenience of size for a city such as London, and losing all the potential advantages of a schematic map in terms of the design objectives of clarity, simplicity, cognitive load, and user engagement. Sixth, for many journeys, the All-Curves map had no particular estimated journey duration disadvantage, or else it had an advantage, compared with the RATP map (7/15 journeys across all three experiments), as well as a planning time advantage (14/15 journeys). Hence, although it may be possible to identify individual situations where there might be a tendency to plan less efficient journeys, the overall results should be considered holistically.

For the All-Curves map, Hoches to Rennes always resulted in a better mean estimated journey duration than the RATP map, and Porte d'Italie to Garibaldi resulted in a better All-Curves map planning time for Experiments 1 and 2 (this journey was not used for Experiment 3). For Hoches to Rennes, the potential usefulness of the orbital lines (Line 2 in this case) is made particularly salient by the All-Curves map. For the other journey, Porte d'Italie is on Line 7, which takes a particularly twisted and complex trajectory through Paris, through all three of the most dense regions on the Metro network (Saint Lazarre, Châtelet-Les Halles, and Gare du Nord/Gare de l'Est). For both of these journeys, it would have been particularly inefficient to attempt to avoid these dense regions: at least one of these would have to be negotiated for each. Hence, it is possible that the All-Curves map enjoyed the greatest advantage precisely where the user needs the most help, in making sense of the most difficult interchange complexes.

5.2. Design recommendations

What can be concluded from the findings in terms of schematic map design? Individual analyses of journeys can identify potential troublespots, with possibilities for fine-tuning on both maps. For more general recommendations about how maps should be designed, we can say that there is no evidence in support of the view that octolinearity – which the RATP map conforms to – comprises some sort of gold standard, guaranteeing, or at least being essential

for the best possible design. Indeed, in the case of a complex interconnected network such as Paris, adherence to octolinearity may prevent an effective map from being created. Here, the inherent constraints of octolinearity may lead to numerous changes of direction, preventing a genuine simplification of the network depiction. Hence, schematizing a network in this way may not reduce the cognitive load of journey planning in some circumstances. Overall, the suggestion is that no level of linearity could enjoy a universal gold-standard status. Instead, the task of the designer is to identify the level that best matches 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 linearity must be a good match for the network, as long as there are few kinks, a low degree of topographical distortion, and a visually coherent design is produced, it does not matter exactly what level is adopted provided that these goals can be satisfied.

In terms of automated schematic design, the findings points towards a more involved process than is currently implemented, in which computers design maps at low levels of linearity, and steadily increase this upwards (tetralinear, hexalinear, octolinear, decalinear etc.). The goal would be to evaluate usability scores – derived from number and severity of corners, uniformity of edge length, etc. – across linearity levels. Linearity level could be incorporated as an additional scoring criterion, such that higher levels carried a penalty, on the basis that an increase in the number of angles raises visual complexity. Software would then generate compromise maps in which the various complexity criteria could be balanced. However, high levels of linearity require particular attention to coherence aspects, such as parallel lines and symmetry (Roberts, 2012). These are not implemented by the current generation of automated systems. Roberts (2012) also identifies certain levels of linearity as being inherently unaesthetic and potentially facing user resistance, such as decalinear, because this has no possibility for perpendicular line crossings or regular shapes such as equilateral triangles. However, the results in the current study imply considerable individual differences in respect of user preferences, and if a particular network matched a particular level of linearity well, then this could override such considerations.

The same objective criteria used to evaluate the relative success of linear maps may also be used in order to determine whether a departure from such designs is warranted and All-Curves versions should be investigated. The angle density of the Paris Metro map is high, with around ten corners per Metro line. The equivalent value for the current official Berlin map is three per U-Bahn line (Roberts, 2012) and, for Madrid, seven per Metro Line (but all at 90 degrees). The higher figures imply poorly-optimized maps, and if these proved resistant to improvement irrespective of linearity level, then it could be called into question whether a linear design is appropriate for the network. However, the complexity threshold is yet to be determined, and automated generation of All-Curves schematic maps is currently in its infancy (Roberts, 2012).

