Metro Map Colour-Coding: Effect on Usability in Route Tracing

Peter B. Lloyd[1][0000-0001-7393-1322], Peter Rodgers[1][0000-0002-4100-3596],
and Maxwell J. Roberts[2]

[1] School of Computing, University of Kent, Darwin Road, Canterbury CT2 7NZ, UK
{pbl,pjr}@kent.ac.uk
[2] Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK
mjr@essex.ac.uk

Abstract. Does the choice of colour-coding scheme affect the usability of metro maps, as measured by the accuracy and speed of navigation? Using colour to differentiate lines or services in maps of metro rail networks has been a common practice around the world for many decades. Broadly speaking, there are two basic schemes: 'route colouring', in which each end-to-end route has a distinct colour, and 'trunk colouring', in which each major trunk has a distinct colour, and the individual routes inherit the colour of the main trunk that they run along. A third, intermediate scheme is 'shaded colouring', in which each trunk has a distinct colour, and each route has a distinct shade of that colour. In this study, 285 volunteers in the US were randomised to these three colour-coding schemes and performed seventeen navigational tasks. Each task involved tracing a route in the New York City subway map. Overall, we found that route colouring was significantly more accurate than the trunk- and shaded-colouring schemes. A planned subset analysis, however, revealed major differences between specific navigational hazards: route colouring performed better only against certain navigational hazards; trunk colouring performed best against one hazard; and other hazards showed no effect of colour coding. Route colouring was significantly faster only in one subset.

Keywords: Metro Maps, Colour Coding, Navigational Hazard, New York City Subway, Vignelli, Usability Testing.

1   Introduction

We report on the first part of a project to study the effect of colour coding on the usability of metro maps. A map designer's choice of which colour-coding scheme to use—in particular, colour-coding by individual route versus trunk—is a long-standing point of controversy, but it has not previously been subject to systematic empirical study.

The subway system of New York City has in the past switched between route-colouring and trunk-colouring, and currently has an official online map that shows individual routes but is coloured by trunk. It is therefore an ideal vehicle for investigating the effect of colour-coding scheme in a realistic system. The New York City subway

has one of the most complex service patterns in the world, and is currently undergoing re-evaluation and development of its information delivery, so it is especially suited to the present study. Although this study specifically addresses a metro map of New York City, we believe the results have wider application to metro maps in other cities.

1.1 Basic concepts and nomenclature

Metro maps exhibit a great variety, as shown by Ovenden (2015). Nomenclature is also varied, and so we will explicitly define the terms that we will be using, as follows.

Metro maps are diagrams of urban rail networks, in which the nodes represent stations, edges represent services, and the layout, colouring, and graphical symbolism are designed to assist the passenger in understanding and navigating the network, in order to use it to travel around the city. Thus the map must show both (a) the topological connectivity between the stations as effected by the physical tracks and the passenger services that run along those tracks, and (b) the approximate relationship of the network to the geographic layout of the city.

A **line** comprises one, two, three, or four tracks running together: the tracks that make up a line are usually laid side-by-side, but sometimes are vertically one above the other. In a **two-track line**, the tracks go in opposite directions, 'up' and 'down'; in a **four-track** line, two tracks go up (one express and one local) and two go down (again, express and local); in a **three-track** line, the third track runs express and alternates between up and down in the morning and evening rush hours; in a **single-track** shuttle, the train alternates direction. A **local** train stops at all stations; an **express** train stops at some and skips others. A **route** is a named individual train service, which normally has a single terminal at each end. A **trunk** is a bundle of routes carried on the same line. A **path** is a series of one or more segments of routes, connected by **transfers**. A **journey** is a project by a rider to travel between origin and destination stations. A **run** is a single-route segment of a path between transfers. More informative metro maps show the routes as well as the lines, and a fully informative metro map will show the variation of route service patterns over time of day and day of week.

Route tracing is the mental activity of following a single route as it wends its way through a metro map, whereas **journey planning** is the mental activity of navigating a journey based on the map; it comprises 'path construction', that is, assembling a path from one or more route segments, and 'path selection', that is, choosing the best one of several alternative paths. A **navigational hazard** is an identifiable local feature of map that increases the likelihood of a user's making a mistake in these activities.

The map to be examined in this study is an official diagram of the subway network of New York City, used on the MTA (Metropolitan Transportation Authority) **Weekender** web site (MTA, 2017). The map was designed by Yoshiki Waterhouse (formerly of Vignelli Associates, now at Waterhouse Cifuentes Design) based on an original 1972 design by Massimo Vignelli, and maintained under Chuck Gordanier, MTA. This diagram was chosen because it is a large and complex map whose structure is suited to experimental changes of colour scheme. Although primarily an online map, the same diagram is increasingly being used in print. This study compares three differently coloured experimental variants of this Weekender map.

Figs. 1 to 3 show excerpts from the three variant maps used in this study. A **route-coloured** map (Fig. 1) shows each route in a distinctive colour, although this may involve re-using colours, as there may be more routes in total than easily distinguishable colours. **Trunk-coloured** maps (Fig. 2) show each trunk in a distinctive colour: all the routes that run along the same trunk share the same colour. Finally, a **shaded-colour** map (Fig. 3) shows each trunk in a distinctive colour, and each route in a different shade of that colour. Note that a line with no trunk-and-branch structure might still have route colours distinct from trunk colours because the line might carry routes with different express and local stopping patterns. **Hybrid** maps have a mix of trunk and route colouring: this is seen in complex maps, e.g. Cologne (KVB, 2017). A hybrid map may be referred to as trunk- or route-coloured if one style predominates.

