

## London taxi drivers exploit neighbourhood boundaries for hierarchical route planning

Eva-Maria Griesbauer<sup>a,b,1</sup>, Pablo Fernandez Velasco<sup>c,1,\*</sup>, Antoine Coutrot<sup>d</sup>, Jan M. Wiener<sup>e</sup>,  
Jeremy G. Morley<sup>b</sup>, Daniel McNamee<sup>f</sup>, Ed Manley<sup>g,h</sup>, Hugo J. Spiers<sup>a,\*</sup>

<sup>a</sup> Institute of Behavioural Neuroscience, Department of Experimental Psychology, University College London, London, UK

<sup>b</sup> Ordnance Survey Ltd, Southampton, UK

<sup>c</sup> University of York, York, UK

<sup>d</sup> Univ Lyon, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69621 Lyon, France

<sup>e</sup> Department of Psychology, Bournemouth University, UK

<sup>f</sup> Neuroscience Programme, Champalimaud Research, Centre for the Unknown, Lisbon, Portugal

<sup>g</sup> School of Geography, University of Leeds, Leeds, UK

<sup>h</sup> Centre for Advanced Spatial Analysis, University College London, London, UK

### ARTICLE INFO

#### Keywords:

Spatial representation  
Real-world evidence  
Hierarchical representations  
Spatial cognition  
Wayfinding  
Executive function

### ABSTRACT

Humans show an impressive ability to plan over complex situations and environments. A classic approach to explaining such planning has been tree-search algorithms which search through alternative state sequences for the most efficient path through states. However, this approach fails when the number of states is large due to the time to compute all possible sequences. Hierarchical route planning has been proposed as an alternative, offering a computationally efficient mechanism in which the representation of the environment is segregated into clusters. Current evidence for hierarchical planning comes from experimentally created environments which have clearly defined boundaries and far fewer states than the real-world. To test for real-world hierarchical planning we exploited the capacity of London licensed taxi drivers to use their memory to construct a street by street plan across London, UK (>26,000 streets). The time to recall each successive street name was treated as the response time, with a rapid average of 1.8 s between each street. In support of hierarchical planning we find that the clustered structure of London's regions impacts the response times, with minimal impact of the distance across the street network (as would be predicted by tree-search). We also find that changing direction during the plan (e.g. turning left or right) is associated with delayed response times. Thus, our results provide real-world evidence for how humans structure planning over a very large number of states, and give a measure of human expertise in planning.

### 1. Introduction

Being able to plan for the future is a critical cognitive ability for humans. How humans might enact spatial planning has been the topic of a wide range of research (Brown & Chrastil, 2019; Chrastil et al., 2015; Ekstrom et al., 2018; Epstein et al., 2017; Hartley et al., 2003; Horner et al., 2016; Mattar & Lengyel, 2022; Fernandez-Velasco and Spiers, 2024; Miller & Venditto, 2021; Newcombe, 2018; Patai & Spiers, 2021; Spiers & Gilbert, 2015; Weisberg & Newcombe, 2016). One common approach is to consider tree-search algorithms. Computationally, tree-search algorithms operate by repeatedly sampling different

alternatives at each decision point according to some heuristic strategies and effectively compare entire sequences of decisions (Shannon, 1950; Daw, Niv, & Dayan, 2005; Huys et al., 2012; van der Ham & Claessen, 2020). More recent approaches involve Monte Carlo tree-searches that randomly sample trajectories potentially guided by state evaluation functions (Browne et al., 2012), pruning strategies that eliminate unfavourable options (Huys et al., 2012), and reinforcement-based strategies that estimate the overall likelihood of a route being chosen (Botvinick et al., 2009; O'Doherty et al., 2015). These models have managed to describe quite closely human and animal trajectories in small-scale and virtual reality environments (Daw et al., 2011; De Cothi

\* Corresponding authors.

E-mail addresses: [p.fernandezvelasco@gmail.com](mailto:p.fernandezvelasco@gmail.com) (P. Fernandez Velasco), [h.spiers@ucl.ac.uk](mailto:h.spiers@ucl.ac.uk) (H.J. Spiers).

<sup>1</sup> Equal contribution

et al., 2022; Miller & Venditto, 2021; Russek et al., 2017). Nevertheless, because the number of possible routes grows exponentially with each extra step to evaluate, tree-search becomes computationally intractable in large-scale environments such as most cities (Gershman et al., 2015).

A promising development that could measure up to the computational demands of real-world navigation is hierarchical route planning (Bast et al., 2016; Botvinick et al., 2009; McNamee et al., 2016; O'Doherty et al., 2015; Tomov et al., 2020). Hierarchical models segregate the environment and represent it through smaller, distinct areas that contain local information about specific places. The distinct areas are referred to as clusters, and the specific places as states. Instead of involving the entire environment, computations are then restricted to a subset of clusters and states. Route planning is first carried out across clusters / subregions to select the relevant ones. Within each cluster, a particular route is planned using the limited number of states available. Ultimately, this results in multiple sequences of shorter routes. This makes intuitive sense when we consider that cities will lend themselves to clustering in the different regions that evolve within its layout (Farzanfar et al., 2023; Filomena et al., 2019; Lynch, 1964).

Empirical studies of human navigation in small regionalized virtual environments have provided evidence of hierarchical route planning (Balaguer et al., 2016; Wiener et al., 2004; Wiener & Mallot, 2003). Experimental work has also shown that humans decompose tasks in optimal ways for planning action sequences (Solway et al., 2014). Evidence for hierarchical route planning has also come from examination of reaction times of participants planning routes in a fictitious transport network, where the number of trainline interconnections, rather than the number of stations, has a significant impact on planning times (Balaguer et al., 2016).

Real-world evidence for hierarchically structured route planning, however, is scant. Taxi drivers in Paris have been studied to explore how they represent the city and plan routes (Chase, 1983; Pailhous, 1969). While this has shown evidence for use of the major street network to apply to planning routes (Pailhous, 1969) and greater overestimation of distances when two reference places were separated by neighbourhood boundaries (Chase, 1983), the studies do not provide an explicit test of hierarchical planning, such as has been used in lab studies (e.g. Balaguer et al., 2016).

One possible approach to exploring hierarchical planning is in the mental simulations of real-world routes, which have shown evidence of clustering of locations into regions and compressed representations of walking times or distances (Arnold et al., 2016; Bonasia et al., 2016; Brunec et al., 2017; Jafarpour & Spiers, 2017). Likewise, the duration of mental planning of real-world routes can indicate regional hierarchies in a spatial environment (McNamee et al., 2016). Nonetheless, these findings concern planning times for a route as a whole. What has not been obtained to date is evidence from step-by-step planning in real-world, ecologically valid settings. The importance of this lacuna comes to the fore when we contrast the small state-spaces employed in controlled experimental setups (e.g. Balaguer et al., 2016; Muhle-Karbe et al., 2023; Wiener et al., 2004; Wiener & Mallot, 2003) and large street networks of the real-world (Griesbauer, Manley, McNamee, et al., 2022). Lab experiments have used bounded spaces with artificially imposed accentuated hierarchies. Real-world cities, on the other hand, lack any straightforward segregation into clusters, often with no straightforward boundaries (Griesbauer, Manley, McNamee, et al., 2022). For instance, the subway network in the study by Balaguer et al. (2016) was designed so as to be hierarchically segregated, and human cognition apparently leverages this structure for planning. It remains unclear where similar response patterns would be evident for route planning in real large-scale urban environments that lack such sharp hierarchical segregation.

The lack of defined boundaries is one of three methodological challenges presented when testing for real-world evidence of hierarchical route planning. The other two are the participants' ability to give precise route descriptions, and familiarity. One can reliably expect participants to remember stops in a relatively small subway network and

give clear descriptions of possible routes, but for an urban street network spanning tens of thousands of streets, there are very likely significant gaps in the spatial knowledge of typical residents. Finally, participants will have varying levels of familiarity with different regions of existing urban settings. Familiarity has been shown to affect performance in spatial tasks (O'Neill, 1992) and to distort representations of distances within an environment (Jafarpour & Spiers, 2017).

Here, we surmount these challenges by testing for hierarchical route planning in London licensed taxi drivers. Licensed London taxi drivers, also known as London cabbies (hereafter referred to as London taxi drivers), are famous for their thorough training on "The Knowledge of London", which covers both main and secondary roads within the six-mile area around Charing Cross in central London, UK (see Griesbauer, Manley, Wiener, & Spiers, 2022). Years of driving experience in London, accumulated during training, as well as post-qualification during work, make such taxi drivers excellent navigators of London, and in contrast to the general population, they are trained to recall routes in all areas of London by giving precise travelling instructions and specific street names for each step of a journey through London (Griesbauer, Manley, Wiener, & Spiers, 2022). In addition, taxi drivers have a familiarity with all areas of London. Previous research with London taxi drivers has provided insights into the dynamics of cognition, neural plasticity and brain function during driving and navigating London (Maguire et al., 2000; Maguire, Woollett, & Spiers, 2006; Maguire, Nannery, & Spiers, 2006; Spiers & Maguire, 2007a,b,c, 2008; see for review Griesbauer, Manley, Wiener, & Spiers, 2022). Here we probe their capacity for planning. Towards this, we use the boundaries in London that are consistently recognised by London taxi drivers, e.g. Soho, Mayfair (based on Griesbauer, Manley, McNamee, et al., 2022). Notably, many boundaries noted on official maps (eg. 'The City of London') were not consistently considered as boundaries. As a result, we were able to reliably segment London into different clusters that would be meaningful to this population.

We examined the response time to state the name of each street in the sequential planning of routes between a given origin and a destination pair (e.g. Kings Cross Station, to Paddington Station). If London taxi drivers employ hierarchical structures when planning routes, then response times should differ depending on the clusters they are planning over (streets on the boundary of regions vs other streets). If a tree-search is used to plan, the planning response time should be higher when there are more streets ahead to be evaluated as part of the plan. If they are exploiting boundary streets to plan they may be faster to respond when they need to select these streets as they would have been pre-selected in a hierarchical plan (McNamee et al., 2016).

