Do Spatial Abilities Have an Impact on Route Learning in Hypertexts?

Markus Kattenbeck[1(✉)], Thomas Jänich[1], and Ludwig Kreuzpointner[2]

[1] Chair for Information Science, University of Regensburg, 93053 Regensburg, Germany
markus.kattenbeck@ur.de, thomas@jaenich.de

[2] Chair for Medical Psychology, Psychological Diagnostics and Research Methodology, University of Regensburg, 93053 Regensburg, Germany
ludwig.kreuzpointner@ur.de

Abstract. Metaphors of navigation have been widely used to describe the behaviour of users surfing the World Wide Web. We present the results of a web-based experiment ($N = 85$ participants) on route learning in Wikipedia. As spatial abilities and sense of direction are known to be important for real-world wayfinding abilities, we examine the extent to which the participants are able to retrace a learned route on their own and the time taken to do this can be predicted using these variables. The tested (G)LMM models, however, show a lower than expected relevance of spatial abilities and sense of direction. The results suggest that both personal factors (such as age and gender) and task are important for the duration of tasks.

1 Introduction

Human spatial abilities and sense of direction are known to be of major importance for the navigational abilities humans show in real-world environments (see e.g. [9,20,65]). While this effect has been repeatedly analyzed, both in real-world (see e.g. [24]) and virtual environments (see e.g. [32] for a very recent study), it has been of less interest in hypertext environments. Metaphors of navigation have been used to describe user behaviour in large hypertexts from the beginning of the concept (see e.g. [16] for a very early account). This metaphor is useful (see [23] for a discussion) due to, in addition to other reasons, the pervasiveness of navigation in everyday life (see e.g. [18, pp. 68–69] and [35, p. 264] for a review of this argument). To date, however, little is known about how factors such as sense of direction and spatial abilities impact on navigational abilities on the web. This is of general interest because of the empirical evidence on users' preferences for navigation over search in digital filing systems (see e.g. [4]) and when trying to re-find mails (see e.g. [8]). Empirical findings suggest, moreover,

We would like to thank 85 persons willing to participate. We are grateful to David Elsweiler for his valuable comments to a draft version of this paper and to 4 anonymous reviewers for their feedback regarding possible improvements.

S. Creem-Regehr et al. (Eds.): Spatial Cognition 2018, LNAI 11034, pp. 211–227, 2018.
https://doi.org/10.1007/978-3-319-96385-3_15

that navigation in folder structures and physical environments use similar brain regions (see [6]). The work presented here, which is exploratory in nature, investigates this aspect by focusing on a specific task on the web: We analyze whether self-reported sense of direction and spatial mental abilities can be used to predict a person's ability to remember a sequence of links in Wikipedia. Choosing route learning is reasonable because it is commonly thought of as the first step of developing a cognitive map (see [49]). Using Wikipedia ensures ecological validity because it is a real-world hypertext system which has successfully been used in other navigation studies.

2 Related Work

Given the goal of this study, we examine two bodies of related work. The first deals with experiments on route learning under different conditions, whereas the second reviews studies dedicated to navigation involving Wikipedia.

Route learning has recently been studied from diverse perspectives. One aspect researchers are interested in is how body movement influences route-learning. Ruddle and colleagues [46,47] provide a strong argument for the importance of body-based information for both navigation tasks and the development of cognitive maps. Other studies examine the importance of gestures for memory recall (see [54,55]) and find evidences that gestures will increase recall across different levels of spatial abilities. A second aspect deals with the influence of age. O'Malley and colleagues [40] find differences with respects to route knowledge (older participants acquire less route knowledge). Lingwood et al. [32] compare the learning and recall capabilities of children of different ages with adults in a virtual environment. They find that even young children can learn routes very quickly in a non-repetitive manner. Similarly, Hartmeyer et al. [22] presents findings suggesting that older adults learn routes more slowly than younger adults. A third aspect of interest is the effect of different environments. Lloyd and colleagues [33] report on a pilot study regarding the equivalence of route learning in real and virtual reality environments. They find evidence indicating that results in terms of error rates and strategies applied by participants are very similar across different environments. This fits with the finding that the mode (treadmill w/o rotation vs. joystick) of conducting experiments on spatial learning has a major impact on landmark, route and survey knowledge acquisition [11,12]. More specifically, decision making fosters graph knowledge whereas idiothetic knowledge increases survey knowledge. Larrue et al. [29] draw the conclusion that "body-centred informations [sic] are more involved in allocentric (distance estimates) than egocentric navigational strategies". This aligns with the findings of van der Ham et al. [21] who compare abilities for learning routes in diverse real and virtual environments, concluding that a combination of walking a route, which is displayed on a mobile device, in an open field yields survey knowledge effects very close to real-world experiments while virtual environment experiments do not. The second body of research, relevant to our work, relates to Wikipedia, where interest in hypertext navigation has increased. A large proportion of the research uses a web-based game called Wikispeedia [61] to collect user data. In this game participants are required to find the shortest path between two random articles. The growing dataset collected by means of this online-game is used for several purposes. West et al. analyze differences from shortest paths [59], indicating a preference of participants to use hubs and extract missing links by analyzing the node centrality of pages and find evidence for the usefulness of these links by human subject ratings [60]. Takes and Kosters [57] try to understand why participants abort a specific navigation task and find evidence that users prefer landmark nodes, i.e. those with high in- and outdegrees in the network (see also [48]). Researchers, moreover, use the dataset to identify semantic relatedness of Wikipedia pages from human paths. Singer and colleagues [53] found evidence that, generally speaking, navigational paths are more useful to calculate semantic relations than given Wikipedia links are. Beyond the use of this particular dataset, methodological advancements have been made in recent years with respect to the comparison of human trails in web interaction in general [50,51] and have been successfully applied to Wikipedia: By matching user traces on Wikipedia with the users' goals – collected through 30,000 responses to a user survey – Singer et al. [52] apply these methods and find different user behaviour for different interests to be identifiable from server logs (e.g. explorers tend to have long sequences of pages at a considerably higher speed than others). Wikipedia server logs are also used to identify presumably useful links which do not yet exist in [41]. Lamprecht et al. [28] analyze eight different language versions of Wikipedia and find that navigability is heavily affected by restricted views on articles (i.e. lead section resp. the first paragraph vs. full article). This is in line with the findings of Dimitrov et al. [17], which indicate a high correlation between the position of a link and its successfulness (i.e. the more on top of the page and the more left on the screen the more successful). Moreover, they find that users prefer links with topical closeness to the current article and links pointing to the outer bounds of the network.