A map can be **route-drawn** (routes drawn as separate lines) or **trunk-drawn** (the trunk is drawn as one line, with branches). A route-drawn map can be route-coloured, trunk-coloured, or shade-coloured; but a trunk-drawn map must be trunk-coloured. The London Underground map is an example of a predominantly trunk-drawn diagram, while the Weekender is a route-drawn, trunk-coloured map.

Fig. 1. Route colours **Fig. 2.** Trunk colours **Fig. 3.** Shaded colours

Use of colour coding. Colour coding of metro services started in 1907, when underground railways had already been in use for half a century (Roberts, 2012). Different cities adopted the concept differently (Lloyd 2012; Ovenden et al., 2008). For example, in London, Paris, and New York, lines were first coloured by operating company, with trunk colouring used within each line. When those lines were subsumed under municipal control (in 1933, 1930, and 1940 respectively) the maps initially retained that historic colour-coding method. In London, that colour scheme mostly continues today; in Paris, it was replaced by route colouring in 1934; in New York, it was replaced by a

route-drawn, route-coloured map in 1967; and by a trunk-drawn, trunk-coloured map in 1979 (Lloyd, 2012). In 2011, the MTA introduced the Weekender, a route-drawn map that preserved the 1979 trunk colouring.

1.2 Usability

Historical usability studies Metro maps have traditionally been created in an artisanal manner without recourse to psychological theory or empirical studies of usability. In recent years, though, a notion of the evidence-based design of metro maps has emerged, combining psychological models with empirical studies of the relationships between observable characteristics of the map and objective measures of the map's usability.

An early study of the usability of a metro map was carried out by the marketing firm Barrington (1966), commissioned by the New York City Transit Authority. Their survey comprised riders' opinions about a prototype design of the map, plus a small set of navigation tests. In the following decade, publicity around the Vignelli map prompted new studies in 1973 by Bronzaft et al. (1976), who selected test journeys with the intention of showing that Vignelli's new schematic map had navigational hazards. She did not intend the study to establish any principles of map design, or to produce an objective evaluation of the Vignelli map in relation to some alternative design.

Although these studies did not provide specific insights into the hazards or design of metro maps, we may at least glean a corroboration of anecdotal reports that enough users find it hard to use metro maps (or, at least, the New York City subway map) to warrant further investigation to characterize and quantify the problem.

Bronzaft & Schachter (1978) and Garland et al. (1979) found that colour-coded maps were more usable, and more strongly preferred, than non-colour-coded maps, but they did not address what effect the choice of particular colour-coding schemes might have.

Modern usability studies. The systematic investigation of metro maps following scientific principles was instigated comparatively recently. Guo (2011) found that the London Underground map layout led to sub-optimal journeys. Roberts (2014a) and Roberts & Vaeng (2016) report on the usability effects of large-scale features of the map layout in the London Underground, the London Docklands Light Railway (DLR), the Paris Metro, and the Berlin U-Bahn. The work described in this paper extends those studies by addressing colour coding and local navigational hazards.

Bronzaft in the 1970s used a mix of *in vivo* travel and pencil-and-paper navigation, and modern studies have mostly used pen-and-paper tasks (Roberts et al., 2016) with some use of touch-sensitive screens (Roberts & Rose, 2016). Guo et al. (2017) extended the methodology by using the automated presentation of map tests remotely over the internet via Mechanical Turk, a technique that we have adopted in the present study.

The main results yielded by these contemporary usability studies (summarised by Roberts, 2014a) are: that objective and subjective evaluations of usability are uncorrelated; that layout affects journey planning; that Beck's octilinear layout is not a 'gold standard' for map layout in all cities, and that instead a framework of more general principles for map layout can be identified, involving the simplicity, coherence, and harmony of line trajectories, and the balance and topographicity of the map layout. This

is in stark contrast to the formerly prevailing Beck-centric conventional wisdom (e.g. Field & Cartwright, 2014). Which aspects of a map have the most impact on objective usability is an open question for empirical research. Factors likely to influence usability are: overall layout; colour coding; symbolism and layout of transfer stations; symbolism of non-transfer stations; junction layout; and positioning and typography of labels.

Theory. Until recently, cognitive theory focused on topographic rather than diagrammatic maps (Robinson, 1952; MacEachren, 1995; Montello, 2002). Stemming from the original work of Miller (1956), the notion of cognitive load is a key to understanding metro maps. For example, Gallotti et al. (2016) have applied cognitive theory to quantify the information overload in map navigation in megacities. In the present project, our model is: (a) tracing a route is essentially a perceptual task, while (b) journey planning is cognitive and involves constructing and comparing alternative paths. Roberts (2014b) outlines two features of map reading: *cognitive load* and *attention capture*. Individuals can handle limited information ('cognitive load'). If the input exceeds that limit, accuracy and speed decline abruptly. Conversely, the salient elements of a task must capture and retain attention. To reduce cognitive load, and heighten attention on salient features, the map should minimise irrelevant noise and be organised to allow the user to find and grasp relevant information. The use of separate route colours is expected to enhance the perceptual task of route tracing but degrade journey planning because their added cognitive load would outweigh the perceptual advantage of distinct colours. Here we test just the first, perceptual part of this theory, that is, route tracing.

Prediction: The perceptual task of visually tracking something is easier if that thing can be picked out with distinctive attributes, in this case distinctive colour; and the countervailing disadvantage, namely that route-colouring adds visual clutter to the map as a whole, does not come into play because the subject is not performing a cognitive task. When we isolate route tracing from the more general problem of journey planning, we expect it to be more accurate and faster with route-colouring than with trunk-colouring map; and shaded-colouring to be intermediate. Note: We consider only route-drawn maps and do not address the effect of route-drawn v trunk-drawn layouts.