The street network also contains additional features that may impact route planning. These include turning left or right, road classifications into main and minor roads, and Euclidean or path distance to the destination. Turns have been found to act as conceptual boundaries that have an impact on memory recall of route features (Brunec et al., 2020; Kuipers et al., 2003; Lloyd, 1989), on the estimation of path distances (Brunec et al., 2017; Hutcheson & Wedell, 2009; Sadalla & Staplin, 1980), and on preferences towards routes with fewer turns (Broach et al., 2012; Elliott & Lesk, 1982; Venigalla et al., 2017). Thus, turns may likely lead to slower response times. As for major roads (e.g. UK 'A roads'), they might be better remembered than minor roads due to increased familiarity, as was the case with Parisian taxi drivers (Pailhous, 1969), which would favour a faster recall. Alternatively, major roads are associated with higher betweenness centrality in the street network (Javadi et al., 2017) and thus more long-range potential options for a plan, leading to predicted slower response times. Past research indicates distortions in the representations of space when a path must circumnavigate a region leading to high circuitry metrics in a path (e.g. long path distance to a destination, but a short Euclidean distance) (Brunec et al., 2017). Thus, it seems possible that response times may be impacted by the degree of circuitry of routes (e.g. straight routes vs U-shaped routes).

## 2. Methods

### 2.1. Participants

44 licensed London taxi drivers were tested after written informed consent. One participant displayed extremely unstructured route recalls that did not allow for a transcription of the routes, thus data from 43 taxi drivers (41 males, 2 females) was analysed. Their mean age was 53.82 years ( $SD = 10.35$ ; range: 34–75 years) and their mean experience driving a taxi was 19.61 years ( $SD = 15.69$ ). 19 taxi drivers participated in the first round of data collection (all male, age = 52.94,  $SD = 9.81$ , experience = 19.97,  $SD = 16.52$ ) and 24 taxi drivers at the second round of data collection (22 male, 2 female; age = 54.50,  $SD = 10.93$ ; experience = 19.30,  $SD = 15.32$ ). All of the taxi drivers were native English speakers. Data was collected in two periods with an interval of six months between the collection times. The sample size corresponded to resource constraints that had to do with recruitment (Lakens, 2022). At the end of the first collection period, sample was judged (heuristically) to be too small for this type of real-world type of experiment, and a second cohort was recruited. All procedures were approved by the ethics committee (CPB/2013/150 and EP/2018/008).

### 2.2. Materials

The task for the first data collection consisted of 12 origin-destination pairs (e.g. from Chelsea harbour to London heliport). For the second data collection, two origin-destination pairs of the original set remained and 6 new pairs were added which increased the variety of route planning problems across London. From the two routes that remained the same, one origin location had to be updated (from Joe Allen's restaurant to Bill's) because the restaurant had moved location. The new origin (Bill's) was a neighbouring restaurant to the original (Joe Allen's restaurant). All pairs were chosen to vary in their geographical properties (e.g. path length, Euclidean distance, direction of travel; see Table S1). Euclidean distance was designed to be relatively decorrelated from the number of streets to the goal and potential boundaries that had to be crossed. This was to allow us to determine whether each type of distance impacted planning. For instance, route 7 and 18 had a similar number of streets (8 expected streets for route 7 and 7 streets for route 18), but varied in their planning distance (~11 km for route 7 and ~1.5 km for route 18). On the other hand, route 7 and route 17 were similar in their planning distance (~11 km vs ~8.4 km) but varied in the number of streets that had to be recalled (8 streets vs 14 streets). Similarly, the number of boundaries that we expected to affect routes varied across routes: Some routes did not cross any boundaries (e.g. route 4, 6 or 10), while other routes required at least partially to mentally travel along boundaries (e.g. route 7, 12 or 14) or crossing several boundaries (e.g. routes 1, 5, 13, 16 or 17). In collaboration with a London taxi driver training school (See Griesbauer, Manley, Wiener, & Spiers, 2022 for details), teachers provided feedback to ensure the validity of the selected route with regards to route planning properties. Two SONY ICD-PX240 Mono Digital Voice Recorders were used in this experiment. One of the recorders was used to replay pre-recorded instructions and the route planning tasks. With the second recorder, the experimenters recorded the whole duration of the experiment, from the initial task presentation to the final route recall.

### 2.3. Experimental design

Licensed London taxi drivers were recruited in the area of Bloomsbury and the borough of Camden, London (UK). Before participating in the study, taxi drivers gave written consent and filled in a personal questionnaire to indicate age, gender, experience, whether they were native speakers, and whether they had taken part in this study on a previous occasion. After participating, taxi drivers received monetary compensation.

The group of taxi drivers who participated in the first study were verbally informed that they were to plan routes (*runs*) between origin-destination pairs (*points*) and that these routes would be presented through audio recordings. Specifically, participants were asked to “call out the run” between the origin and the destination. This terminology (“calling out a run”) is in line with the activity that taxi drivers undergo in their training, and it involves the quick recall of the street names and turns, in order, along the shortest route between the origin and the destination. Participants were asked to listen carefully to the instructions, and they were warned that repeating instructions was not possible. If they did not understand or know either of the points, they could skip the route or carry it out where they perceived the points to be. In order to avoid interfering effects from obstructions unrelated to street network properties, participants were instructed to disregard possible congestion and any temporary obstructions in the street network. As some traffic rules can change in dependence of day and time, participants were prompted to imagine they were carrying out the route planning task on a typical Monday morning around 11.00 AM, to keep conditions consistent. Finally, in order to avoid distractions from the planning process and to elicit a structured recall across all drivers, participants were instructed to focus on the route planning as if they were in a ‘knowledge examination’ situation (Griesbauer, Manley, Wiener, & Spiers, 2022) and to refrain from questions, comments or explanations.

Instructions to follow the structured recall as in the examination situation were given only before the first and the second route plan, but drivers received a reminder of day and time (i.e. Monday morning at 11.00 AM) before each route. These reminders had been audio recorded and were presented together with the audio recordings of the twelve route planning in the following format:

“Please remember to do this under appearance conditions. So, no questions or clarifications, when you hear the points. If you're unsure, start or go where you think it is, or skip the run.” (Before route 1 and 2).

“Please remember, you're doing the next run on a Monday morning around 11 am.” (Before each route).

“Please call out the run from [Point, Street] to [Point, Street].” (Task presentation).

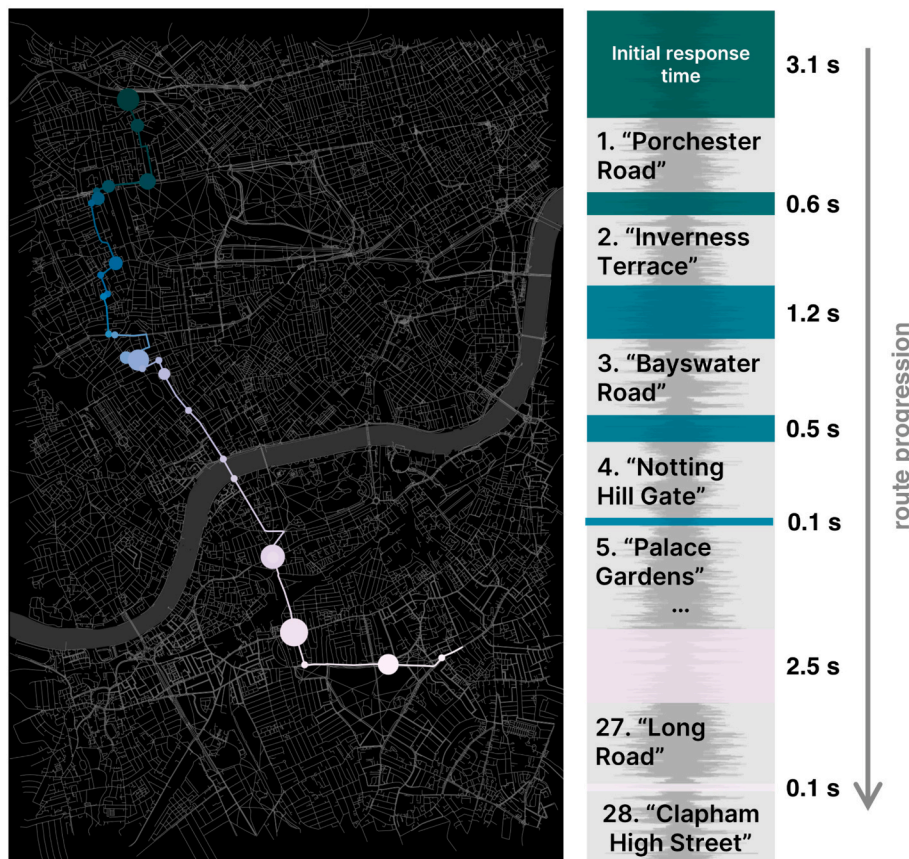
Drivers listened to the set of instructions and then planned the route before moving on to the next set of instructions and route planning task (for an illustrative example of a recall block, see Fig. 1). The whole sequence from the first instruction to the final route recall was recorded on a second audio-recording device.

For the second group of taxi drivers, who were presented with a modified set of eight origin-destination pairs, the following additional modifications were put in place to improve the procedure: During the first period of data collection, several taxi drivers repeatedly asked for clarifications of the origins or destinations or could not remember which location was named. Thus, several routes were skipped by drivers. Therefore, ‘flashcards’ were provided stating the location and the corresponding street for both origin and destination. These flashcards were shown to the drivers directly after they had listened to the audio recorded task and stayed visible during the entire route recall of a task. Additionally, after completing the first and second routes, drivers were asked for confirmation that the recalled route reflected what they would have done on a Monday morning at 11.00 am, which all taxi drivers confirmed.

### 2.4. Data transcription

The collected audio data of the recorded route recalls was transcribed in terms of street names and response times. Initial response times (i.e. the time between task instructions and the first street named) reflected route planning behaviour that included a variety of actions and

## “Please call out the run from the Royal Oak tube station, to Revolution, Clapham High Street”



**Fig. 1.** Example of one taxi driver’s plan for one of the eighteen origin-destination pairs probed. Taxi drivers received pre-recorded audio instructions of the origin and destination location and planned the route before being presented with the next route. All instructions and route recalls were audio recorded (top) and transcribed to extract response times and visualise planned routes (bottom right). The time between given instructions and the first recalled street was defined as the *initial response time*. Time between recalled streets of the route were defined as *response times between streets*. See Fig. 2 for all the routes taken for this origin-destination pair. This figure was generated using Open Street Map data and the *osmdata* package in R.

processes inconsistent across drivers or routes (e.g. affirmative questions concerning locations). These actions were not consistent across or within taxi drivers or routes, and could not be separated from each other analytically, because most planning was carried out silently to ensure natural planning behaviour of the drivers. Therefore, initial planning times have been excluded from the analysis of response times. In contrast, *response times between streets* (i.e. pauses between two consecutively named streets) were part of the sequential recall of street names and reflected planning behaviour directly related to each point of the recall process, not leaving opportunities for unrelated planning actions (see Figs. 1 & 2).