This overview of related work reveals while navigation in Wikipedia is a popular area of research, route learning and the effect of spatial abilities on navigation abilities – two aspects shown to be important in real world environments - have not yet been studied. This is of interest to our community because this kind of knowledge can help in the design of navigational aids for browsing behaviour based on personal factors. These can be part of a user profile, i.e. which is independent from the information repository currently browsed. It can, furthermore, have an impact on study designs, as large hypertexts are even more readily available than virtual environment setups.

3 Hypotheses

In their 2010 review article Wolbers and Hegarty [65, p. 141] provide a strong argument that real-world navigational abilities, which route learning is a part of, rely on spatial abilities and sense of direction. According to their model, spatial abilities can be measured by mental rotation ability, embedded figures and spatial memory span. Based on these insights about the omnipresence of navigation metaphors for user behaviour in hypertext and the fact that the influence of spatial abilities and sense of direction has not yet been analyzed, we derive the following two hypotheses for this exploratory study:

H1. The better an individual's spatial abilities and sense of direction the faster he or she refinds the memorized path.

H2. The poorer an individual's spatial abilities and sense of direction the more errors will be made during refinding.

4 Method

Data was collected by means of an online experiment. Participants were presented with routes (ordered sequences of Wikipedia pages reached by clicks on links) and then asked to refind these paths. We analyzed different facets of their ability to do this based on participant characteristics, i.e. dependent on their self-reported sense of direction and how they performed on spatial ability tests. Thus, this section has three parts. First, the routes in Wikipedia are described; second the way spatial abilities were assessed is presented, and, finally, the experimental setup is described. The experiments were conducted in German.

4.1 Determining Wikipedia Routes

We used the random article function of the German language Wikipedia to find four lemmas used as starting pages. For each starting lemma the destination page was found according to the following procedure. If the desired length of the path was not reached, a page was randomly selected from all content links the current Wikipedia page had. Table 1 presents the resulting random paths. Each of these paths comprises seven steps following the oft-cited 7 ± 2 rule for working memory capacity (see [34]). Path 1, however, comprises four steps only to familiarize participants with the task, i.e. this data was excluded from the analysis. By using random walks between start and destination lemma, we ensure that diverging interests in topics should not have an impact.

4.2 Measuring Spatial Abilities

Inspired by Wolbers and Hegarty [65, p. 141], we utilise measures of sense of direction, spatial memory and broad visual perception which include mental rotation and embedded figures tests:

Sense of direction. Participants were required to fill a self-report survey on their sense of direction. The scale presented by Münzer and Hölscher [37] comprises three different subscales: allocentric orientation strategy (7 questions), egocentric orientation (10), and cardinal direction orientation strategy (2). This survey is well-established, including norm data on a 4,000 participants sample published in 2016 [36,38].

Table 1. The paths participants had to learn by navigating Wikipedia.