Motivation. The study's impetus comes from a 2014 workshop of wayfinding professionals, which suggested that route-coloured maps would have higher usability. Its relevance comes from metros that switch between colour schemes (e.g. Frankfurt, New York City), plus those that use shaded-colour schemes (Kick, 2017; KVB, 2017). So there is theoretic and practical interest in the question: what effect does colour scheme have on metro map usability—especially its accuracy, as this has more real-life impact?

2 Method

Map sourcing. The MTA provided a vector image file of the subway map of New York City, current at the start of this study (September 2015), and we modified the colours in Adobe Illustrator for this project. We wanted our results to apply to both screen and

print media, but the latter has a narrower range of colour tones than screens, so we adjusted the trunk colours of the Weekender map (see below). Our three schemes were:

- **Route-coloured map** used RGB colours approximating the Pantone colours of the 1972 Vignelli map. Some routes did not exist in 1972 and have no 'correct' Unimark colours and so were assigned subjectively reasonable colours. (Fig. 1.)
- **Trunk-coloured map** closely matched the colours of the Weekender, toned down to the printable range, and adjusted to maintain visibility of station markers. (Fig. 2.)
- **Shade-coloured map** was inspired by Jabbour's KickMap, but the colours were toned down to the printable range, and adjusted to maintain tonal separation. In each trunk, light and dark shades were alternated for greater contrast. (Fig. 3.)

We omitted route labels from the map to prevent the route-tracing task becoming trivial. And we omitted background details that were not pertinent to the navigation task (wheelchair access and bus routes) to reduce possible sources of extraneous variance.

We defined test journeys #1 to #17 by selecting seventeen pairs of stations, so that each journey would present one main navigational hazard (Table 1) characterised thus:

- **Route slips:** *flipping* ('*Slip F*') where two or more routes switch places (Fig. 4); *joining* ('*Slip J*') where two or more routes converge into one trunk on a shared segment of line (three instances are shown in the orange routes in Fig. 5), *underpassing* ('*Slip U*'),where one trunk passes underneath another (Fig. 6), the *twin hazard* ('*Slip T*') of flipping on an underpass (Fig. 7), and *parallel running* ('*Slip P*') where two or more routes have a long run together (Fig. 8).
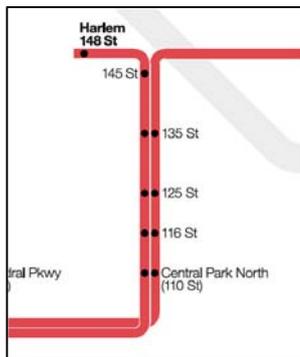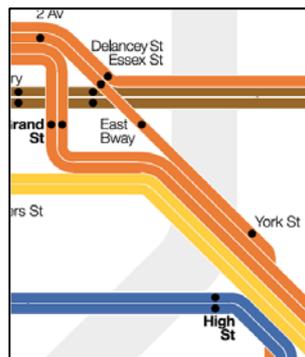

Fig. 4. Slip F, #12

Fig. 5. Slip J, #8/#9
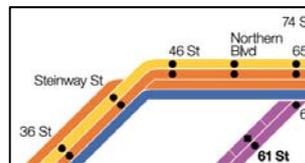

Fig. 7. Slip T, #14/#15

Fig. 6. Slip U, #10/#11

Fig. 8. Slip P, #7

- **Branch jumps:** these comprise *route jumps* ('*Jump R*') where the rider 'jumps' across a branched route, in this case Beach 90 St to Beach 44 St (Fig. 9) and *trunk jumps* ('*Jump T*') where the rider jumps across a branched trunk, 138 St-Grand Concourse to 3 Av-138 St in this case (Fig. 10).
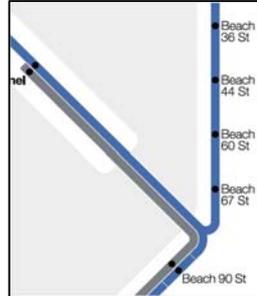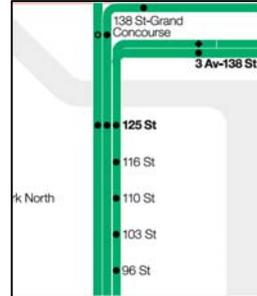

Fig. 9. Jump R, #2   **Fig. 10.** Jump T, #4

- **Mistransfers:** The only mistransfer considered was the *phantom transfer* ('*Trans P*'), where the user believes a transfer is needed when it is not, in this case when travelling from 168 St to 157 St (Fig. 11).
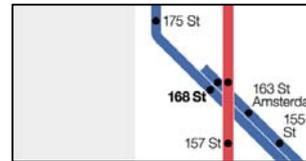
Fig. 11. Trans P, #17

Four attention exercises were added, which were trivial journeys to check that the user was paying attention and not just clicking at random, #18 to #21.

Table 1. Routes.