The transcription of all response times was carried out with the free and open-source audio software Audacity, versions 2.2.2 and 2.3.1, which allowed for an accuracy of up to 0.1 s. The street names were transcribed and corrected for mistakes (e.g. “Townsend Rd” to “Townmead Rd”) or unified (e.g. “Charles the 1st” to “Trafalgar Square”) to ensure comparability of response times at each street or place. The analysis of the complete dataset was carried out in R (version 4.0.2). To ensure the reliability of transcribed response times, an inter-rater reliability test was carried out for both studies. The intraclass correlation coefficient (ICC) for the first study (transcribed by two coders), was assessed through a two-way, mixed effects, absolute agreement, single-measures model. The ICC ( $ICC = 0.98, p < .001$ ) was in the excellent range, suggesting a similar transcription of response times due to a high agreement between the two coders. For the second study (transcribed by

four coders), a one-way random effects model with absolute agreement was used. The ICCs indicated a moderate range of agreement ( $ICC = 0.61, p < .001$ ), which was considered acceptable for a rating involving four coders. A Wilcoxon Signed-rank test indicated no group differences between the two sets of taxi drivers for routes 7 and 8 for log-transformed ( $Mdn(S1) = -0.22, IQR = 1.18, Mdn(S2) = -0.16, IQR = 1.24, p = .252, r = 0.046$ ) or z-standardised ( $Mdn(S1) = -0.29, IQR = 1.11, Mdn(S2) = -0.33, IQR = 1.24, p = .688, r = 0.016$ ) response times between streets. Accordingly, the data from both instances of data collection was treated as one data set.

### 2.5. Data analysis

We used a linear mixed effects model to test the effect of street network variables on the log-transformed response times (c.f. Coutrot et al., 2018). Note that these are response times between streets, not overall response times. Taxi drivers and their routes were entered as random effects to account for individual differences and potential correlations between repeated measures. The fixed effects variables of the model were *boundaries (B)*, *turn actions (T)*, *number of streets (N)*, *road type (R)* and *Euclidean distance (E)*. Here, the *boundaries* reflected agreement rates across taxi drivers in percentages. This was based on a previous study (Griesbauer, Manley, McNamee, et al., 2022), in which taxi drivers were asked to indicate streets they considered as boundaries for London districts or dividing areas. The agreement on boundary



**Fig. 2.** Overlay of the routes planned by all taxi drivers for a single origin-destination pair. Taxi drivers received pre-recorded audio instructions of the origin (Royal Oak tube station) and destination (Revolution, Clapham High Street) locations and planned the route. All instructions and route recalls were audio recorded and transcribed to extract response times and visualise planned routes (in yellow). Time between recalled streets of the route were defined as *response times between streets*. Larger circumferences on the map (in yellow) correspond to longer response times between streets. All the responses from all participants for a single origin-destination pair have been overlaid on the map of London. This figure was generated using Open Street Map data and the `osmdata` package in R. For a. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

streets (between taxi drivers) varied considerably, from some boundaries providing almost no consensus to some boundaries consistently noted as boundaries (see Fig. 3). Here, we use the agreement rate for each boundary as a fixed effect variable.

The *number of turns* was extracted from the route recall, and corresponded to the number of turns on the route. Turns were coded categorically as *turn* where a change in direction occurred between consecutive streets or as a *forward* action where streets continued straight without a change of direction. Note that taxi drivers are required in the exam (and in our experiment) to state either forward or turn. The *number of streets* to the destination was counted down, with the last street having the value 1 and the first street the value  $n$ , if a driver recalled  $n$  numbers of streets for a route. *Road type* was categorised as either a main road (either trunk roads or primary roads), or other roads (i.e. all remaining road types, such as secondary, tertiary, or residential roads). Here, data about road type classifications was extracted from the OS MasterMap Integrated Transport Network (ITN) Layer (2018). *Euclidean distance* to the destination was calculated from each

intersection of two consecutive streets to the destination (i.e. this was the distance between the reported step on the route corresponding to the response time in the regression, and the destination, not between origin and destination). Data was extracted from OS MasterMap Integrated Transport Network (ITN) Layer (2018). These five fixed effects were decorrelated as all variance inflation factors were below 2.5. The basic model, which was used to describe log-transformed response times between streets, had the following structure:

$$\text{Log (RTs)} \sim 1 + B + T + N + R + E + (1 | \text{Taxi Driver}) + (1 | \text{Route}) \quad (1)$$

*Model 1* was designed to provide an initial assessment of the data based on factors found in the literature. Follow up analysis was carried out to examine in more detail the effect of boundaries, and two alternative models were considered. One alternative model (*Model 2*) replaced boundary streets with all the streets in the London district of *Soho* (*S*), including its boundaries. For each route, it is not about whether the route as a whole passes through Soho, but about the section of the route that passes streets in Soho. We chose to focus on Soho because it was the only district that conceptually appears as an ‘island’ with sharply defined boundaries in the central London street network (Griesbauer, Manley, McNamee, et al., 2022) and appears to be treated distinctly by London taxi drivers (Maguire, Nannery, & Spiers, 2006). In contrast, Mayfair and Belgravia are not entirely surrounded by other urban areas because they share boundaries with Hyde Park, a green space that is conceptually different from urban spaces. The second alternative model (*Model 3*) examined the impact of *circuitry*, a similar concept to U-turn costs. This analysis was carried out because Balaguer et al. (2016) had found a cost of U-turns where participants had to head back along a subway line to reach the destination. Here, circuitry was defined as the fraction of path distance to the destination divided by Euclidean distance to the destination, both calculated from each street. The closer the circuitry value to 1, the more similar is the travelled route (path distance) to a straight line (Euclidean distance). The larger the circuitry value, the more deviation there is of the path from a straight line. Since circuitry and Euclidean distance to the goal were not independent from each other, Euclidean distance was replaced by circuitry in Model 3. Hence, the two alternative models were (changed variable in bold):

$$\text{Log(RTs)} \sim 1 + \mathbf{S} + T + N + R + E + (1 | \text{Taxi Driver}) + (1 | \text{Route}) \quad (2)$$

$$\text{Log(RTs)} \sim 1 + B + T + N + R + \mathbf{C} + (1 | \text{Taxi Driver}) + (1 | \text{Route}) \quad (3)$$

**S** = Soho streets, replacing boundaries from Model 1.

**C** = Circuitry, replacing Euclidean distance in Model 1.

### 3. Results

#### 3.1. Overview of the data

The data, which was collected from  $N = 43$  licensed London taxi drivers (green badge holders), included the recall of 354 routes (first study: 173, second study: 181). The mean for *initial planning times* was  $M = 13.83$  s ( $SD = 13.40$ ; Table 1) over  $N = 315$  routes. Data of some initial planning events had to be removed as drivers asked for clarifications and tried to engage in conversations. On average, taxi drivers recalled 9.1 out of 12 routes during the first study, and 7.5 out of 8 routes during the second study.

There were a total of 3398 responses, with a mean *response time between streets* of  $M = 1.82$  s ( $SD = 3.24$ ). The *total response duration* for routes, a measure to reflect the total planning duration based on responses between streets per driver and route, was  $M = 17.53$  s ( $SD = 16.47$ ). See Fig. 2 for an example of the routes chosen for an origin-destination pair. See Fig. 6 for examples of actual routes where groups of taxi drivers took the same path to show the dynamics in response times. For the raw dataset, these response times between streets were skewed towards minimal response times (see Fig. S1). Log-



**Fig. 3.** Overlay of boundaries. Adapted from Griesbauer, Manley, McNamee, et al. (2022). The process to generate the boundaries was as follows: The original paper drawings (a) were digitised in geoJSON.IO (b). A layover of all drawings (c) was created. Lines with higher opacity, e.g. boundaries around Mayfair and Soho (e), indicate a higher agreement across drivers on a boundary street than for streets with low opacity, e.g. City of London (d). Sources: (a) © Crown copyright and database rights [2021] Ordnance Survey (100025252); (b) Hanson & Seeger (2016); (c–e) Overlaid mapping created by Melda Salhab.

**Table 1**  
Variable Summary for recalled data.

	N <sup>1</sup>	Mean	SD	Min	Max
Initial Planning Time <sup>2</sup> (sec)	315	13.83	13.40	1.00	102.00
Mean Response Times between Streets <sup>2</sup> (sec)	3398	1.82	3.24	0.10	61.4
Total Response Duration <sup>2,3</sup> (sec)	354	17.53	16.47	0.20	89.40
Total no of Streets Recalled	354	10.71	5.71	3.00	30.00

<sup>1</sup> Number of occurrences in data set.  
<sup>2</sup> Minimum coding accuracy of response times: 0.1 s.  
<sup>3</sup> Mean of the sum across all response times between Streets per taxi driver and per planned route.