	Path 1	Path 2	Path 3	Path 4
Start	Feiburger Beschwerdenliste	Kreis Samter	Subnetz	Kavadh I.
Step 1	Albert-Ludwigs-Universität Freiburg	Herzogtum Warschau	Supernetting	Belagerung
Step 2	Franz-Joseph-Bob	Erste Polnische Teilung	Netzklassen	Magister offociorum
Step 3	Volkswirtschaft	Weichsel	Netzmaske	Odoaker
Step 4	n/a	Nieszawa	Oktette	Noricum
Step 5	n/a	Estland	Digitaltechnik	Otto Helmut Urban
Step 6	n/a	Einheitssteuer	Datenbus	Hainburg
Destin.	n/a	Beitragsbe-messungsgrenze	Omnibus	Retz

Spatial memory. The spatial memory span of participants was assessed by a regular 4×4 field without any labeling variant of the corsi-block test [13], implemented in javascript. Participants have to repeat the order of colored highlighted fields starting with two fields. After a correct trail a new order with one more field was presented. After two mistakes the test ended. The measured value was number of field of the last correct trail. As Berch et al. [7, p. 330] pointed out, Corsi developed the block task as alternative to measure memory for verbal sequences, thereby noting "that the mental representations of verbal and spatial information in serial short-term memory are functionally equivalent".

Spatial abilities. Participants worked on a computer adapted version of three subtests of a general test of cognitive abilities (LPS-2, see [25]): No. 6: mental rotation, No. 7: visualization of 3D geometric bodies, No. 8: identify forms in line-pattern. These tests were constructed to represent *space* defined by Thurstone [58], which was the base for the *broad visual perception* of Carroll's Three-Stratum-Theory [10].

4.3 Experimental Setup

We developed a web application (see Fig. 2) as a Wikipedia-based data acquisition tool using Python 2.7 [1] and its framework called Django [2] to allow for a maximum of flexibility in the experimental design. The application captured the clicks on links and input of tests and questionnaires in a database. To restrict the number of potential participants, Internet Explorer, Chrome and Firefox were supported as browsers. The web application was not usable in mobile browsers in order to avoid confounding effects (e.g. differences in task time stemming from different rendering of Wikipedia pages on mobile devices).

A cookie ensured that an experiment was taken once and only once from within the same browser. All navigational meta elements (e.g. the navigation on the left-hand side and the search bar) were rendered non-clickable even though they were still part of the page. This means, category pages or description pages of pictures were not reachable to participants. Moreover, the change of color for clicked links from blue to purple was disabled. There was no way for the participants to leave Wikipedia from within the application.

Figure 1 shows the way experiments were conducted. Participants were first asked to provide their informed consent about the goals and the data analysis associated with this study. Next, the participants entered the learning phase of path 1, i.e. the start page was opened and participants were informed about the names of the start and destination page. On the instruction part of the page (see Fig. 2) participants were asked to remember the sequence of links they were guided. They were further instructed not to use CTRL+F or other shortcuts, but asked to find their way to the current decision point link (see Table 1) on their own. This had been repeated until they reached the destination page.

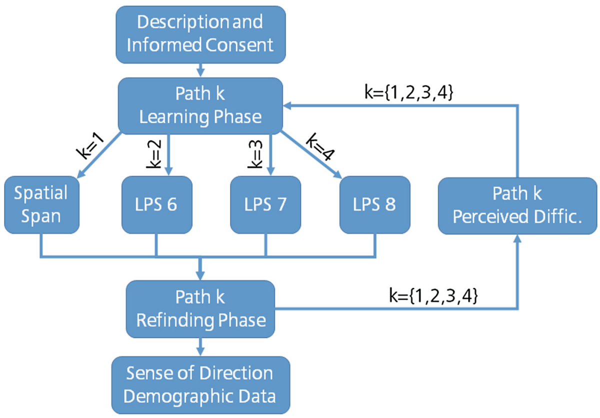


Fig. 1. A flowchart explaining the experimental setup. Every participant was subjected to $k = 4$ tasks, consisting of a route learning task, a test of spatial abilities, and a refinding phase. Having finished all tasks they were subject to a self-report sense of direction questionnaire and asked to provide demographic data.

Except for the first path, the paths were randomly assigned to each learning phase for every participant. Having finished a learning phase participants were asked to do one of the psychological tests (see above) to test their spatial abilities (see Fig. 1). Having finished the current test, participants entered the refinding phase. They were instructed to retrace the route they learned. During refinding the names of the start and the landing page were displayed on top of the page. In case of a mistake (i.e. when participants clicked on a wrong link) participants were instructed to return to the previous page via browser's back button. If participants did not reach the destination page within fifteen links, they were given the opportunity to stop the current task and continue with the remaining part of the experiment. On completion of each refinding phase, participants answered two questions, one about the perceived difficulty finding the links in the learning phase, the other about the difficulty of the refinding task. Having finished all tasks, participants were asked to provide demographic data as well as to fill in the sense of direction questionnaire (see [37]).