#	From	To	Hazard
1	104 St (A)	Aqueduct | North Conduit Ave (A)	Jump R
2	Beach 90 St (A)	Beach | 44 St (A)	"
3	Nereid Av (5)	Eastchester | Dyre Av (5)	"
4	138 St-Grand | Concourse (4,5)	3 Av | 138 St (6)	Jump T
5	72 St (B,C)	5 Av-53 St (M,E)	"
6	175 St (A)	155 St (B,D)	"
7	Prince St (R,N)	Queensboro | Plaza (N,Q,7)	Slip P
8	Coney Island | Stillwell Av (D,N,F,Q)	Wood|haven Blvd (R,M)	Slip J
9	Coney Island | Stillwell Av (D,N,F,Q)	Forest Hills|-71 Av (R.F.M,E)	"
10	Spring St (C,E)	Lafayette|Av (C)	Slip U
11	Spring St (C,E)	Nostrand Av (A,C)	"
12	New Lots Av (3)	Harlem | 148 St (3)	Slip F
13	New Lots Av (3)	149 St-Grand Concourse (2,4,5)	"
14	Jamaica-179 St (F)	Court Sq | 23 St (M,E)	Slip T
15	Jamaica-179 St (F)	Roosevelt | Island (F)	"
16	City Hall (R)	Queensboro | Plaza (N,Q,7)	"
17	168 St (A,C,1)	157 St (1)	Trans P
18	Euclid AV (A,C)	Grant Av (A)	Attention
19	Kings Hwy (N)	Coney Island | Stillwell Av (D,N,F,Q)	"
20	Fulton | St (G)	Clark St (3)	"
21	Kingsbridge | Rd (4)	Woodlawn (4)	"

Recruitment. Subjects were recruited from across the USA through the Amazon *Mechanical Turk* crowdsource service (Amazon, 2017) and paid $7.50 for up to an hour's work. Such crowdsourcing tools have been established as a platform for conducting behavioural science experiments (Crump et al., 2013; Paolacci et al., 2010). Whilst laboratories provide a controlled space for experimenters, they are typically restricted in participant numbers and demographics (Gadiraju et al., 2017). Crowdsourcing provides access to a large number of diverse participants (Mason & Suri, 2012). Issues can arise: first, a lack of control of participants' attention to the task, who might be clicking randomly to complete the task as quickly as possible; second, the lack of a means of interacting with the participant for training. We mitigated these, in the first case through attention-checking questions that could be performed without much cognitive effort, and in the second case through careful design of the study software to include a sizeable automated training phase (which was developed in the supervised pre-pilot study).

Study software. We presented the maps as Scalable Vector Graphic (SVG) images in JavaScript within Mechanical Turk. As each subject began, the software randomly allocated a colour scheme (trunk, route, or shaded). It then presented four predefined training trials, which had the same appearance as the test trials, but the subject was told whether his/her answer was correct. For each test trial, the software randomly selected one of the twenty-one origin-destination pairs, and displayed them in a zoomed rectangle of the map, with blue and red outlines around the start and end station names. The words "Start" and "Finish" were displayed (in blue and red respectively) in a large font (40 pixels) to draw attention to these stations, these two words fading over ten seconds.

Above the map was the question, "Can you travel from station $<S_1>$ to station $<S_2>$ without changing trains?" (e.g. Fig. 12), followed by radio buttons labelled "Yes" and "No". When the user clicked on one them, a third button appeared, labelled "Finish", which the user clicked to close that trial and go to the next. There were three clickable symbols immediately above this, "+" and "-" for zoom in and out, and "Reset", and users could pan the map by clicking to grab it—all reproducing the interface on the Weekender site. In a laboratory pre-pilot study, we found that users had difficulty using these controls, so zooming by thumbwheel was added. An online pilot study using Mechanical Turk showed no further changes were needed, so we moved to the full study.
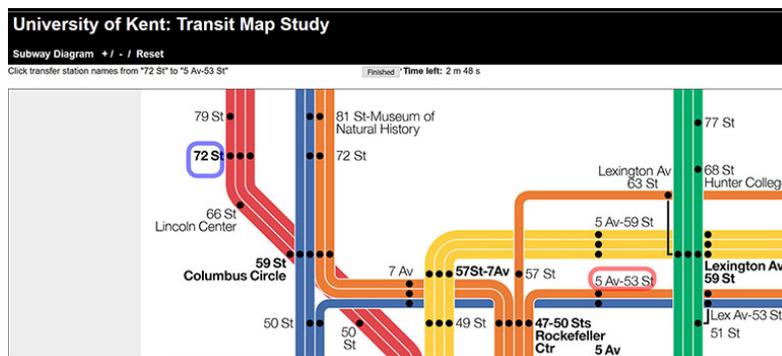
Software and data: `https://www.cs.kent.ac.uk/projects/metromap/`.

Fig. 12. Sample screenshot (for test journey from 72 St to 5 Av–53 St)

Each subject was allowed three minutes per trial: any timeouts were logged as a 'failure by timeout' (and counted as a non-success in the statistics), and the subject moved to the next trial. A total of 10 trials were timed out (4 trunk, 4 route, 2 shaded).

Exclusions. Neither the training trials nor the attention exercises were included in the statistics presented here. We excluded all data from any subject who failed in two of the four attention exercises, or who failed to complete the full set of tasks, or who completed the tasks but failed to upload the data to Mechanical Turk. Of 305 subjects who were randomised, 20 were excluded for those reasons.

3   Results

Navigation accuracy when following a route in the New York City diagrammatic subway map is significantly affected by the colour-coding scheme. Table 2 (and Fig. 13) shows the analysis of 4,845 trials by 285 subjects (98 trunk, 93 route, 94 shade colouring). Each subject's accuracy is scored as the percentage of correct answers in 17 trials.