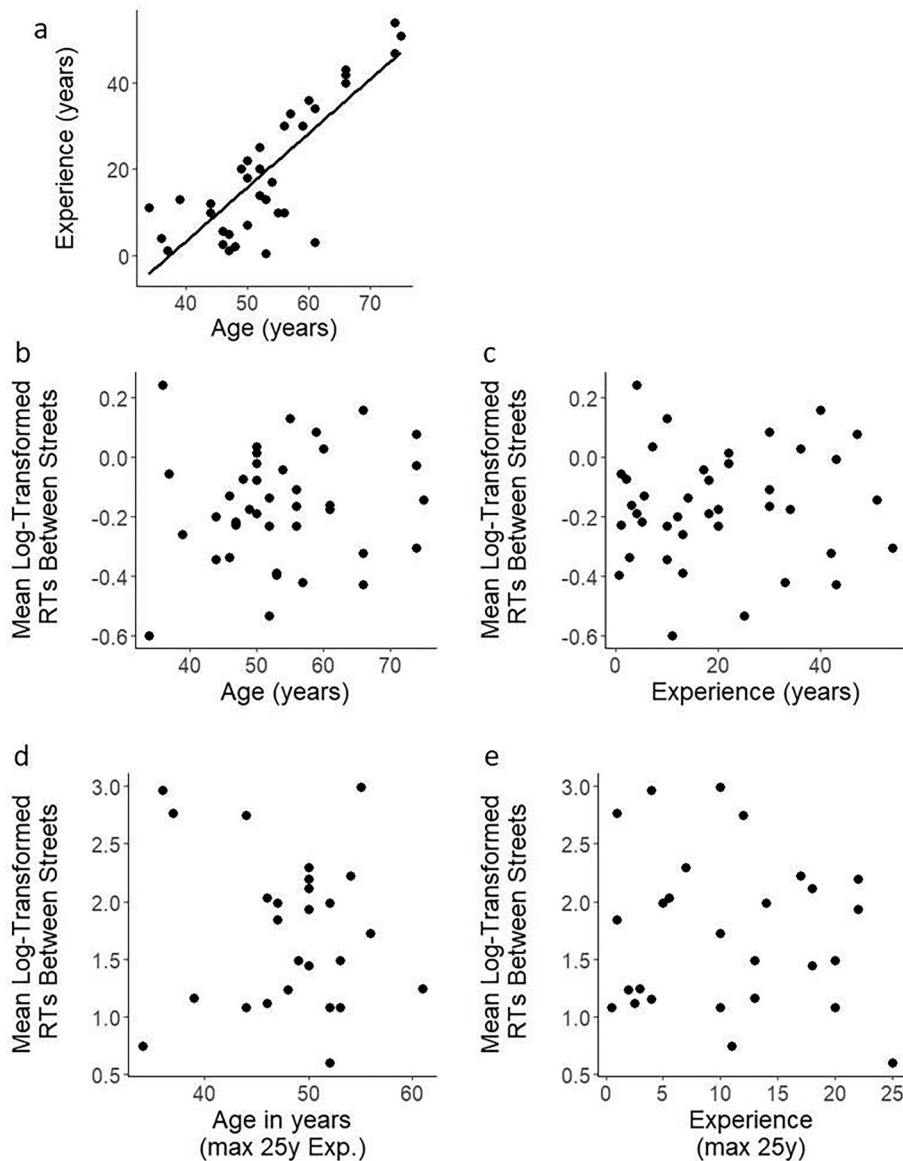
transformation (using the natural logarithm) and z-standardisation provided a better fit with a normal distribution, with the exception of high density at minimal response times of the log-transformed data. This was also reflected across routes (see Fig. S2). Mean response times between streets ranged from 0.7 s for route 10 to 2.8 s for route 9 and violin plots indicated skewness of raw data towards fast recalls between streets. These were ordered by ascending means of z-transformed data to allow for comparison across routes whilst accounting for individual differences between drivers. A high number of outliers for the raw and z-transformed data at each route highlighted slow responses between named streets with up to 60s. Log-transformed violin plots showed two high density peaks of data, one near fast response times around 0.1 s (log-values around -1) and a second density peak near values of 1 s (i.e. log-values around 0). For this study, the high density of fast response

times (around 0.1 s) was thus considered as a measurable lower bound on response times from the audio recordings. In contrast, deviations towards slower responses (i.e. second density peak of log-transformed data and outliers in the raw and z-transformed data set) were expected to convey information about spatial structures indicative of potential planning of new sequences (and therefore of hierarchical planning). Similar to response times between streets, the means of response times between streets by route, the total number of recalled streets per route, and initial response times were all left-skewed for raw data. After log-transformation and z-standardisation initial response times fit a normal distribution (see Fig. S1). No relation was found between raw or transformed initial response times and mean response times between streets. Thus, there was no evidence that those who engaged in longer initially thinking about the route were then faster at calling it out street-by-street.

### 3.2. Age and experience

We considered that with experience taxi drivers might improve with

their navigation, and conversely with age they might decline (see e.g. Coutrot et al., 2018). Thus, we studied the relationship between age ( $M = 53.82$  years,  $SD = 10.35$ ) and experience driving a taxi ( $M = 19.61$  years,  $SD = 15.69$ ) for the group of 43 taxi drivers (see Fig. 4). A Spearman correlation indicated a strong, significant positive relation of age and experience ( $r_s(43) = 0.73$ ,  $p < .001$ ). No relation was found between means of log-transformed response times by route and experience ( $r_s(40) = 0.08$ ,  $p = .623$ ) or age ( $r(37) = 0.13$ ,  $p = .445$ ) for the entire group of taxi drivers. Age and experience were decorrelated in a group of taxi drivers with 25 years of experience or less ( $r(24) = 0.23$ ,  $p = .564$ ). In that sub-group of taxi drivers there was also no relation between means of log-transformed response times by route and experience or age (experience:  $r(25) = -0.11$ ,  $p = .584$ ; age:  $r(24) = -0.11$ ,  $p = .603$ ). Thus, when separating age and experience, we find no evidence of decline with age and no gain with experience.

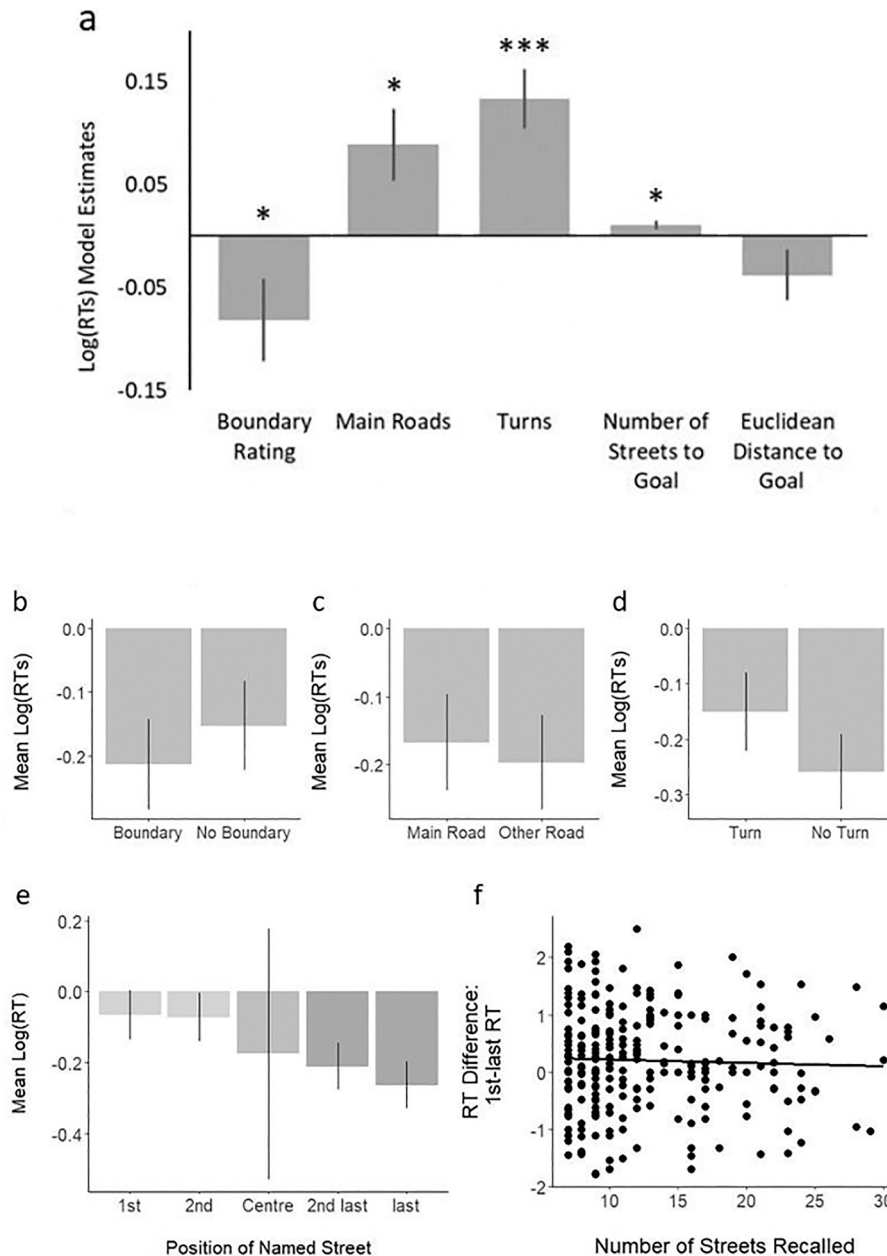


**Fig. 4.** Age-experience relationship. Age and experience across taxi drivers was strongly correlated for the entire group of participants (a), but decorrelated for a subset of drivers with 25 years of experience or less. Neither the entire group of taxi drivers (b, c) nor the subset of drivers with 25 years of experience or less (d, e) showed any relation to mean log-transformed response times between streets.

### 3.3. Turns, boundaries and number of streets to the goal impact response times

Model 1 revealed that *boundaries* ( $b = -0.082, p < .05, 95\% \text{ CI} = -0.047 - -0.12$ ), *turns* ( $b = 0.13, p < .001, 95\% \text{ CI} = 0.17 - 0.093$ ), *number of streets to the destination* ( $b = 0.011, p < .05, 95\% \text{ CI} = 0.040 - -0.018$ ), and *main roads* ( $b = 0.089, p < .05, 95\% \text{ CI} = 0.12 - 0.054$ ) but not *Euclidean distance* ( $b = -0.038, p = .126, 95\% \text{ CI} = 0.013 - 0.063$ ) had significant effects on *response times* (Fig. 5). Coefficients of the model are reported in Fig. 5a, these parameters were additionally visualised individually in Fig. 5b-f. Thus, streets that required a turn

into, or that were main roads, were associated with slower response times (Fig. 5). Streets classified as boundary streets had a faster call out than other streets (Fig. 5d). There was also a small but consistent effect that the greater the number of streets to the destination the slower response times (Fig. 5a,e). This difference in response speed was most evident for initial streets and final streets (Fig. 5e), but not strongly evident when the full range of streets was examined (Fig. 5f). Euclidean distance did not contribute to the outcome of the model. Thus, it is not the distance ‘as the crow flies’ that will determine planning speed, but rather the number of streets to be travelled. For the 3398 responses, the number of boundaries is 923, the number of turns is 1858, and the



**Fig. 5.** Evidence for hierarchical planning revealed by faster responses on boundaries. Linear mixed model estimates (a) indicate a highly significant effect ( $* < 0.05$ ,  $*** < 0.001$ ) of turns, main roads, the number of streets to the goal and boundaries significantly impact on log-transformed response times. Euclidean distance does not reach significance, nor did circuitry when used in the model (model 3 see methods). (b-d) Differences in the recall of turns and non-turns (b), main roads and minor roads (c) and boundary streets and non-boundary streets (d) are plotted to allow visualisation of the response patterns. Additionally, the position of the street along the sequential recall highlights differences between initial and final streets (e). The difference between the first and last street is plotted for individual responses for routes containing 3 to 30 streets, revealing little evidence of slowing when the route planned is extensive (e.g. 30 streets) vs when it is short (e.g. 3 streets). (f) The difference between the first and last street is plotted for individual responses for routes containing 3 to 30 streets, revealing little evidence of slowing when the route planned is extensive (e.g. 30 streets) vs when it is short (e.g. 3 streets).



number of main roads is 2330. It therefore seems unlikely that there is an outsize effect from small numbers.