A departure from linearity via an All-Curves design is by no means only suited to Paris (Roberts, 2009, 2012) but is controversial from the point of view of potential user-resistance. In terms of user preferences, support for the All-Curves Paris map was somewhat mixed, with 55% of people choosing this rather than the RATP map across all three experiments. There is evidence to suggest that, in general, people have a tendency to prefer curved shapes to angular ones with straight lines (Silvia & Barona, 2009), but this may possibly interact with knowledge. With the current studies conducted in England (primarily London and also Essex, a county adjacent to London), the majority of participants would have been reasonably familiar with the octolinear London Underground map, which could possibly have influenced their expectations concerning good design. However, interestingly, across studies, there was a significant age effect, with the age of people who selected the All-Curves map (mean 36.4 yrs, *SD* 16.1) slightly exceeding those who selected the RATP map (mean 30.3 yrs, *SD* 15.3), F $(1,150) = 5.52$, $p = .020$, which might go against a familiarity hypothesis. In general, map preference tasks seem to evoke strong individual differences, with all-curves maps, straight-line based maps, and topographically accurate maps splitting people three ways, but the sources of these differences are difficult to pin down. For example, across all three experiments, 56% of males selected the all-curves map in preference to the RATP map, and 54% of females.

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 network properties certainly lend themselves to the All-Curves approach, and although it is possible to find octolinear commercial Paris Metro maps which have fewer changes of direction than the RATP 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 octolinear 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, 2012). Overall though, the All-Curves design is an improvement over the current RATP map, and on average its aggregate questionnaire 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.

5.3. Objective measures versus subjective ratings

Turning to subjective measures (aggregate questionnaire ratings and map preferences), overall, these were not-at-all, very weakly or, at best, inconsistently related to objective measures of planning. This 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, and particularly applies to managers of transport undertakings commissioning designs from third parties. If they lack expertise in schematic map design, then objective, evidence-based criteria will be essential in order to ensure that a delivered design has adequate usability.

Why might objective and subjective measures be so weakly related? First, we should note that the people who took part in the study were unlikely to have any particular expertise in designing or evaluating maps. In the context of a short planning task, and in the absence of

explicit feedback they would be unlikely to be able to acquire such expertise either, and so would have difficulty in identifying the essentials of effective design in terms of the configuration of a map. Without such scaffolding, more salient but superficial indicators are likely to be used instead, such as familiarity of design principles, expectations of how a map should be designed, and whether it gives the impression of topographical accuracy (e.g., see Lindell & Mueller, 2011). More generally, superficial evaluation by novices is one of the recurring findings in the expertise literature (e.g., Chi, Feltovich, & Glaser, 1981). Also, it should be noted that although participants undertook planning tasks, the level of feedback available from their performance was relatively low. For Experiments 1 and 2, the only indication of a poorly functioning map would have been the planning of invalid routes, but these were uncommon, only reaching 10% for the Commercial map in Experiment 1. Hence, performance at these tasks is unlikely to yield sufficient feedback cues as to whether or not the map is functioning adequately (see also Roberts & Newton, 2003). Even for Experiment 3, where planning time difficulties were more salient because of the deadline procedure, of the eight people who exceeded at least one deadline, and/or failed to complete the final trial within the alloted time (owing to an insufficient time-bank), just 5/8 opted to switch to the alternative map in preference. One particular problem faced by participants is that it is difficult to make an evaluation in the absence of some sort of baseline. With participants planning using just one map, and subsequently choosing between designs, they had only attempted to use – and hence directly experienced – one of the options. Had both been experienced, and one had yielded more perceived difficulty than the other, this might well have re-calibrated their evaluations. To investigate this would therefore require repeated measures designs, possibly with a more stringent planning deadline than used in Experiment 3.

It is important to note that the aggregate questionnaire ratings and map preferences were not completely divorced from objective usability measures. All participants rejected the Commercial map in Experiment 1, 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. Hence, it is possible that only a particularly bad design

with obvious cues to its failings, for example one that is very difficult to read and for which many invalid routes are made, would be expected to yield aggregate ratings in line with its usability for non-experts. However, we should exercise some caution here. Although this map was objectively the worst in Experiment 1, the gap between this and the intermediate map (RATP) in terms of usability was smaller than the gap between the intermediate and the best map (All-Curves). The magnitudes of the usability differences between the three maps does not reflect the magnitudes of the rating differences between them. It is possible that participants rejected the Commercial map on the basis of aesthetic cues, rather than an awareness that it was difficult to use.