Table 2. Effect of colour coding on accuracy & speed of navigation.

| Mean score (%), S $\pm SE$				Mean time (seconds), T $\pm SE$			
Trunk	Route	Shaded		Trunk	Route	Shaded	
66.8 ± 1.4	71.9 ± 1.2	65.7 ± 1.3	$p < 0.01$, $F(2,282) = 5.7$	26.8 ± 1.0	24.5 ± 1.1	27.3 ± 1.4	NS, $F(2,282) = 1.5$
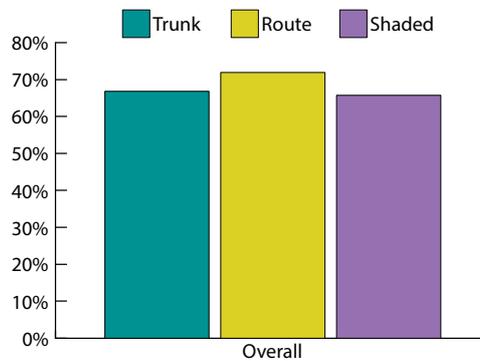
Fig. 13. Mean score

Route colouring was the most accurate, while trunk colouring was only slightly more accurate than shaded. A one-factor, three-level analysis of variance showed a statistically significant effect of colour scheme on accuracy ($p < 0.01$), although only one pairwise comparison (route v shaded) was significant ($p < 0.05$). Route colouring was also the fastest, but the effect of colour coding on time was not statistically significant.

The main navigational hazard presented in each trial affected the accuracy of navigation. We examined two classes of hazard: slippage between routes, and jumping

across branches (Table 3, Fig. 14). In slippage, route colouring's advantage was more pronounced, but in branch jumping, the effect was inverted and trunk performed best.

Table 3. Subset analysis by general class of navigational hazard.

	Mean score (%), S $\pm SE$				Mean time (seconds), T $\pm SE$			
	Trunk	Route	Shaded		Trunk	Route	Shaded	
Slip	68.4±*1.5*	81.1±*1.5*	71.3±*1.7*	$p < 0.01$, F(2,282) = 15.8	31.5±*1.2*	29.4±*1.3*	31.8±*1.6*	NS, F(2,282) = 0.8
Jump	69.0±*2.3*	60.0±*1.9*	61.9±*2.3*	$p < 0.01$, F(2,282) = 4.7	21.0±*1.0*	18.0±*1.0*	21.8±*1.4*	NS, F(2,282) = 3.0
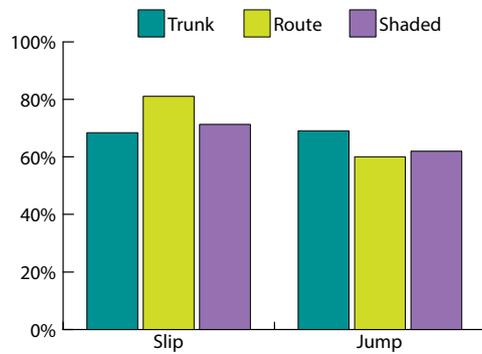
Fig. 14. Mean score, by class of navigational hazard

In both subsets Slip and Branch, a three-level analysis of variance showed a significant effect of colour coding (both $p < 0.01$). In the Slip subset, pair-wise analysis of variance showed that route colouring is significantly more accurate than either trunk or shaded (both $p < 0.01$) but trunk and shaded are not significantly different from each other. In the Jump subset, pair-wise analysis of variance shows only that trunk colouring is significantly more accurate than route ($p < 0.05$).

Analysis of eight specific types of navigational hazard indicated a strong dependence of the colour-coding effect on the main navigational hazard that is presented in a trial: three out of five slippage hazards showed large and significant advantages in route colouring, while the other two showed no significant dependence on colour coding (Table 4, Fig. 15). The two branch-jumping hazards showed a significant effect, but with opposite directions–one favouring trunk colouring, the other route. Phantom transfers showed worse-than-chance error rates, with a significant advantage in route colouring.

In most subsets, route colouring yielded the fastest navigation, but only in one ('Jump T') did this attain statistical significance.

Where the main navigational hazard is route slippage due to flipping (Slip F), joining (Slip J) and the twin hazard of flipping on an underpass (Slip T), route colouring shows a strong and statistically significant advantage in accuracy of navigation. Trunk colouring is worst, and shaded colouring has intermediate accuracy. In each of these three subsets, three-level analysis of variance shows a significant effect of colour scheme (p

< 0.01); and pair-wise analyses show that route colouring is significantly more accurate than trunk ($p < 0.01$) or shaded ($p < 0.05$) but trunk and shaded are not significantly different from each other.