Next, we focused our analysis on the most consistent bounded region of London: Soho (Griesbauer, Manley, McNamee, et al., 2022). Model 2 revealed significant effects of *Soho* streets ( $b = -0.16, p < .05, 95\% \text{ CI} = -0.094 - -0.23$ ), *turns* ( $b = 0.14, p < .001, 95\% \text{ CI} = 0.17-0.11$ ) and *number of streets* to the destination ( $b = 0.0095, p < .05, 95\% \text{ CI} = 0.014-0.0052$ ). The effect of *main roads* was marginally significant ( $b = 0.089, p < .05, 95\% \text{ CI} = 0.95-0.029$ ) and *Euclidean distance* was again not significant ( $b = -0.037, p = .135, 95\% \text{ CI} = -0.012 - -0.061$ ). We found the response times of streets recalled within Soho were faster than for the rest of the environment. All other fixed effects variables indicated similar results, except for main roads, which only reach marginal significance. The second alternative, Model 3, examining *circularity* revealed that this variable was not significant ( $b = 0.014, p = .272, 95\% \text{ CI} = 0.026-0.001$ ), see Fig. S3. In this model there was again a significant effect of *boundaries* ( $b = -0.085, p < .05, 95\% \text{ CI} = -0.045 - -0.12$ ), *turns* ( $b = 0.14, p < .001, 95\% \text{ CI} = 0.17-0.11$ ), *number of streets* to the destination ( $b = 0.0061, p < .05, 95\% \text{ CI} = 0.0091-0.0032$ ), and *main roads* ( $b = 0.084, p < .05, 95\% \text{ CI} = 0.12-0.049$ ). Thus, from our extended models we show that planning across a set of streets in a well known bounded region is faster compared to other streets (matching predictions) and that routes that include higher circuitry were not slower to provide responses in planning (not matching predictions).

### 3.4. Streets with longer names did not have longer response times

We considered the possibility that it could take longer to recall the exact name of street names which are longer (e.g. 'Stoke Newington Church Street', is long, while 'Oak Row', which is much shorter). We found no relationship between the two normally distributed variables - short and long street names (see Fig. S5). A linear mixed model that additionally accounted for the number of characters as a fixed effect in the original model indicated no significant impact of the length of the street name ( $b = 0.0033, p = .410, 95\% \text{ CI} = -0.00071 - 0.0074$ ). All other fixed effects variables were in line with previous model results (*boundaries*:  $b = -0.082, p < .05, 95\% \text{ CI} = -0.12 - -0.042$ ; *turns*:  $b = 0.13, p < .001, 95\% \text{ CI} = 0.11-0.16$ ; *number of streets*:  $b = 0.011, p < .05, 95\% \text{ CI} = 0.0062-0.015$ ; *road type*:  $b = 0.087, p < .05, 95\% \text{ CI} = 0.052-0.12$ ; *Euclidean distance*:  $b = -0.0384, p = .122, 95\% \text{ CI} = -0.063 - -0.014$ ). Thus, the length of street names did not impact responses of the recall of individual streets.

## 4. Discussion

In this study, we asked licensed London taxi drivers to plan the route between a set of origin-destination pairs and recorded the time taken to state each street name (response time) they would use in the route. We tested the effect of spatially embedded street network variables on their response times using linear mixed effects models. Our results indicate a consistent pattern in modulations of response times in relation to spatial features, including the impact of hierarchical structures such as boundaries. We first discuss the impact of boundaries on response times and then consider the impacts of other variables, such as turns, circuitry and number of streets in the planning process.

In contrast to prior studies of hierarchical route planning using artificial environments, which relied on visually distinct boundaries between districts / clusters in very small networks (e.g. Wiener & Mallot, 2009; Wiener et al., 2004; Schick et al., 2019), our study used boundary agreement rates for London districts to reflect the gradual differences in boundaries, characteristic of real-world cities (Cohen et al., 1978; Filomena et al., 2019; Griesbauer, Manley, McNamee, et al., 2022; Manley, 2014). We found that taxi drivers were faster to select the next street when that street was a boundary rather than a non-boundary street. Our interpretation of these results is that faster response times correspond to the prioritization process in the pre-planning phase,

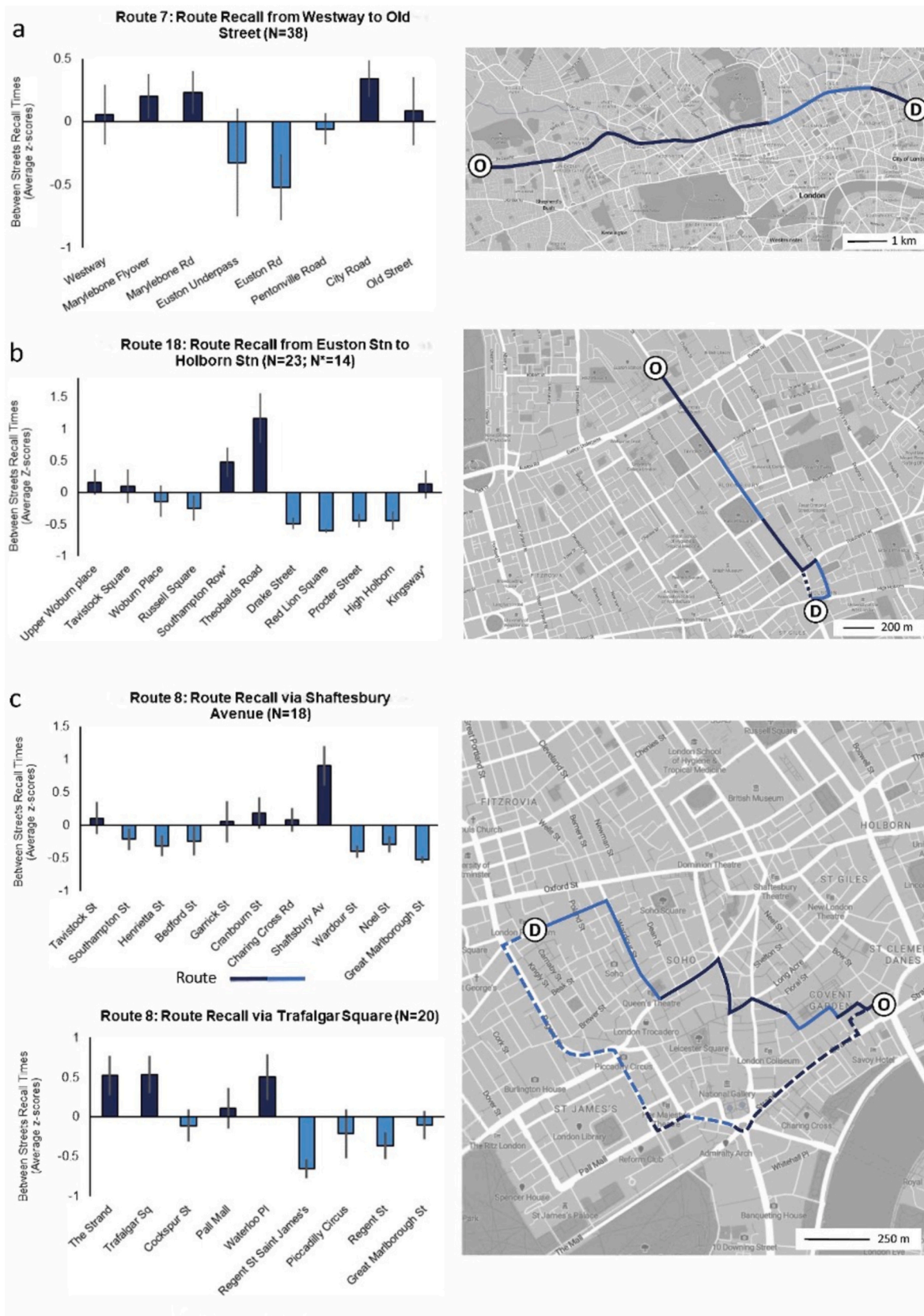
before the taxi drivers start to describe the streets in the itinerary step-by-step. In other words, the boundary streets are faster to be recalled because they have already been processed in this preplanning phase (McNamee et al., 2016). This is different than 'fine-to-coarse' planning (with fine spatial information for the close environment and coarse spatial information for distant locations), in which a participant might plan up to a boundary and then decide the next actions within a region (Spiers & Maguire, 2008; Wiener & Mallot, 2003). In fact, our findings run contrary to a previous study in which participants navigated fictional subway networks, where responses were slower at stations on the boundary between subway lines (Balaguer et al., 2016). A plausible explanation is that, in subways, boundaries and the number of options for next states are confounded, whereas in real-world city streets, boundaries can occur across sets of streets with no change in the number of path options.

There are some caveats regarding the generalization of these results to planning in non-expert populations. London taxi drivers learn the streets and regions of London in an idiosyncratic way, which might align with hierarchical planning processes (Griesbauer, Manley, Wiener, & Spiers, 2022). The best candidate for this potential training-performance connection is the method of spanning a string of cotton between a given origin-destination pair and then considering sub-goals along the route (Griesbauer, Manley, Wiener, & Spiers, 2022). This is specifically taught in London taxi driver training schools, and often such subgoals correspond to major roads, bridges, or boundary streets. Nevertheless, it is important to note that this is just one learning technique, and that taxi drivers also follow many learning techniques (e.g. mnemonics, mental simulation, method of loci...) that do not seem to rely on hierarchical segmentation (for a full list, see Griesbauer, Manley, Wiener, & Spiers, 2022). It is also important to highlight that the boundary rates employed in our study were obtained from a previous experiment with taxi drivers (Griesbauer, Manley, McNamee, et al., 2022) and that London taxi driver training schools do not explicitly teach this set of boundaries to the trainees.