Even if the aggregate questionnaire 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. In all three experiments, subjective usability ratings were related to whether participants wished to *persist* with using a map. Hence, for both RATP and All-Curves maps in Experiments 1 and 2, those who rejected the map that they had used for planning, in favor of the alternative, had previously rated the map poorly via the questionnaire. Experiment 3 only replicated this finding for the All-Curves map (but with the largest effect size of all three experiments) and for the RATP map, the small effect was in the opposite direction to expected. This differential effect is difficult to explain, but suggests that the faster planning necessary for the deadline task and, for the RATP map, the greater intrusion of the impending deadline owing to the longer planning times, prevented a stable overall assessment of this design even from a subjective point of view. Hence, people's final choices reflected the prejudices brought to the experiment even more.

Overall, the engagement of a map is an important consideration when formulating information provision: If a design is rejected by passengers, 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 questionnaire ratings were almost identical across experiments), which suggests that to cater properly for map engagement, it would be necessary to provide for individual differences. Unlike alternatives to the QWERTY keyboard design, however, familiarity with traditional schematic maps does 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 evidence-based decisions derived from objective usability data. 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, 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. One important point to note is that because of the relative instability of the Paris Metro map design, Parisians are far less likely to be committed to a particular mapping approach than Londoners.

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. Computer science/graph drawing researchers tend to label this type of map as *octilinear* rather than *octolinear* because this contrasts with the other major type of graph, one without diagonal lines, which is known as *rectilinear* (e.g., as used for electrical circuit diagrams). However, Roberts (2009, 2012) shows that there are other important rule-sets with potential utility, and maps which are *hexalinear* and *dodecalinear* can be easily constructed. The terminology is more elegant if the Greek prefixes are not modified.

3. 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. The assumption underlying this work is that the optimization of the basic line trajectories is a fundamental aspect of usability that can be considered in isolation from these. 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, 2008b, 2012).

4. 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 disorganized network.*

5. At the time of writing, of the eight complex networks listed earlier, the only exceptions to octolinearity are the New York Subway and Mexico City Metro. Other notable exceptions in America include the Chicago Transit Authority network (the 'El'), 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 octolinearity; 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 octolinear schematics as they grow and mature, sometimes with irregular linearity as an intermediate step. The overwhelming majority of network maps are therefore either octolinear, or topographical.

APPENDIX: RESPONSES TO INDIVIDUAL QUESTIONNAIRE QUESTIONS

***** INSERT TABLE 8 HERE *****

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., & 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.

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.

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.

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.

Hong, S.-H., Merrick, D., & do Nascimento, H. A. D. (2006). Automatic visualization of metro maps. *Journal of Visual Languages and Computing*, *17*, 203-224.

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.

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

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

Mijksenaar, P., & Vroman, R. (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. (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. (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., & 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. (2008a). *Underground maps after Beck* (2nd ed). Harrow Weald, UK: Capital
      Transport Publishing.

Roberts, M. J., (2008b). Information pollution on the Underground Map.
      http://www.tubemapcentral.com/information_pollution/ip.html [accessed 11/05/2012]

Roberts, M. J. (2009). Henry Beck rules, not OK? Breaking the Rules of Diagrammatic Map
      design. Unpublished manuscript, available for download at
      http://www.tubemapcentral.com/Breaking_the_rules.pdf [accessed 11/05/2012]

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

Roberts, M. J., & Newton, E. J. (2003). Individual differences in the development of
      reasoning strategies. In D. Hardman & L. Macci (Eds.), *Thinking: Psychological
      perspectives on reasoning, judgment, and decision making* (pp. 23-43). Chichester:
      John Wiley.

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., & 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.

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.

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. 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.

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*, *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 aggregate questionnaire ratings, for

Experiments 1, 2, and 3.