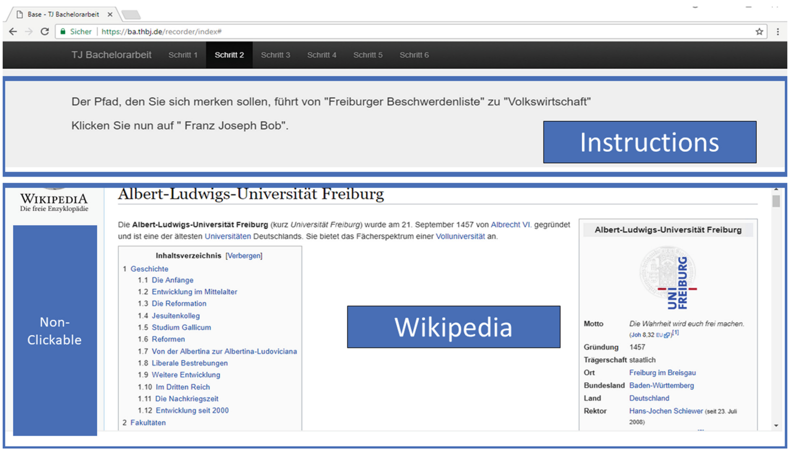


Fig. 2. A screenshot of the app used for data acquisition. The upper third of the application's website was used to provide instructions to the participants. The rest of the page contained the modified version of the Wikipedia.

4.4 Statistical Modeling

We use GNU R [42] and its packages dplyr [64], stringr [63], lavaan [45], psych [43], lme4 [5], lmerTest [26], xtable [14], ggplot2 [62] and texreg [31] to conduct our analysis. To explore if the Wikipedia-path-refind-paradigma is suitable for examination navigational abilities, we use duration (*dur*), i.e. the time needed to refind the right path, and the number of errors (*errors*) participants make as dependent variables. The independent variable we use to predict *dur* and *errors* are, first, the composite score of visual perception built with the three LPS-2 subtests (*LPS*); second, the composite overall score of the sense of direction questionnaire (*frs*), and third the spatial memory operationalized by the best level in the corsi task (*corsi*). To model the effects of spatial abilities, memory and sense of direction on working time we use a linear mixed model (LMM) in order to take individual working speed (*id*) and *task* specific attributes into account [19]. Additionally, we include the number of *errors* to control time consumption by mistakes. We also control for *age* and *gender* to take the ongoing discussion about different wayfinding strategies (see e.g. [22] resp. [30,56]) into account.

It would have been inappropriate to model *errors* with this approach due to the distribution of the error counts: 155 of 231 tasks (67%) were faultless. On average we find 1.6 errors ($SD = 3.5$, $skewness = 3.5$ and $kurtosis = 15.4$, see Fig. 3 and Table 2). To handle this floor effect, we modeled the probability of an error occurring as dependent variable by means of a logistic regression with random intercept within the Generalized Linear Mixed Model (GLMM) framework [15].

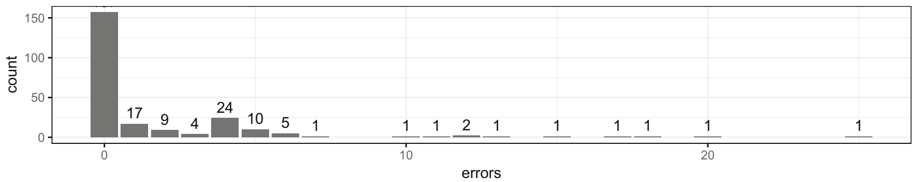


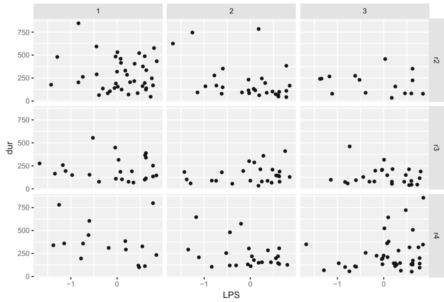
Fig. 3. Distribution of errors in all tasks (frequency of zero mistakes: 157)

5 Result

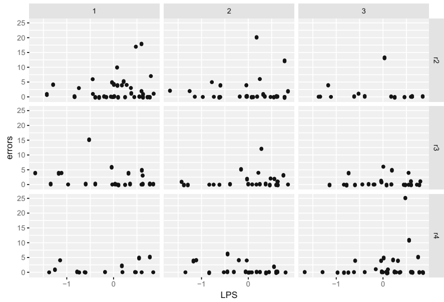
5.1 Descriptive Statistics

The data was collected from the August 25th to September 15th 2017. The application's disclaimer was viewed by approx. 300 different potential participants. 99 of the visitors completed the experiment. Of these, 13 participants were removed because they interrupted the experiment for more than 15 min. Furthermore, one record was not usable due to technical issues with storing the answers given to questionnaires. Therefore, 85 participants remain in the sample (46 female, 52 students, $\overline{x}_{age} = 27.34$ years, see Table 2). 78 out of 85 people participated from home, 6 from their office and one from a university building.