While the impact of boundaries provides support for hierarchical planning, our results do not negate a tree-search plan being deployed. Akin to Balaguer et al. (2016) we also found the number of states to the goal (in our case streets) was positively associated with response time. A tree-search approach to planning predicts slower responses (more options to consider) for states with more states to the goal (Elliott & Lesk, 1982; Miller & Venditto, 2021; Streeter & Vitello, 1986). Thus, our study provides both evidence for hierarchical and tree search planning. However, the impact of the number of streets to the goal was small. When we examined the response times as a function of the number of streets on the route (3 to 30) the impact was very minimal (for a visualisation, see Fig. 5f). Routes requiring responses with 30 states to reach the goal had similar response time differences between the first and last street as routes with 3 streets. If the taxi drivers needed to do a tree-search over 30 streets to the goal, they would be very slow on the first steps and then get quicker, and if they needed to do a tree search over just 3 streets, there should be very little difference between start and end response times. We see no such effect, so it suggests there is minimal tree search happening when the taxi drivers are planning routes. This result strongly supports hierarchical planning as the dominant means by which London taxi drivers compress their representations of London to aid planning, but further work is required to clarify the connection between hierarchical and tree-search planning.

### 4.1. Impact of road network structure and turning on planning demands

Our finding that turning left or right caused slowing in response times is consistent with results from previous studies exploring the impact of turns on spatial memory (Brunec et al., 2020). This slowing is consistent with taxi drivers visualising the route as they progress, because visualising a turn would plausibly be more demanding than continuing straight ahead. A previous study in which taxi drivers gave



**Fig. 6.** Routes where taxi drivers chose very similar routes. Taxi drivers planned routes between origin (O) and destination (D) pairs. A selection of three routes with a high agreement on the route across drivers is displayed to highlight differences in z-standardised response times between streets. Agreement was found for linear routes (a, b) and routes with two alternatives (c). Corresponding streets with faster (light blue) and slower (dark blue) than average recall speed are highlighted on the right. Note that in (b) only nine drivers planned to do a loop past Red Lion Square, the other 14 taxi drivers planned to go straight from Southampton Row to Kingsway. Map source: Mapbox. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

post-task verbal reports of their thought process indicated that many taxi drivers, years after initial training, visualise the different streets to the goal as part of the planning process (Spiers & Maguire, 2008). Our investigation of the training process revealed that visualising the route was one of the key methods used to train taxi drivers in their spatial memory and learning of London (Griesbauer, Manley, Wiener, & Spiers, 2022). Notably, the response time for left and right turns is not trivially due to the additional delay that would be incurred by having to add the words “left into..” or “right into..” before street names; all responses required a statement about the direction, e.g. “forward into..” if there was no turn.

We found that taxi drivers were slower on average when a major road was the next step in the planned route. Major roads are associated with more long-distance connections than minor roads, as formalised by a higher betweenness centrality (for London, see Javadi et al., 2017). As such, this means the number of possible routes from such a street to other locations is substantially increased, driving a greater demand on planning. In contrast, an amnesic London taxi driver was able to navigate by such streets despite losing memory for minor roads (Maguire, Nannery, & Spiers, 2006), and Parisian taxi drivers have been previously reported to prioritise such streets for planning (Pailhous, 1969). We take it that there are two contrary effects at play here: major roads have more long-distance connections, but they are also more easily remembered. Generally speaking, one would expect major roads to be more memorable, as is reflected in the study of Parisian taxi drivers. There is a big difference, however, between London taxi drivers and taxi drivers in other cities such as Paris. In Paris (and in most other European capitals), taxi drivers are just required to learn the major roads of the street network. In London, taxi drivers are required to learn all streets across the city, not just major roads. This, we suggest, dampens the memorability of major roads for London taxi drivers, which, together with their higher betweenness centrality, corresponds to the slower response times associated for major roads that we find in our study.

The only spatial variables that had no significant impact on response times were Euclidean distance and circuitry (i.e. path distance to the goal divided by Euclidean distance to the goal). For many real-world routes, Euclidean distance and number of streets to the goal are correlated. Here we designed routes that decorrelated these, helping to reveal the specific effect of ‘streets to the goal’ from ‘Euclidean distance to the goal’. The absence of an effect of Euclidean distance suggests that, while taxi drivers might visualise turns into new streets, or have higher demands in rotating travel in a mental model, they do not consistently create mental images of views along the entire real route to the goal. Rather, it would be consistent with a sequence of mental ‘snapshots’ of street-to-street views being used for planning. This again mirrors the training of taxi drivers in which they learn a visual view for each street and exploit this for planning (Griesbauer, Manley, Wiener, & Spiers, 2022). This suggests that at least some elements of route planning are done using an egocentric frame of reference (Török et al., 2014). Indeed, when first asked about a route, taxi drivers tend to feel directions between the origin and the destination (e.g. “As soon as he said Guy’s Hospital ... I know Guy’s Hospital is there [subject points in the correct direction to Guy’s Hospital]”, in Spiers & Maguire, 2008). Here, the surprising absence of a circuitry effect (i.e. deviation of the path from a straight line) on planning reaction times suggests that it is the event of turning in a new direction that is demanding, and that once mentally simulating a route in a given direction, there is no specific mental cost to continuing in that direction, even though it may be away from the goal.

#### 4.2. Limitations

There are several limitations to our study. Firstly, we recruited drivers in the areas of Bloomsbury and King’s Cross, and they showed preferences for working in Central London, which could have impacted their knowledge of areas with greater distance to central London (e.g. West London, south of the River Thames). Future studies could recruit

taxi drivers at different taxi ranks across London to account for this possibility. Secondly, even though the study included several geo-spatial properties (e.g. distance measures, boundaries, road types, turns), there is a range of information that we did not account for in the current analysis. For instance, planning directions, angular deviations at each street, spatial analytics or perceptual input of building use were not included (even though angular deviation and circuitry are related metrics). These variables went beyond the scope of the study, but they should be considered in the future.

Potentially confounding factors, such as linguistic factors, age, or experience, were not found to impact response times. We considered whether the length of street names may interfere with response times during the verbal recall, but no such effect was found. For the entire group of taxi drivers, age and experience were correlated, and as age increased, so did experience. Age would be expected to impair route planning (van der Ham & Claessen, 2020), and experience should have the opposite effect (Chase, 1983; Pailhous, 1969). Moreover, training of spatial navigation abilities is expected to cause changes in hippocampal volume, the neural centre for spatial navigation: in a series of neuro-imaging studies, volume changes in the posterior hippocampal grey matter were correlated with experience, and with greater experience one might expect better performance (Maguire et al., 2000; Maguire, Woollett, & Spiers, 2006). However, the current data does not seem to point in this direction, as there were no correlations between experience and response times. We considered that for the entire group of taxi drivers, age and experience might have cancelled out any response time related effects. Even for a subset of drivers, where age and experience were decorrelated, no impact of age or experience was found. It is possible that the thorough training and daily use of navigation skills among London taxi drivers might protect them from age-related decline in spatial skills (Lövdén et al., 2012). Ultimately, this might explain why neither the entire group, nor the subset of taxi drivers showed any ageing effects for response times of route recalls.

While we controlled for word length, other linguistic factors, such as word complexity or familiarity of street names could have acted as confounds. Or taxi drivers might be aware of a street but struggle to recall it by name. Additional validation of findings could be achieved through alternative approaches with designs that do not rely on verbal recall of street names. For instance, video recorded route drawings on maps could provide supporting evidence for the current results. However, such a design would draw less on mental representations because visual features of the map (differently highlighted routes) are likely to impact route planning, given the availability of external spatial information. It is also not the manner in which taxi drivers are formally tested on competency. Therefore, study designs might have to restrict visual information to a small area around the streets that are being highlighted to ensure planning to rely on mental planning (e.g. showing only an area smaller than a quarter mile on a map at any time). In turn, implementing such a study would require technical or technological approaches that could come with other drawbacks regarding motor actions and a preference for paper maps over digitised maps with taxi drivers. Alternatively, drivers might be prompted to highlight key points that they would pass through on a map, or to indicate if specific streets would be part of their route. Under time pressure, these prompts might highlight response time differences for street network structures that are part of a hierarchical representation when contrasted with streets that are non-hierarchical. Here, planning would have to depend on predefined conditions, such as the most direct route rather than general preferences as in this study. However, such approaches would not allow for rich information on street level, focusing instead on coarse route level information, because attention is drawn to key points rather than entire routes. While each method comes with its own drawbacks, a diversity of approaches in future work might provide a useful contrast to the current study.

### 4.3. Future directions

In this study, we tested a total of 43 taxi drivers across two sets of data collection, and we analysed a total of almost 3400 response times between streets in relation to different street network properties. This data allowed us to identify interesting effects in the route planning process of licensed London taxi drivers, such as the impact of boundaries, turns or road type. There is room, however, for further spatial analysis of the environment under study. Space syntax, for instance, could provide vital insight into the role of major roads and street network properties in route planning (Hölscher et al., 2006; Javadi et al., 2017; Peponis et al., 1990; Yesiltepe et al., 2023). The current approach did not allow for this because most spatial network analytics use an approach that segregates entire streets into several clusters and then attributes spatial measures to each cluster, rather than to an entire street and their corresponding response times. Additionally, there is currently no approach to automatically assigning cluster-based information from the space syntax dataset of London to the street network, as clusters were parts of an unlabelled graph (i.e. no assignment of street names) and only contained graph network related information (e.g. betweenness centrality). Future analysis could consider a manual approach and focus on just a subset of routes or a specific area such as Soho.

Another question that remains open is how the spatial variables under study can be used to predict route planning at places in which multiple features interfere with each other. For instance, we would have predicted a fast recall for Shaftesbury Avenue in route 8, but the response time turned out to be extremely high. It appears that the properties that it was both a main road and a turning outweighed boundary factors in this particular case. Additionally, the intersection of Charing Cross Road with Shaftesbury Avenue might have turned it into a decision point, as both roads are main roads and several alternative options have to be considered, such as whether to cut through the north-eastern corner of Leicester Square. Thus, when forming predictions for a complex, real-world environment it will be useful to develop approaches to weigh many factors that might impact planning at specific places e.g. using judiciously chosen interaction variables for linear modelling, and this is an area that requires further research.