		Experiment 1			Experiment 2		Experiment 3	
		RATP	All-Curves	Commer-cial	RATP	All-Curves	RATP	All-Curves
Planning time (seconds per journey)	Mean	66.3	52.4	63.9	50.0	25.2	28.9	19.8
	SD	28.5	28.9	26.3	26.8	12.9	13.5	4.7
Drawing time (seconds per journey)	Mean				30.6	21.6		
	SD				8.9	7.6		
Planning time (log transformed data)	Mean	4.02	3.78	4.01	3.63	3.00	3.22	2.93
	SD	0.43	0.55	0.40	0.58	0.54	0.43	0.21
Planning time (inv-log, geometric mean)	Mean	55.7	43.8	55.1	37.7	20.1	25.0	18.7
Drawing time (log transformed data)	Mean				3.33	2.95		
	SD				0.31	0.39		
Drawing time (inv-log, geometric mean)	Mean				27.9	19.1		
Estimated journey duration, invalid routes deleted (minutes)	Mean	57.4	59.9	60.1	56.8	58.0	65.9	64.1
	SD	3.4	4.2	5.1	4.2	7.5	4.1	2.4
Estimated journey duration, invalid routes substituted (minutes)	Mean	58.9	60.2	62.1	57.0	58.0	66.3	64.6
	SD	3.4	3.8	4.9	4.1	7.5	4.5	2.2
Invalid routes (percent)	Mean	6.5	2.0	10.0	2.9	1.1	3.8	1.3
	SD	13.1	6.1	17.5	7.2	4.6	8.1	5.0
Aggregate questionnaire rating (11 to 77, high scores better)	Mean	56.4	56.0	43.4	52.6	57.4	56.6	57.7
	SD	12.9	13.0	10.9	11.6	10.2	12.0	8.5

Table 2. Mean usability measures by map type and journey for Experiment 1.

			Hoches	Glacière	Porte d'Italie	Rambuteau	Ségur
			Rennes	Richelieu Drouot	Garibaldi	Convention	Rue St-Maur
Planning time (log transformed data)	RATP	Mean	4.34	4.07	4.16	3.66*	3.90
		SD	0.51	0.59	0.64	0.53	0.45
Planning time (inv-log, geometric mean)		Mean	76.3	58.4	64.2	38.7*	49.2
Planning time (log transformed data)	AC	Mean	4.07*	3.71*	3.57*	3.74	3.81*
		SD	0.58	0.61	0.67	0.60	0.69
Planning time (inv-log, geometric mean)		Mean	58.8*	40.9*	35.7*	42.1	45.2*
Planning time (log transformed data)	COM	Mean	4.25	3.90	4.06	3.90	3.94
		SD	0.50	0.53	0.55	0.53	0.51
Planning time (inv-log, geometric mean)		Mean	70.3	49.3	58.1	49.5	51.3
Estimated journey duration, invalid routes substituted (minutes)	RATP	Mean	72.0	51.3	66.8*	48.8*	55.9*
		SD	11.4	6.0	8.2	2.6	3.8
	AC	Mean	70.1*	49.6*	67.8	53.9	59.6
		SD	9.9	6.5	8.4	5.9	9.0
	COM	Mean	73.0	54.8	70.6	51.0	61.0
		SD	11.5	4.9	7.6	6.2	12.1
Invalid routes (percent)	RATP		22.5	5.0	5.0	0.0	0.0
	AC		7.5	2.5	0.0	0.0	0.0
	COM		20.0	12.5	10.0	0.0	7.5

Notes. RATP: Official Paris RATP Metro map; AC: All-Curves design; COM: Commercial map

Asterisks denote the best-performing map for each planning time and estimated journey duration.

Table 3. Usability measures as a function of Map Type and Map Choice for Experiment 1.

		Switch RATP (N = 24)	Switch All-Curves (N = 17)	Keep RATP (N = 16)	Keep All-Curves (N = 23)
Planning time	Mean	3.97	3.78	4.10	3.79
(log transformed data)	*SD*	0.41	0.54	0.46	0.56
Planning time	Mean	53.0	43.6	60.5	44.2
(inverse-log, geometric mean)					
Estimated journey duration,	Mean	59.2	59.7	58.6	60.5
invalid routes substituted (minutes)	*SD*	3.8	3.3	2.9	4.1
Invalid routes	Mean	6.7	1.2	6.3	2.6
(percent)	*SD*	15.2	4.9	9.6	6.9
Aggregate questionnaire rating	Mean	52.1	51.8	62.8	59.1
(11 to 77, high scores better)	*SD*	13.7	13.4	8.6	12.0

Table 4. Mean usability measures by map type and journey for Experiment 2.