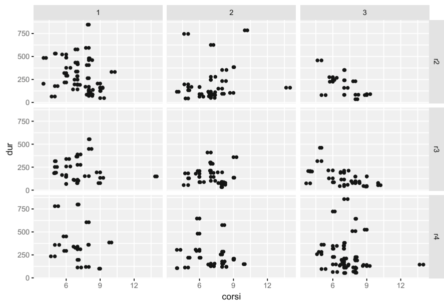
Due to the within groups design the 85 participants yield 255 trails. Of these, 18 trails were excluded from the analysis. 17 because the participants found a shorter path to the same target destination (1 for task 2, 11 trails for task 3, and 5 in task 4) and one trail because it took the participant longer than 40 min to finish. All statistical results reported below are based on the resulting $N_{trails} = 237$ usable trails. The figures presented in Table 2 reveal that task 3 is solved faster on average and with less variability compared to task 2 and 4 (see also Fig. 4). For all tasks the mean *duration* when solving it as first (position 1) is approximately 80 s longer than on position 2 and 3 ($M_{p1} = 280.37$ ($SD = 180.37$), $M_{p2} = 197.9$ ($SD = 156.85$), $M_{p3} = 199.89$ ($SD = 157.4$)). A 3×3 ANOVA shows main effects for *task* ($F(2, 228) = 8.22, p < .001$) and *position* ($F(2, 228) = 6.75, p < .01$), but not for their interaction ($F(4, 228) = 0.24, p = .92$). Neglecting the assumption for ANOVAs and apply it in the same way for *errors*, there are no effects for task or position (all F-values <1). The dependent variables are right-skewed, which is expected for time-based and



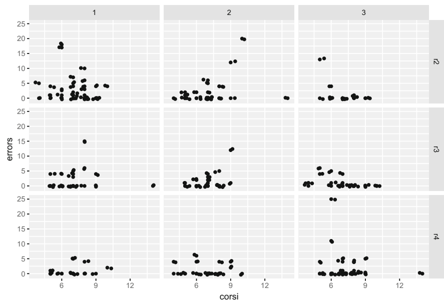
(a) LPS x duration



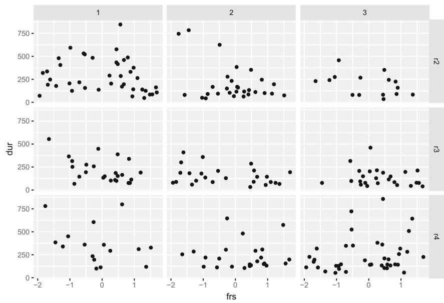
(b) LPS x errors



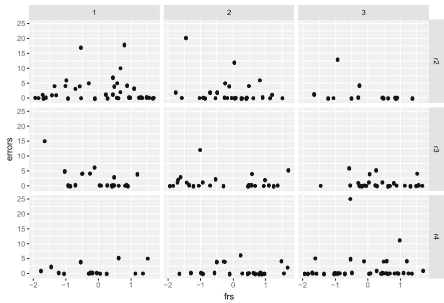
(c) corsi x duration



(d) corsi x errors



(e) frs x duration



(f) frs x errors

Fig. 4. Scatterplots of the independent variables against the time needed to find the path (*duration*) for each task (r2, r3, r4) at the particular position of the task (1, 2, 3).

error variables. Similarly, the *corsi* task results show a leptokurtic distribution because of high distribution around the mean of seven memorized steps and some outliers (see Fig. 4(c). The correlations between *duration* and number of *errors* for tasks 2 and 3 are rather high, whereas this is not the case for task 4. *frs* shows, furthermore, a similar pattern with lower coefficients. The remaining independent variables do not show any significant bivariate correlations neither with *duration* nor *error*.

Table 2. The correlations and descriptive statistics of the independent and dependent variables. Significant correlations ($\alpha = .05$) are bold-faced. *rX_dur* is the duration of refinding task X, *rX_err* the number of errors made during refinding for task X. *LPS* denotes the composite score of the three LPS-2 subtests indicating spatial abilities, *frs* the composite score of the sense of direction survey; *corsi* means the highest level achieved in the corsi-block-test. Level 0 of variable *gender* denotes females, 1 denotes males. Please note: The composite scores are centered.

	r2_dur	r3_dur	r4_dur	r2_err	r3_err	r4_err	LPS	frs	corsi	gender	age
r2_dur	1.00										
r3_dur	.11	1.00									
r4_dur	.13	.05	1.00								
r2_err	**.61**	.02	.04	1.00							
r3_err	.03	**.73**	−.01	.00	1.00						
r4_err	.14	.08	**.42**	**.30**	.09	1.00					
LPS	−.20	.00	−.09	.09	−.06	.06	1.00				
frs	**−.25**	**−.26**	.02	−.12	−.21	.03	**.23**	1.00			
corsi	−.05	−.21	−.15	.03	.01	−.09	.00	.11	1.00		
gender	−.10	−.02	−.20	.11	.04	−.15	.02	.14	.13	1.00	
age	.14	.21	.12	−.07	−.09	−.03	.17	.16	**−.30**	−.11	1.00
n	85	85	84	84	74	80	85	85	85	85	85
mean	241.07	164.03	260.28	2.11	1.36	1.31	0.00	0.00	7.12	0.46	27.34
sd	181.49	107.09	182.07	4.13	2.69	3.37	0.65	0.98	1.55	0.50	11.34
min	34.83	32.99	53.07	0.00	0.00	0.00	−1.67	−1.94	4.00	0.00	14.00
max	846.67	554.29	856.98	20.00	15.00	25.00	0.86	1.69	14.00	1.00	57.00
skew	1.30	1.39	1.46	2.64	2.85	4.73	−0.73	−0.25	0.77	0.16	1.59
kurtosis	1.23	1.68	1.63	7.01	9.78	28.41	−0.46	−1.00	3.08	−2.00	1.10