Table 4. Subset analysis by specific type of navigational hazard

	Mean score (%), S $\pm SE$				Mean time (seconds), T $\pm SE$			
	Trunk	Route	Shaded		Trunk	Route	Shaded	
Slip F	70.9±3.1	89.2±2.4	75.0±3.3	$p < 0.01$, F(2,282) = 9.9	38.1±2.1	38.9±2.4	39.7±2.9	NS, F(2,282) = 0.1
Slip J	52.6±2.3	68.8±2.6	56.4±2.7	$p < 0.01$, F(2,282) = 10.1	53.6±3.1	43.6±3.0	52.3±3.7	NS, F(2,282) = 2.6
Slip T	65.3±3.0	85.7±2.1	71.3±2.2	$p < 0.01$, F(2,282) = 15.8	22.5±1.1	20.9±1.2	22.4±1.3	NS, F(2,282) = 0.5
Slip P	79.6±4.1	75.3±4.5	80.9±4.1	NS, F(2,282) = 0.5	23.1±1.5	26.0±2.2	26.4±2.1	NS, F(2,282) = 0.8
Slip U	80.6±3.5	81.2±3.5	77.7±3.7	NS, F(2,282) = 0.27	20.3±1.0	19.8±1.1	20.2±1.3	NS, F(2,282) = 0.0
Jump R	46.3±3.7	23.7±3.6	35.8±3.5	$p < 0.01$, F(2,282) = 9.25	21.4±1.2	18.6±1.1	21.6±1.6	NS, F(2,282) = 1.6
Jump T	91.8±1.9	96.4±1.5	87.9±2.2	$p < 0.01$, F(2,282) = 4.8	20.6±1.1	17.4±1.1	22.1±1.5	$p < 0.05$, F(2,282) = 3.5
Trans P	37.8±4.9	51.6±5.2	33.0±4.8	$p < 0.05$, F(2,282) = 3.59	14.2±1.0	15.4±1.1	15.8±1.5	NS, F(2,282) = 0.5
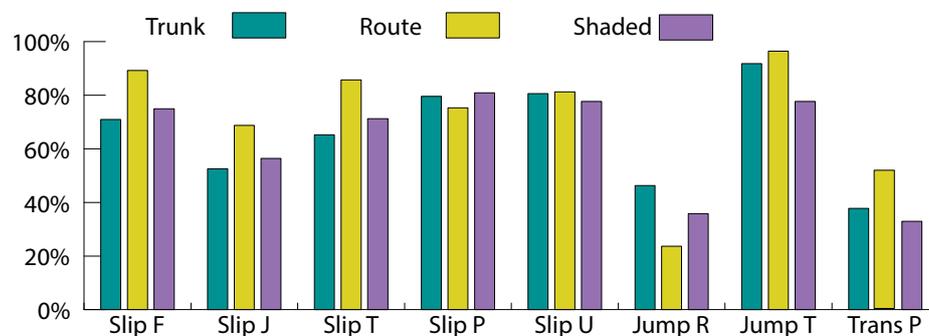
Fig. 15. Mean score, by individual type of navigational hazard

In the trials where the main navigational hazard is route slippage due to parallel running (Slip P) or underpassing (Slip U), neither the accuracy nor the speed of navigation exhibited any significant dependence on the colour coding.

Branch jumping exhibits quite different behaviour from route slippage. In the trials where the main navigational hazard is jumping across a route branch (Jump R), both the accuracy and speed of navigation are strongly dependent on colour scheme, but—unlike in the slippage hazard—the trunk colour scheme is most accurate and route colouring worst. Note that these results are all worse than chance: in effect, the shaded-

and route-coloured maps are positively misleading, while the trunk-coloured map is almost as good as tossing a coin. In a three-level analysis of variance, colour coding has a significant effect on accuracy ($p < 0.01$), and pair-wise analysis of variance shows that trunk is significantly better than route colouring ($p < 0.01$). The other pair-wise comparisons are not statistically significant.

In the trials where the main navigational hazard is jumping across a trunk branch (Jump T), the error rate is low, but the pattern is the same as in the slippage hazards: route colouring is the most accurate. Three-level analysis of variance shows that colour coding has a significant effect on accuracy ($p < 0.01$); pair-wise analysis of variance shows that route colouring is significantly better than shaded ($p < 0.05$). The other pair-wise comparisons are not statistically significant.

In both of these Jump subsets, route colouring is the fastest. Three-level analysis of variance shows a significant effect of colour coding only for Jump T. Pair-wise analyses of variance are not statistically significant.

In the trials where the only hazard is a phantom transfer, we have a situation where, like Jump R above, the map is positively misleading with success rates substantially below chance, and there is a significant dependence on colour-coding scheme. In contrast with Jump R, however, in Trans P, route colouring is most accurate. Three-level analysis of variance showed a significant effect of colour coding ($p < 0.05$); but the pair-wise analyses of variance are not significant. Speed of navigation shows no clear pattern in this subset, and analysis of variance yields nothing statistically significant.

4   Discussion

4.1   Navigational hazards

Route slippage. The results confirm the central plank of our theoretical platform: the navigational task of tracing a line as it weaves through the subway map is more accurately performed when done with a map that employs route colouring. It was expected that shaded colouring would yield an intermediate accuracy between those of route colouring and trunk colouring. In fact, in these tasks there is no real difference in accuracy between the trunk and shaded colouring.

Route colouring's advantage is mainly manifested in two recognisable navigational hazards: flipping, where two routes swap their positions within a trunk (referred to as 'Slip F'); and joining, where routes join a trunk ('Slip J'). Two other hazards that were expected to be comparable were also analysed: underpassing, where one trunk passes under another trunk ('Slip U'); and parallel running, where a trunk runs for a long distance with multiple bends but without any of the other hazards ('Slip P'). In all four hazards, we expected the perceptual task of tracing one route in a trunk of identically coloured routes would be less accurate than doing so when each trunk is distinctly coloured. In fact, the hazards Slip U and Slip P showed no differential effect of colour coding. A fifth subset was analysed, whose principal risk was the twin hazard of flipping immediately next to an underpass: as expected, this yielded a higher error rate than

Slip F or Slip U alone, and exhibited a significant effect of the colour coding, but this appears to be due entirely to the flipping, not the underpassing.

It appears that when the parallel routes closely follow each other (as in Slip P and Slip U), it is easier for the eye to follow them, and route colouring offers no advantage; but when the routes behave differently (as in Slip F and J), they are harder to follow as they split the user's attention, and then route colouring offers an advantage.