			Hoches	Glacière	Porte d'Italie	Rambuteau	Ségur
			Rennes	Richelieu Drouot	Garibaldi	Convention	Rue St-Maur
Planning time (log transformed data)	RATP	Mean	3.80	3.39	3.78	3.53	3.63
		SD	0.80	1.01	0.66	0.86	0.67
Planning time (inv-log, geometric mean)		Mean	44.9	29.6	43.9	34.2	37.6
Planning time (log transformed data)	AC	Mean	3.11*	2.80*	2.90*	3.13*	3.04*
		SD	0.66	0.65	0.56	0.69	0.88
Planning time (inv-log, geometric mean)		Mean	22.4*	16.4*	18.2*	22.9*	21.0*
Estimated journey duration, invalid routes substituted (minutes)	RATP	Mean	72.0	49.0*	63.7*	48.6*	51.6*
		SD	10.7	5.1	10.6	9.7	3.0
	AC	Mean	64.6*	51.2	66.2	52.3	55.8
		SD	11.0	11.9	9.2	11.2	9.1
Invalid routes (percent)	RATP		0.0	9.5	4.8	0.0	0.0
	AC		0.0	5.3	0.0	0.0	0.0

Notes. RATP: Official RATP Paris Metro map; AC: All-Curves design.

Asterisks denote the best-performing map for each planning time and estimated journey duration.

Table 5. Usability measures as a function of Map Type and Map Choice for Experiment 2.

		Switch RATP (N = 7)	Switch All-Curves (N = 11)	Keep RATP (N = 14)	Keep All-Curves (N = 8)
Planning time	Mean	3.67	3.16	3.60	2.78
(log transformed data)	SD	0.48	0.54	0.64	0.50
Planning time	Mean	39.4	23.5	36.7	16.1
(inverse-log, geometric mean)					
Estimated journey duration,	Mean	58.6	59.2	56.2	56.4
invalid routes substituted (minutes)	SD	5.3	9.3	3.2	4.1
Invalid routes	Mean	2.9	1.8	2.9	0.0
(percent)	SD	7.6	6.0	7.3	0.0
Aggregate questionnaire rating	Mean	45.9	52.5	56.0	64.0
(11 to 77, high scores better)	SD	11.2	9.3	10.6	7.8

Table 6. Mean usability measures by map type and journey for Experiment 3.

			Avenue du President Kennedy	Hoches	Pelleport	Rambuteau	Richard Lenoir
			Danube	Rennes	Charles Michel	Convention	St Francois Xavier
Planning time (log transformed data)	RATP	Mean	3.45	3.22	3.36	3.01	3.08
		SD	0.61	0.44	0.52	0.45	0.45
Planning time (inv-log, geometric mean)		Mean	31.6	25.0	28.7	20.3	21.7
Planning time (log transformed data)	AC	Mean	3.19*	3.04*	2.85*	2.88*	2.68*
		SD	0.36	0.28	0.27	0.28	0.27
Planning time (inv-log, geometric mean)		Mean	24.2*	21.0*	17.2*	17.8*	14.6*
Estimated journey duration, invalid routes substituted (minutes)	RATP	Mean	70.6	72.9	82.5*	52.4	52.9
		SD	12.0	13.6	7.5	4.7	9.3
	AC	Mean	70.3*	67.6*	84.6	51.6*	48.8*
		SD	6.6	8.0	8.6	5.5	4.4
Invalid routes (percent)	RATP		0.0	12.5	0.0	6.3	0.0
	AC		0.0	0.0	6.3	0.0	0.0

Notes. RATP: Official RATP Paris Metro map; AC: All-Curves design.

Asterisks denote the best-performing map for each planning time and estimated journey duration.

Table 7. Usability measures as a function of Map Type and Map Choice for Experiment 3.