5.2 Modeling Errors and Duration – (G)LMM Results

Duration. Model 3 (see Table 3), i.e. including all variables, yields the best LMM fit with respect to Akaike's information criterion ($\Delta AIC_{Model1-Model3} = 92.39$) and the LR-test ($\Delta\chi^2_{Model1-Model3}(df = 7) = 106.39, p < .001$). Comparing the conditional R^2 (cR^2, i.e. the variance explained by both fixed and random factors [39]) reveals a significant increase when the number of *errors*, *gender* and *age* are used as control variables. The marginal R^2 (mR^2, variance explained by solely fixed factors) does not change for Model 3, but the coefficients of *LPS*, as well as *age* are now rendered significant. All coefficients, however, show the expected sign. The random effects *task* and *errors* are highly negatively correlated ($r = -.999$), a fact also evident from level 1 coefficients ($b_{r2} = -190.60 + 27.29 * errors$, $b_{r3} = -241.89 + 34.63 * errors$ and $b_{r4} = -140.26 + 20.08 * errors$).

Table 3. The fit of the LMM (model 1–3, left) to predict the time spent for a single task (*dur*) and the GLMM (model 4–6, right) results used to predict the probability of the occurrance of errors during a refinding task (*errordich*). Model 1 and 4 include only random factors (*id* and *task*); model 2 and 5 adds the fixed effects *LPS*, *frs* and *corsi*; model 3 includes *gender*, *age* and *errors* and model 6 just *gender* and *age*.

	Model 1	Model 2	Model 3
(Intercept)	225.57^{***}	308.53^{***}	330.52^{***}
	(23.44)	(56.53)	(55.97)
LPS		−22.77	-38.62^{**}
		(17.65)	(13.87)
frs		−21.50	−13.53
		(11.90)	(9.52)
corsi		−11.66	−3.95
		(7.24)	(5.92)
age			3.13^{***}
			(0.85)
gender			−33.27
(1=male, 0=female)			(17.88)
AIC	3103.20	3100.39	3010.81
BIC	3117.07	3124.67	3048.96
Log Likelihood	-1547.60	-1543.20	-1494.41
Num. obs.	237	237	237
Num. groups: id	84	84	84
Num. groups: task	3	3	3
Var: id (Intercept)	3257.24	2010.36	787.75
Var: task (Intercept)	1224.69	1251.98	38419.28
Var: Residual	24061.00	24106.22	15716.09
Var: task errors			787.37
Cov: task (Intercept) errors			-5500.00
marginal R^2		.04	.04
conditional R^2	.16	.16	.73

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

	Model 4	Model 5	Model 6
(Intercept)	-0.82^{***}	0.28	−0.08
	(0.00)	(0.88)	(1.17)
LPS		−0.14	−0.16
		(0.29)	(0.29)
frs		−0.27	−0.29
		(0.20)	(0.20)
corsi		−0.16	−0.14
		(0.12)	(0.13)
age			0.01
			(0.02)
gender			0.00
(1=male, 0=female)			(0.38)
AIC	303.59	305.03	308.81
BIC	314.00	325.84	336.56
Log Likelihood	-148.80	-146.51	-146.41
Num. obs.	237	237	237
Num. groups: id	84	84	84
Num. groups: task	3	3	3
Var: id (Intercept)	0.91	0.84	0.84
Var: task (Intercept)	0.01	0.01	0.01
marginal R^2		.08	.08
conditional R^2	.49	.50	.50

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

Errors. When modeling the probability of an *error* using GLMM, the fixed effects do not show any significant relevance. Taking just the random effects *task* and the personal factor into account (model 4, Table 3) results in the best fit looking at *AIC*. It shows, furthermore, a $\Delta\chi^2$ which is not statistically significantly worse than comparing the two other models χ^2 with more variables ($\Delta\chi^2_{Model1-Model2}(df = 3) = 4.65, p = .20$), but it explains just $cR^2 = .49$ variance. If fixed effects were, nevertheless, taken into account, they could explain $mR^2 = 8\%$ variance.