Branch jumping. We expected that colour coding might affect the reading of branches, and that subjects might mistakenly believe they can 'jump' across junctions without changing trains. We therefore included six trials to investigate 'branch jumping': three to test for jumping across trunk junctions, where a route splits off from a trunk ('Jump T'), and three for jumping across route junctions, where a route splits into two limbs ('Jump R'). A clear result was obtained: Jump R had a high error rate, and route colours performed worst; Jump T had a low error rate, and route colours performed best. It is curious that Jump R is affected by the colour scheme, despite the fact that the junction itself looks the same whichever scheme is used. Only the context provided by the rest of the map changes. It is hypothesised that in a route-coloured map, the user is confused by the inconsistent use of the colour coding in the branch, with two undifferentiated routes given the same colour, and the confusion disrupts the thinking and causes errors.

Mistransfers. Besides the test trials, four trivial trials were included to provide a check on the subject's attention. In the laboratory pre-pilot study, however, one of these tasks, the phantom transfer ('Trans P'), had surprisingly high error rates and showed a strong effect of colour coding. This was therefore added as the seventeenth test trial in the online pilot study and the main study; and a replacement attention trial added. The cause of the errors seems to be a misreading of the Vignelli transfer symbolism at 168 St station, at the junction of the 7 Ave Broadway Line (formerly IRT) and the Washington Heights line (BMT), where the three station dots are not aligned. As this navigational task does not involve tracking routes over any long distances or through any perceptual hazards, it is surprising that accuracy in this task was strongly affected by colour coding. Its high error rates mean that, in a trunk- or shaded-coloured map, this junction is positively misleading. (With route colours, the passenger has an even chance of reading it correctly.) We speculate that trunk and shaded colouring increases the perceived separation of station markers on the two lines, as the two Broadway routes are linked by the same colour, while the Washington Heights route is differentiated by its colour.

Speed. Our original hypothesis was that route colouring would enhance the speed of navigation as well as its accuracy. The results neither prove nor disprove the hypothesis, but they do suggest the alternative hypothesis that the tracking component of following a route takes the same time whether it is done correctly or incorrectly. The only subset where colour coding had a significant effect on speed was Jump T, which involves no tracking but the more cognitive task of understanding a junction. It is expected that a correlation between accuracy and speed will emerge in the more cognitive task of journey planning, which we will address in the second part of this project.

4.2   Implications for design

The original impetus for this research was the practical design question of whether to draw a metro map with route or trunk colour coding. It was hypothesised that route colouring would make it easier to carry out the navigational task of tracing a route, but that the visual clutter it creates would make it harder to carry out complex journey planning. It was an open question whether one or the other of those two factors would predominate and therefore whether it is possible to recommend one or the other colour scheme. The present study examines only the route-tracing task; a further study of the effect of colour coding on journey planning is in preparation.

We have found that colour coding does not have a consistent effect across the board. It depends on navigational hazard. Route colouring offers its greatest advantage in tracing routes against the hazards of flipping and joining. If a metro map has a lot of those hazards, but is not so complex that it requires difficult journey planning, then we would expect route colouring to yield a more usable map.

A more general outcome is the identification of distinct navigational hazards, which a designer may address in mapmaking, and software should avoid in automated design.

- Flipping causes the user to lose track of a route, especially in a trunk-coloured map. Change the layout to avoid flips or develop less hazardous flips.
- Joining routes within a trunk causes the user to lose track of a route: make the separation clearer by changing the layout or the colouring.
- If routes are drawn separately in the map, then splitting a route creates a high risk of branch jumping. Never split routes. This argues against hybrid maps.
- If transfers are shown with Vignelli-style dot proximity then always keep the dots aligned, and try to avoid drawing such stations at junctions.

4.3   Implications for automated design

Work on the automated design of metro maps (e.g. Stott & Rodgers, 2004; Wolff, 2007) has centred on algorithmic solutions to the mathematical problem of laying out a metro map subject to multiple criteria of what constitutes of a good layout. The results of the present study indicate that besides the overall layout, the colour coding and specific local navigational hazards need to be addressed. Algorithmic methods for automatically finding and fixing local navigational hazards could have an important role to play.

4.4   Implications for further research.

Inspection of the data from this study reveals that users often make poor choices, either journeys that are longer than necessary, or that involve more transfers than necessary, or that involve attempts to execute non-existent routes or transfers. Work by Roberts has led to a framework for the 'macro design' of a map layout. The present work points to a need for a comparable investigation of 'micro design' to establish the effect of colour, transfer layout and symbolism, and junction layout on usability.

5   Conclusions

There is a strong interaction between the colour-coding scheme and usability in the diagrammatic New York City subway map, as measured by accuracy of navigation when carrying out route-tracing tasks. This can be understood in terms of specific navigational hazards. In a broad class of hazards (Slip F, Slip J, Slip T, Jump T, and Trans P), the route colouring has a large, statistically significant advantage in accuracy. In one type of hazard (Jump R), the pattern is reversed, and trunk-colouring outperforms route-colouring; in two cases (Slip P, Slip U), colour coding has no significant effect.

In terms of the psychological model: We conclude that route-coloured maps become more accurate than trunk- and shaded-coloured maps when routes that belong to the same trunk behave differently by branching and switching places, and thereby divide the user's attention, as opposed to simply running in parallel. We also conclude that route-coloured maps become less accurate if the colour coding is applied inconsistently (for example, route splitting), thereby confusing the user's thinking.

These results indicate that local navigational hazards can play a major role in the overall usability of a metro map. And that the choice of a colour-coding scheme that would yield the most usable map in real-life usage must depend on the comparative frequency of different kinds of hazard in the map, and the comparative frequency of different navigations (route tracing *versus* journey planning) undertaken by passengers.