		Switch RATP (N = 10)	Switch All-Curves (N = 5)	Keep RATP (N = 6)	Keep All-Curves (N = 11)
Planning time	Mean	3.22	2.96	3.22	2.91
(log transformed data)	SD	0.48	0.19	0.37	0.23
Planning time	Mean	25.1	19.3	25.1	18.4
(inverse-log, geometric mean)					
Estimated journey duration,	Mean	65.1	64.5	68.1	64.6
invalid routes substituted (minutes)	SD	3.8	1.3	5.2	2.5
Invalid routes	Mean	2.0	0.0	6.7	1.8
(percent)	SD	6.3	0.0	10.3	6.0
Aggregate questionnaire rating	Mean	57.7	49.6	54.7	61.4
(11 to 77, high scores better)	SD	12.3	5.9	12.2	6.9

Table 8 (overleaf). Mean scores for individual questionnaire questions for all three experiments.

Note: for each question, the maximum score is 7, and the minimum is 1, 4 is the neutral point. For asterisked questions, scores have been reversed to that a high score is favourable (disagrees with statement) and a low score is unfavourable (agrees with the statement). For all remaining questions except the final one, a high score indicates agreement with the statement, and a low score indicates disagreement.

	Experiment 1			Experiment 2		Experiment 3	
	All-Curves	RATP	Comm-ercial	All-Curves	RATP	All-Curves	RATP
I found the journeys easy to plan	5.5 (1.2)	5.2 (1.3)	4.8 (1.4)	5.6 (1.1)	4.9 (1.1)	5.4 (1.1)	5.4 (1.0)
*The routes were difficult to discriminate (identify)	5.0 (1.7)	5.1 (1.4)	3.3 (1.5)	4.8 (1.5)	4.6 (1.5)	4.7 (1.0)	4.6 (1.6)
The station names were easy to identify	5.2 (1.6)	5.4 (1.6)	4.5 (1.6)	5.0 (1.8)	5.1 (1.6)	5.6 (1.1)	5.3 (1.4)
*Station interchanges were difficult to negotiate	4.6 (1.6)	4.6 (1.7)	3.2 (1.3)	4.9 (1.7)	4.2 (1.6)	4.5 (1.6)	4.5 (1.4)
Line trajectories were easy to follow using this map	5.2 (1.4)	5.4 (1.5)	4.4 (1.7)	5.3 (1.7)	5.0 (1.3)	5.5 (0.9)	5.6 (1.4)
*I found this map disorientating to use	5.4 (1.6)	5.4 (1.6)	3.7 (1.8)	5.4 (1.6)	4.8 (1.8)	5.2 (1.4)	5.4 (1.4)
I would be happy to use this map to plan real-life journeys around Paris	5.6 (1.7)	5.4 (1.6)	4.3 (1.6)	5.5 (1.5)	5.3 (1.7)	5.7 (1.1)	5.4 (1.1)
*With this map design, I would rather walk or take a Taxi than use the Metro	5.4 (1.6)	5.5 (1.6)	3.1 (1.8)	5.9 (1.1)	5.1 (1.7)	5.7 (1.3)	5.6 (1.3)
*I found the map visually 'disturbing'	5.2 (1.8)	5.3 (1.7)	3.7 (1.9)	5.6 (1.7)	5.1 (2.1)	6.0 (1.3)	5.5 (1.8)
*I found the map cluttered	4.1 (1.9)	4.2 (1.8)	2.6 (1.7)	4.1 (1.8)	3.7 (2.0)		
I found the map clean and uncluttered						4.4 (1.4)	4.7 (1.5)
*I would look for another design of Paris Metro map to use at the earliest opportunity	4.8 (1.8)	4.7 (1.9)	2.8 (1.5)	5.2 (1.4)	4.8 (1.9)	5.0 (1.2)	4.8 (1.4)
I preferred to look for a direct route no matter how many interchanges were required	4.3 (1.9)	4.2 (1.9)	4.3 (1.6)	4.2 (1.5)	5.3 (1.3)	3.9 (1.6)	4.3 (1.6)
Some parts of the map looked confusing, and I planned journeys that avoided them	3.5 (1.9)	3.1 (1.8)	3.4 (1.7)	3.2 (1.7)	2.5 (1.4)	3.6 (1.7)	3.2 (1.7)
The best routes for me had the fewest station stops along the way						4.8 (1.4)	3.8 (0.9)
This map is intended for planning journeys, but I think that it is also probably geographically accurate	4.5 (1.6)	3.6 (1.7)	4.4 (1.6)	4.2 (1.5)	4.1 (1.5)	4.8 (1.4)	3.8 (0.9)
Frequency of urban rail travel: (7) every day; (6) a few days every week; (5) a few times every month; (4) about once a month; (3) a few times a year; (2) once a year or less; and (1) never/not for years						1: N = 9 2: N = 3 3: N = 1 4: N = 1 5: N = 1 6: N = 1	N = 8 N = 1 N = 4 N = 0 N = 2 N = 1