6 Discussion

A reasonable model fit for predicting *duration* can be achieved only by including the control variables age and gender ($cR^2 = .73$, model 3) – a finding in line with the controversy about the impact of gender and age mentioned above (see Sect. 4.4). In particular, considering the number of *errors* of each task shows an impact on *duration*. The variance of the *task* intercept increases heavily in model 3 compared to model 2 (see Table 3); similarly, with each error *duration* increases on average by 27, 34 or 20 s for task r2, r3 and r4, respectively. In model 3, however, *LPS* shows a statistically significant negative effect on *duration*, i.e. participants with lower abilities need more time. Spatial memory (*corsi*) has a lower impact for *duration* – presumably due to the floor effect. The weight of sense of direction (*frs*) shows a large standard error and is consequently not rendered statistically significant. The sign of the weight, however, is in line with prior expectations, i.e. better orientation skills yield shorter refinding time. *Men* seem to be a slightly faster than *women*, but high variability was again observed. Furthermore, *age* shows a statistically significant effect, with older participants taking on average 3 s longer for every year of age difference. These results are generally in line with the findings of studies in non-hypertext navigation contexts (see e.g. [22, 40]), which consistently report weaker route learning abilities in older persons. In the logistic GLMM (models 4 to 6), however, the impact of *LPS*, *frs* and *corsi* on error probability is much weaker. All independent variables show weights signed as expected (i.e. the higher the score the lower the *error probability*), but the large standard errors render them statistically insignificant. Based on the informational criterion (*AIC*) model 4 is to be preferred. Even in a model with no personal factors at all, however, neither of the independent variables shows a significant effect. Taken together, these results indicate that personal factors beyond the measured abilities have a major impact. The subjects' ability to memorize words may be one of those aspects (see [3, p. 833ff.] for pointers to this idea). To test whether the low predictive power of all models in this study is in general caused by a differential effect on the different tasks, we calculate a model where the fixed effects for each task are taken into account. Compared with model 3 no statistically significant increase in likelihood ($\Delta\chi^2(df = 25) = 4.04, p > .99$) is found. The mean task duration is different and there is an influence if participants do the task for the first time (it took longer), but there is no interaction. There must be traits beside those measured interacting with tasks attributes, e.g. topical interest in the Wikipedia article.

From a methodological perspective there are two main conclusions from these results: First, longer paths need to be used in order to avoid the floor effect by inducing greater variance with respect to the number of errors. While the average length of seven in the *corsi* block test reinforces the finding by Miller [34], a path length of seven steps does not yield the theoretically and statistically desirable spread of error counts. Second, participants should be made aware of the importance of speed in refinding – either via the instructions given or through a visualization on the refinding pages. This will help to reduce biases which might, for example, be introduced by topical interest, i.e. participants who are interested in a topic might start reading the Wikipedia article during the refinding phase but ought to reach the current link as quickly as possible.

7 Conclusion and Future Work

Based on the importance of sense of direction and spatial abilities for real-world wayfinding performance we analyze the importance of these factors for route learning performance in hypertexts. We report on an online user study based on Wikipedia to research this question and fitted several models to predict refinding duration and whether errors are made by subjects using a LMM and a GLMM approach, respectively. The analysis yields a weak fit for all models discussed. These results suggest that personal factors other than spatial abilities and sense of direction and, in particular, task related aspects play an important role. We have also seen that these results leave room for methodological improvements. There are four lines of future work we would like to explore. First, we want to assess the influence landmarks have on route learning in hypertexts. Existing evidence suggests that landmarks are of high importance in gaining location knowledge (see [44, pp. 41–108] for an overview of cognitive aspects that make landmarks important). Second, we plan experiments to further investigate the role of other personal traits (e.g. Big five traits, in particular conscientiousness, [27]) in both, route learning and path finding. Third, we will conduct experiments based on different page structures presented to users, thereby analyzing if e.g. location of contents has an effect on the usefulness of spatial abilities. Finally, we plan to compare real-world wayfinding abilities of participants with their abilities in hypertexts in further within-subjects design studies.

References

1. Python Language Reference, version 2.7 (2010). http://www.python.org. Accessed 18 Feb 2018
2. Django (Version 1.11) (2017). https://djangoproject.com. Accessed 18 Feb 2018
3. Baddeley, A.: Working memory: looking back and forward. Nat. Rev. Neurosci. **4**, 829–839 (2003)
4. Barreau, D., Nardi, B.A.: Finding and reminding: file organization from the desktop. ACM SIGCHI Bull. **27**, 39–43 (1995)
5. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. J. Stat. Softw. **67**(1), 1–48 (2015)
6. Benn, Y., Bergman, O., Glazer, L., Arent, P., Wilkinson, I.D., Varley, R., Whittaker, S.: Navigating through digital folders uses the same brain structures as real world navigation. Sci. Rep. **5**, 14719 (2015)
7. Berch, D.B., Krikorian, R., Huha, E.M.: The Corsi block-tapping task: methodological and theoretical considerations. Brain Cogn. **38**(3), 317–338 (1998)
8. Bergman, O., Gradovitch, N., Bar-Ilan, J., Beyth-Marom, R.: Folder versus tag preference in personal information management. J. Am. Soc. Inform. Sci. Techno. **64**(10), 1995–2012 (2013)
9. Burte, H., Montello, D.R.: How sense-of-direction and learning intentionality relate to spatial knowledge acquisition in the environment. Cogn. Res. **2**(1), 18 (2017)
10. Carroll, J.B.: The three-stratum theory of cognitive abilities. In: Flanagan, D.P., Genshaft, J.L., Harrison, P.L. (eds.) Contemporary Intellectual Assessment: Theories, Tests, and Issues, pp. 122–130. Guilford Press, New York (1997)