London taxi drivers are expert planners trained to navigate a particular urban environment. It is increasingly clear that there is a wide diversity of expert navigation strategies across the world (Farzanfar et al., 2023). It would be interesting to employ the methods developed here to investigate the factors that impact planning response times in other expert navigators (e.g. those navigating river system networks). In our study, we found no evidence of decline with age (and no gain with experience), which is remarkable, as navigation performance generally declines with age (Coutrot et al., 2018). Future work could study if this is the case in other navigation experts. Future work could also investigate more explicitly how navigation choices in the network related to metrics from reinforcement learning agents (De Cothi et al., 2022) or metrics about the space such as centrality measures (Lancia et al., 2023) and street network entropy (Coutrot et al., 2022).

Recently, there have been some important advances concerning the brain regions involved in hierarchical representations (Andermane et al., 2021; Peer & Epstein, 2021). Expert planning of the kind studied in this paper is likely to involve prefrontal-hippocampal interactions (Patai & Spiers, 2021; Simons & Spiers, 2003). Previous work has shown differences in hippocampal engagement in familiar environments when many vs few turns are required to reach a goal (Patai et al., 2019). Lateral prefrontal activity has been found to scale with the number of future options in streets when re-planning was required (Javadi et al., 2017) and with detour demands (Javadi et al., 2017). It seems likely for taxi drivers the prefrontal cortex would be engaged when it is difficult to exploit a hierarchy, such as instances of planning routes circuitous routes within a region where no hierarchy can be exploited. Future work could adapt our methodology for brain imaging studies in order to

clarify the neural structures involved in expert planning.

## 5. Conclusion

The current study provided real-world evidence of hierarchical route planning. The vast complexity of existing street networks, such as London, renders these environments computationally intractable for traditional tree-search models of situated route planning. Hierarchical models, which segregate the environment and represent it through smaller clusters containing local information (McNamee et al., 2016), are a promising approach, but it has been challenging to find real-world evidence for them. Previous work had employed virtual environments, which have hierarchically organised clusters by design, in contrast to urban networks, which often lack clearly defined boundaries. Here, we bridged this gap by relying on agreement rates across taxi drivers as an indicator for boundaries. We were then able to test the impact of boundaries and of other street network properties on route recall tasks performed by London taxi drivers. Our ecologically valid data shows that boundaries impact response times during route recall, as do other street network properties such as turning actions and road types. Taken together, our findings support the view that taxi drivers employ a hierarchical representation of London when planning routes.

### CRedit authorship contribution statement

**Eva-Maria Griesbauer:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pablo Fernandez Velasco:** Writing – review & editing, Writing – original draft, Visualization, Methodology. **Antoine Coutrot:** Methodology, Formal analysis, Data curation. **Jan M. Wiener:** Writing – review & editing, Investigation, Conceptualization. **Jeremy G. Morley:** Supervision, Investigation, Funding acquisition, Conceptualization. **Daniel McNamee:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Ed Manley:** Writing – review & editing, Formal analysis, Data curation, Conceptualization. **Hugo J. Spiers:** Writing – review & editing, Supervision, Project administration, Investigation, Conceptualization.

### Declaration of competing interest

The authors have no conflicts of interest to declare.

### Acknowledgements

We thank the Engineering and Physical Sciences Research Council UK (EPSRC) and Ordnance Survey Ltd. for funding our research. The British Academy supported the work of PFV (PFSS23\230053). We thank Iva Brunc for her help preparing figures.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2024.106014>.

### Data availability

Data is available as raw data in a supplementary file

### References

- Andermane, N., Joensen, B. H., & Horner, A. J. (2021). Forgetting across a hierarchy of episodic representations. *Current Opinion in Neurobiology*, 67, 50–57. <https://doi.org/10.1016/j.conb.2020.08.004>
- Arnold, A. E. G. F., Iaria, G., & Ekstrom, A. D. (2016). Mental simulation of routes during navigation involves adaptive temporal compression. *Cognition*, 157, 14–23. <https://doi.org/10.1016/j.cognition.2016.08.009>

- Balaguer, J., Spiers, H., Hassabis, D., & Summerfield, C. (2016). Neural mechanisms of hierarchical planning in a virtual Subway network. *Neuron*, 90(4), 893–903. <https://doi.org/10.1016/j.neuron.2016.03.037>
- Bast, H., Dellling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., ... Werneck, R. F. (2016). Route planning in transportation networks. In L. Kliemann, & P. Sanders (Eds.), *Algorithm engineering: Selected results and surveys* (pp. 19–80). Springer International Publishing. [https://doi.org/10.1007/978-3-319-49487-6\\_2](https://doi.org/10.1007/978-3-319-49487-6_2)
- Bonasia, K., Blommestein, J., & Moscovitch, M. (2016). Memory and navigation: Compression of space varies with route length and turns. *Hippocampus*, 26(1), 9–12. <https://doi.org/10.1002/hipo.22539>
- Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113(3), 262–280. <https://doi.org/10.1016/j.cognition.2008.08.011>
- Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46(10), 1730–1740. <https://doi.org/10.1016/j.tra.2012.07.005>
- Brown, T. I., & Chrastil, E. R. (2019). Editorial: Spatial navigation: Memory mechanisms and executive function interactions. *Frontiers in Human Neuroscience*, 13. <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00202>
- Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., ... Colton, S. (2012). A survey of Monte Carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1), 1–43. <https://doi.org/10.1109/TCAIG.2012.2186810>
- Brunec, I. K., Javadi, A.-H., Zisch, F. E., & Spiers, H. J. (2017). Contracted time and expanded space: The impact of circumnavigation on judgements of space and time. *Cognition*, 166, 425–432.
- Brunec, I. K., Ozubko, J. D., Ander, T., Guo, R., Moscovitch, M., & Barense, M. D. (2020). Turns during navigation act as boundaries that enhance spatial memory and expand time estimation. *Neuropsychologia*, 141, Article 107437. <https://doi.org/10.1016/j.neuropsychologia.2020.107437>
- Chase, W. G. (1983). Spatial representations of taxi drivers. In D. Rogers, & J. A. Sloboda (Eds.), *The Acquisition of Symbolic Skills* (pp. 391–405). Springer US. [https://doi.org/10.1007/978-1-4613-3724-9\\_43](https://doi.org/10.1007/978-1-4613-3724-9_43)
- Chrastil, E. R., Sherrill, K. R., Hasselmo, M. E., & Stern, C. E. (2015). There and back again: Hippocampus and retrosplenial cortex track homing distance during human path integration. *Journal of Neuroscience*, 35(46), 15442–15452.
- Cohen, R., Baldwin, L. M., & Sherman, R. C. (1978). Cognitive maps of a naturalistic setting. *Child Development*, 1216–1218.
- Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., ... Spiers, H. J. (2022). Entropy of city street networks linked to future spatial navigation ability. *Nature*, 604(7904), 104–110.
- Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., ... Hornberger, M. (2018). Global determinants of navigation ability. *Current Biology*, 28(17), 2861–2866.
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on Humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204–1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, 8(12), 1704–1711.
- De Cothi, W., Nyberg, N., Griesbauer, E.-M., Ghanamé, C., Zisch, F., Lefort, J. M., ... Spiers, H. J. (2022). Predictive maps in rats and humans for spatial navigation. *Current Biology*, 32(17), 3676–3689.e5. <https://doi.org/10.1016/j.cub.2022.06.090>
- Ekstrom, A. D., Spiers, H. J., Bohbot, V. D., & Rosenbaum, R. S. (2018). *Human spatial navigation*. Princeton University Press. <https://books.google.fr/books?hl=en&lr=&id=pwZKdWAAQBAJ&oi=fnd&pg=PP1&dq=Ekstrom+et+al.+2018+navigation&ots=e6j-zhAgBR&sig=dngjh0VicDvFYkXeMEGNyVvWVs>
- Elliott, R. J., & Lesk, M. E. (1982). Route finding in street maps by computers and people. *Proceedings of the Second AAAI Conference on Artificial Intelligence*, 258–261.
- Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, 20(11), Article 11. <https://doi.org/10.1038/nn.4656>
- Farzanfar, D., Spiers, H. J., Moscovitch, M., & Rosenbaum, R. S. (2023). From cognitive maps to spatial schemas. *Nature Reviews Neuroscience*, 24(2), Article 2. <https://doi.org/10.1038/s41583-022-00655-9>
- Fernandez-Velasco, P., & Spiers, H. J. (2024). Wayfinding across ocean and tundra: What traditional cultures teach us about navigation. *Trends in Cognitive Sciences*, 28(1), 56–71. <https://doi.org/10.1016/j.tics.2023.09.004>
- Filomena, G., Versteegen, J. A., & Manley, E. (2019). A computational approach to 'the image of the City'. *Cities*, 89, 14–25.
- Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245), 273–278. <https://doi.org/10.1126/science.aac6076>
- Griesbauer, E.-M., Manley, E., McNamee, D., Morley, J., & Spiers, H. (2022). What determines a boundary for navigating a complex street network: Evidence from London taxi drivers. *The Journal of Navigation*, 75(1), 15–34. <https://doi.org/10.1017/S0373463321000679>
- Griesbauer, E.-M., Manley, E., Wiener, J. M., & Spiers, H. J. (2022). London taxi drivers: A review of neurocognitive studies and an exploration of how they build their cognitive map of London. *Hippocampus*, 32(1), 3–20. <https://doi.org/10.1002/hipo.23395>
- van der Ham, I. J. M., & Claessen, M. H. G. (2020). How age relates to spatial navigation performance: Functional and methodological considerations. *Ageing Research Reviews*, 58, Article 101020. <https://doi.org/10.1016/j.arr.2020.101020>
- Hanson, B. A., & Seeger, C. J. (2016). *Creating Geospatial Data with geojson.io. Extension and Outreach Publications* (p. 128).
- Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well-worn route and the path less traveled: Distinct neural bases of route following and wayfinding in humans. *Neuron*, 37(5), 877–888.
- Hölscher, C., Meilinger, T., Vrachliotis, G., Brösamle, M., & Knauff, M. (2006). Up the down staircase: Wayfinding strategies in multi-level buildings. *Journal of Environmental Psychology*, 26(4), 284–299. <https://doi.org/10.1016/j.jenvp.2006.09.002>
- Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like processing of imagined navigation. *Current Biology*, 26(6), 842–847. <https://doi.org/10.1016/j.cub.2016.01.042>
- Hutcheson, A. T., & Wedell, D. H. (2009). Moderating the route angularity effect in a virtual environment: Support for a dual memory representation. *Memory & Cognition*, 37(4), 514–521. <https://doi.org/10.3758/MC.37.4.514>
- Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: How the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Computational Biology*, 8(3), Article e1002410. <https://doi.org/10.1371/journal.pcbi.1002410>
- Jafarpour, A., & Spiers, H. (2017). Familiarity expands space and contracts time. *Hippocampus*, 27(1), 12–16. <https://doi.org/10.1002/hipo.22672>
- Javadi, A.-H., Emo, B., Howard, L. R., Zisch, F. E., Yu, Y., Knight, R., ... Spiers, H. J. (2017). Hippocampal and prefrontal processing of network topology to simulate the future. *Nature Communications*, 8(1), Article 1. <https://doi.org/10.1038/ncomms14652>
- Kuipers, B., Tecuci, D. G., & Stankiewicz, B. J. (2003). The skeleton in the cognitive map: A computational and empirical exploration. *Environment and Behavior*, 35(1), 81–106. <https://doi.org/10.1177/0013916502238866>
- Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267.
- Lancia, G. L., Eluchans, M., D'Alessandro, M., Spiers, H. J., & Pezzulo, G. (2023). Humans account for cognitive costs when finding shortcuts: An information-theoretic analysis of navigation. *PLoS Computational Biology*, 19(1), Article e1010829. <https://doi.org/10.1371/journal.pcbi.1010829>
- Layer. (2018). OS MasterMap Integrated Transport Network (ITN). In *Coverage: London. Ordnance Survey, GB*. Using: EDINA Digimap Ordnance Survey Service.. Retrieved from: <https://digimap.edina.ac.uk>
- Lloyd, R. (1989). The estimation of distance and direction from cognitive maps. *The American Cartographer*, 16(2), 109–122. <https://doi.org/10.1559/152304089783875415>
- Lövdén, M., Schaefer, S., Noack, H., Bodammer, N. C., Kühn, S., Heinze, H.-J., ... Lindenberger, U. (2012). Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiology of Aging*, 33(3), 620.e9–620.e22. <https://doi.org/10.1016/j.neurobiolaging.2011.02.013>
- Lynch, K. (1964). *The image of the city*. MIT Press. <https://books.google.fr/books?hl=en&lr=&id=phRPWSpAgC&oi=fnd&pg=PA1&dq=Lynch+the+image+of+the+city&ots=Jy66b5Gok&sig=qoJCNEL6mEJk1eJkHRH6Zafuaw>
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97(8), 4398–4403. <https://doi.org/10.1073/pnas.070039597>
- Maguire, E. A., Nannery, R., & Spiers, H. J. (2006). Navigation around London by a taxi driver with bilateral hippocampal lesions. *Brain*, 129(11), 2894–2907. <https://doi.org/10.1002/hipo.20233>
- Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus*, 16(12), 1091–1101. <https://doi.org/10.1002/hipo.20233>
- Manley, E. (2014). Identifying functional urban regions within traffic flow. *Regional Studies, Regional Science*, 1(1), 40–42. <https://doi.org/10.1080/21681376.2014.891649>
- Mattar, M. G., & Lengyel, M. (2022). Planning in the brain. *Neuron*, 110(6), 914–934.
- McNamee, D., Wolpert, D. M., & Lengyel, M. (2016). Efficient state-space modularization for planning: Theory, behavioral and neural signatures. *Advances in Neural Information Processing Systems*, 29. <https://proceedings.neurips.cc/paper/2016/hash/10907813b97e249163587e6246612e21-Abstract.html>
- Miller, K. J., & Venditto, S. J. C. (2021). Multi-step planning in the brain. *Current Opinion in Behavioral Sciences*, 38, 29–39. <https://doi.org/10.1016/j.cobeha.2020.07.003>
- Muhle-Karbe, P. S., Sheahan, H., Pezzulo, G., Spiers, H. J., Chien, S., Schuck, N. W., & Summerfield, C. (2023). Goal-seeking compresses neural codes for space in the human hippocampus and orbitofrontal cortex. *Neuron*, 111(23), 3885–3899.
- Newcombe, N. S. (2018). Individual variation in human navigation. *Current Biology*, 28(17), R1004–R1008. <https://doi.org/10.1016/j.cub.2018.04.053>
- O'Doherty, J. P., Lee, S. W., & McNamee, D. (2015). The structure of reinforcement-learning mechanisms in the human brain. *Current Opinion in Behavioral Sciences*, 1, 94–100. <https://doi.org/10.1016/j.cobeha.2014.10.004>
- O'Neill, M. J. (1992). Effects of familiarity and plan complexity on wayfinding in simulated buildings. *Journal of Environmental Psychology*, 12(4), 319–327. [https://doi.org/10.1016/S0272-4944\(05\)80080-5](https://doi.org/10.1016/S0272-4944(05)80080-5)
- Pailhoux, J. (1969). Représentation De L'espace Urbain Et Cheminement. *Le Travail Humain*, 32(1/2), 87–139.
- Patai, E. Z., Javadi, A.-H., Ozubko, J. P., O'Callaghan, A., Ji, S., Robin, J., ... Moscovitch, M. (2019). Hippocampal and retrosplenial goal distance coding after long-term consolidation of a real-world environment. *Cerebral Cortex*, 29(6), 2748–2758.
- Patai, E. Z., & Spiers, H. J. (2021). The versatile wayfinder: Prefrontal contributions to spatial navigation. *Trends in Cognitive Sciences*, 25(6), 520–533.
- Peer, M., & Epstein, R. A. (2021). The human brain uses spatial schemas to represent segmented environments. *Current Biology*, 31(21), 4677–4688.e8. <https://doi.org/10.1016/j.cub.2021.08.012>