Figure 1. Examples of matrix items with complex, difficult to identify shapes/patterns (upper) and simpler, easier to identify shapes/patterns (lower). The underlying logic for these is the same, each combining one *constant in a row* rule – the type of line – with two *distribution of three values* rules – the numbers of dots, and the pattern of line linkage (Carpenter, Just, & Shell, 1990). In the upper item, the superficial complexities of the lines obscure the underlying patterns. Image © Maxwell J. Roberts, 2011. Reproduced with permission, all rights retained.
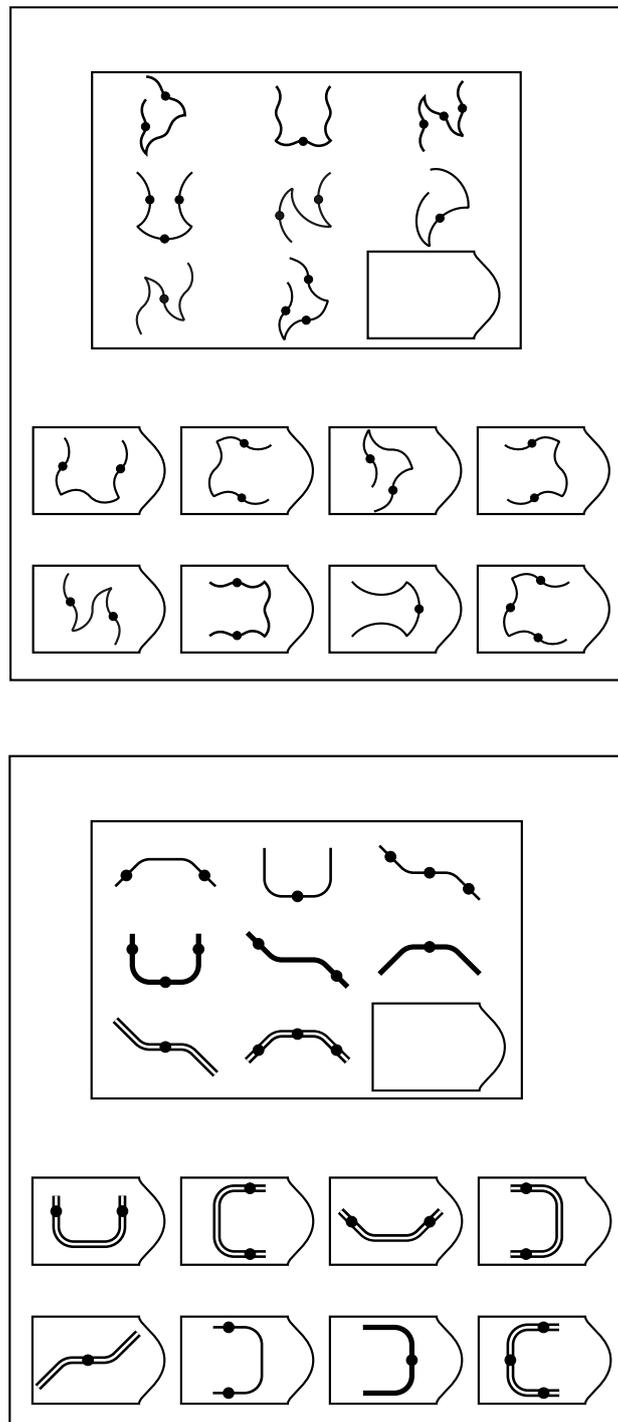
Figure 2. Monochrome image of the All-Curves map used in this study. © Maxwell J. Roberts, 2007.

Reproduced with permission, all rights retained. Copies of all stimuli can be obtained from the first

author. The official RATP Paris Metro map can also be downloaded from www.ratp.fr and a section of

the Commercial map is given at www.streetwisemaps.com/metro-map/paris-metro-map.html

[accessed 11/05/2012]