References

Amazon.com, Inc.: Mechanical Turk, `https://www.mturk.com/mturk/`, last accessed 2017/11/28

Barrington and Company: A Survey of Prospective User Reaction to the New Subway Map, May 31, 1966, report submitted to the NYCTA. Collection of Raleigh D'Adamo (1966)

Bronzaft, A.L., Dobrow, S.B., O'Hanlon, T.J.: Spatial orientation in a Subway System, Environment & Behavior, 8:575-594 (1976)

Bronzaft, A.L., Schachter, J.: memo to MTA Subway Map Committee, March 3, 1978. New York Transit Museum (1978)

Crump, M.J.C., McDonnell, J.V., Gureckis, T.M.: Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. PLoS ONE8(3): e57410 (2013). doi: 10.1371/journal.pone.0057410

Gadiraju, U., Möller, S., Nöllenburg, M., Saupe, D., Egger-Lampl, S., Archambault, D., Fisher, B.: Crowdsourcing Versus the Laboratory: Towards Human-Centered Experiments Using the Crowd. In: Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments. pp. 6-26. Springer, Berlin (2017). doi:10.1007/978-3-319-66435-4_2

Field, K., Cartwright, W.: Becksploitation: The Over-Use of a Cartographic Icon. The Cartographic Journal 51(4), 343-359 (2014). doi: 10.1179/0008704114Z.000000000150

Gallotti, R., Porter, M.A., Barthelemy, M.: Lost in Transportation: Information measures and cognitive limits in multilayer navigation. Science Advances 2, e1500445 (2016). doi:10.1126/sciadv.1500445

Garland, H.C., Haynes, J.J., Grubb, G.C.: Transit Map Color Coding and Street Detail: Effects on Trip Planning Performance. Environment and Behavior 11(2), 162-184 (1979)

Guo, Z.: Mind the map! The impact of transit maps on path choice in public transit. Transportation Research Part A 45, 625-639 (2011). doi:10.1016/j.tra.2011.04.001

Guo, Z., Zhao, J., Whong, C., Wyman, L.: Redesigning subway map to mitigate bottleneck congestion: An experiment in Washington DC using Mechanical Turk. Transportation Research Part A 106, 158-169 (2017). doi:10.1016/j.tra.2017.09.017

Kick Design: KickMap, `http://kickmap.com/`, last accessed 2017/11/29

KVB (Kölner Verkehrs-Betriebe AG): Bahnen in Köln. `http://www.kvb-koeln.de/german/fahrplan/linienplan.html`, last accessed 2017/11/29 (2017)

Lloyd, P.B.: Vignelli: Transit Maps. RIT Press, Rochester, NY (2012)

MacEachren, A.M.: How maps work: Representation, Visualization, and Design. Guilford Press, New York, NY (1995)

Mason, W., Suri, S.: Conducting behavioral research on Amazon's Mechanical Turk. Behavioral Research Methods 44(1), 1–23 (2012). doi:10.3758/s13428-011-0124-6

Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63(2), 81–97 (1956)

Montello, D.M.: Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches. Cartography and Geographic Info. Science 29 (3), 283-304 (2002)

MTA (Metropolitan Transportation Authority, New York, NY): The Weekender. `http://web.mta.info/weekender.html`, last accessed 2017/11/29

Ovenden, M.: Transit Maps of the World. Penguin, New York, NY (2015)

Ovenden, M., Lloyd, P.B., Pepinster, J.: Paris Métro Style. Capital Transport, Harrow (2008)

Paolacci, G., Chandler, J., Ipeirotis, P.G.: Running experiments on Amazon Mechanical Turk. Judgment and Decision Making 5(5), 411–419 (2010)

Roberts, M.J.: Underground Maps Unravelled. Self-published, Wivenhoe, UK (2012)

Roberts, M.J., Newton, E.J., Canals, M.: Radi(c)al departures: Comparing conventional octolinear versus concentric circles schematic maps for the Berlin U-Bahn/S-Bahn networks using objective and subjective measures of effectiveness. Information Design Journal 22(2), 92-115 (2016). doi:10.1075/idj.22.2.04rob

Roberts, M.J.: Schematic maps in the laboratory. In: First International Schematic Mapping Workshop University of Essex, 2nd–3rd April 2014. `http://www.tubemapcentral.com/articles/Roberts_empirical_SMW2014_2014.pdf`, last accessed 2017/11/29 (2014a)

Roberts, M.J.: What's your theory of effective schematic map design? ibid. `http://repository.essex.ac.uk/11368/1/Roberts-Theoretical.pdf`, last accessed 2017/11/29 (2014b)

Roberts, M. J., Rose, D.: Map-induced journey-planning biases for a simple network: A Docklands Light Railway study. Transportation Research Part A 94, 446–460 (2016). doi:10.1016/j.ijhcs.2016.06.003

Robinson, A.H.: The look of maps. University of Wisconsin Press, Madison, WI (1952)

Stott, J.M., Rodgers, P.: Metro Map Layout Using Multicriteria Optimization. In: Proceedings of 8th International Conference on Information Visualisation (IV04), IEEE, 14th-16th July 2004. 355-362. `https://kar.kent.ac.uk/14133`, last accessed 2017/11/29 (2004). doi:10.1109/IV.2004.1320168

Wolff, A.: Drawing Subway Maps: A Survey. Informatik—Forschung und Entwicklung 22, 23-44 (2007). doi:10.1007/s00450-007-0036-y