- Peponis, J., Zimring, C., & Choi, Y. K. (1990). Finding the building in wayfinding. *Environment and Behavior*, 22(5), 555–590. <https://doi.org/10.1177/0013916590225001>
- Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Computational Biology*, 13(9), Article e1005768. <https://doi.org/10.1371/journal.pcbi.1005768>
- Sadalla, E. K., & Staplin, L. J. (1980). The perception of traversed distance: Intersections. *Environment and Behavior*, 12(2), 167–182. <https://doi.org/10.1177/0013916580122003>
- Schick, W., Halfmann, M., Hardiess, G., Hamm, F., & Mallot, H. A. (2019). Language cues in the formation of hierarchical representations of space. *Spatial Cognition & Computation*, 19(3), 252–281. <https://doi.org/10.1080/13875868.2019.1576692>
- Shannon, C. E. (1950). Programming a computer for playing chess. In *First presented at the national IRE convention, march 9, 1949, and also in Claude Elwood Shannon collected papers* (pp. 637–656). <https://cir.nii.ac.jp/crid/1571417124017508096>.
- Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, 4(8), 637–648.
- Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., & Botvinick, M. M. (2014). Optimal Behavioral Hierarchy. *PLoS Computational Biology*, 10(8), Article e1003779. <https://doi.org/10.1371/journal.pcbi.1003779>
- Spiers, H. J., & Gilbert, S. J. (2015). Solving the detour problem in navigation: A model of prefrontal and hippocampal interactions. *Frontiers in Human Neuroscience*, 9. <https://www.frontiersin.org/articles/10.3389/fnhum.2015.00125>.
- Spiers, H. J., & Maguire, E. A. (2006a). Spontaneous mentalizing during an interactive real world task: An fMRI study. *Neuropsychologia*, 44(10), 1674–1682.
- Spiers, H. J., & Maguire, E. A. (2006b). Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage*, 31(4), 1826–1840.
- Spiers, H. J., & Maguire, E. A. (2007a). A navigational guidance system in the human brain. *Hippocampus*, 17(8), 618–626. <https://doi.org/10.1002/hipo.20298>
- Spiers, H. J., & Maguire, E. A. (2007b). Neural substrates of driving behaviour. *Neuroimage*, 36(1), 245–255.
- Spiers, H. J., & Maguire, E. A. (2007c). The neuroscience of remote spatial memory: A tale of two cities. *Neuroscience*, 149(1), 7–27.
- Spiers, H. J., & Maguire, E. A. (2008). The dynamic nature of cognition during wayfinding. *Journal of Environmental Psychology*, 28(3), 232–249.
- Streeter, L. A., & Vitello, D. (1986). A profile of Drivers' map-Reading abilities. *Human Factors*, 28(2), 223–239. <https://doi.org/10.1177/001872088602800210>
- Tomov, M. S., Yagati, S., Kumar, A., Yang, W., & Gershman, S. J. (2020). Discovery of hierarchical representations for efficient planning. *PLoS Computational Biology*, 16(4), Article e1007594. <https://doi.org/10.1371/journal.pcbi.1007594>
- Török, Á., Nguyen, T. P., Kolozsvári, O., Buchanan, R. J., & Nadasdy, Z. (2014). Reference frames in virtual spatial navigation are viewpoint dependent. *Frontiers in Human Neuroscience*, 8, 646.
- Venigalla, M., Zhou, X., & Zhu, S. (2017). Psychology of route choice in familiar networks: Minimizing turns and embracing signals. *Journal of Urban Planning and Development*, 143(2), 04016030. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000364](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000364)
- Weisberg, S. M., & Newcombe, N. S. (2016). How do (some) people make a cognitive map? Routes, places, and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(5), 768–785. <https://doi.org/10.1037/xlm0000200>
- Wiener, J. M., & Mallot, H. A. (2003). 'Fine-to-coarse' route planning and navigation in regionalized environments. *Spatial Cognition & Computation*, 3(4), 331–358. [https://doi.org/10.1207/s15427633scc0304\\_5](https://doi.org/10.1207/s15427633scc0304_5)
- Wiener, J. M., & Mallot, H. A. (2009). Fine-to-coarse' route planning and navigation in regionalized environments. *Spatial Cognition and Computation*, 3(4), 331–358.
- Wiener, J. M., Schnee, A., & Mallot, H. A. (2004). Use and interaction of navigation strategies in regionalized environments. *Journal of Environmental Psychology*, 24(4), 475–493. <https://doi.org/10.1016/j.jenvp.2004.09.006>
- Yesiltepe, D., Velasco, P. F., Coutrot, A., Torun, A. O., Wiener, J. M., Holscher, C., ... Spiers, H. J. (2023). Entropy and a sub-group of geometric measures of paths predict the navigability of an environment. *Cognition*, 236, Article 105443.