11. Chrastil, E.R., Warren, W.H.: Active and passive contributions to spatial learning. Psychon. Bull. Rev. **19**(1), 1–23 (2012)
12. Chrastil, E.R., Warren, W.H.: Active and passive spatial learning in human navigation: acquisition of graph knowledge. J. Exp. Psychol. Learn. Mem. Cogn. **41**(4), 1162–1178 (2015)
13. Corsi, P.M.: Human memory and the medial temporal region of the brain. Diss. Abstr. Int. **34**(2–B), 895 (1972)
14. Dahl, D.B.: xtable: Export Tables to LaTeX or HTML, r package version 1.8-2 (2016)
15. Demidenko, E.: Mixed Models: Theory and Applications with R. Wiley, Hoboken (2013)
16. Dillon, A., Richardson, J., McKnight, C.: Navigation in hypertext: a critical review of the concept. In: Diaper, D., Gilmore, D., Cockton, G., Shackel, B. (eds.) Human-Computer Interaction-INTERACT 1990, pp. 587–592. North Holland, Amsterdam (1990)
17. Dimitrov, D., Singer, P., Lemmerich, F., Strohmaier, M.: What makes a link successful on Wikipedia? In: Proceedings of the 26th International Conference on World Wide Web, WWW 2017, pp. 917–926 (2017)
18. Fabrikant, S.I.: Spatialized browsing in large data archives. Trans. GIS **4**(1), 65–78 (2000)
19. Gałecki, A., Burzykowski, T.: Linear Mixed-Effects Models Using R: A Step-by-Step Approach. Springer, New York (2013)
20. van der Ham, I.J.M., Claessen, M.H.: Navigation ability. In: Neuropsychology of Space. Spatial Functions of the Human Brain, Chap. 8, pp. 267–308. Academic Press, Amsterdam (2017)
21. van der Ham, I.J.M., Faber, A.M.E., Venselaar, M., van Kreveld, M.J., Löffler, M.: Ecological validity of virtual environments to assess human navigation ability. Front. Psychol. **6**, 637 (2015)
22. Hartmeyer, S., Grzeschik, R., Wolbers, T., Wiener, J.M.: The effects of attentional engagement on route learning performance in a virtual environment: an aging study. Front. Aging Neurosci. **9**, 235 (2017)
23. Hochmair, H.H., Luttich, K.: An analysis of the navigation metaphor - and why it works for the world wide web. Spat. Cogn. Comput. **6**(3), 235–278 (2006)
24. Hölscher, C., Meilinger, T., Vrachliotis, G., Brösamle, M., Knauff, M.: Finding the way inside: linking architectural design analysis and cognitive processes. In: Freksa, C., Knauff, M., Krieg-Brückner, B., Nebel, B., Barkowsky, T. (eds.) Spatial Cognition 2004. LNCS (LNAI), vol. 3343, pp. 1–23. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-32255-9_1
25. Kreuzpointner, L., Lukesch, H., Horn, W.: Leistungsprüfsystem 2. LPS-2. Manual. Hogrefe, Göttingen (2013). http://epub.uni-regensburg.de/27958
26. Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B.: lmerTest package: tests in linear mixed effects models. J. Stat. Softw. **82**(13), 1–26 (2017)
27. Kyritsis, M., Blathras, G., Gulliver, S., Varela, V.A.: Sense of direction and conscientiousness as predictors of performance in the Euclidean travelling salesman problem. Heliyon **3**(11), e00461 (2017). http://www.sciencedirect.com/science/article/pii/S2405844017306072
28. Lamprecht, D., Dimitrov, D., Helic, D., Strohmaier, M.: Evaluating and improving navigability of Wikipedia: a comparative study of eight language editions. In: Proceedings of the 12th International Symposium on Open Collaboration, OpenSym 2016, pp. 17:1–17:10. ACM, New York (2016)

29. Larrue, F., Sauzeon, H., Wallet, G., Foloppe, D., Cazalets, J.R., Gross, C., N'Kaoua, B.: Influence of body-centered information on the transfer of spatial learning from a virtual to a real environment. J. Cogn. Psychol. **26**(8), 906–918 (2014)
30. Lawton, C.A.: Strategies for indoor wayfinding: the role of orientation. J. Environ. Psychol. **16**(2), 137–145 (1996)
31. Leifeld, P.: texreg: conversion of statistical model output in R to LaTeX and HTML tables. J. Stat. Soft. **55**(8), 1–24 (2013)
32. Lingwood, J., Blades, M., Farran, E.K., Courbois, Y., Matthews, D.: Using virtual environments to investigate wayfinding in 8- to 12-year-olds and adults. J. Exp. Child Psychol. **166**, 178–189 (2018)
33. Lloyd, J., Persaud, N.V., Powell, T.E.: Equivalence of real-world and virtual-reality route learning: a pilot study. Cyberpsychol. Behav. **12**(4), 423–427 (2009)
34. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **101**(2), 343–352 (1994)
35. Montello, D.R.: Navigation. Cambridge Handbooks in Psychology, pp. 257–294. Cambridge University Press, Cambridge (2005)
36. Münzer, S., Fehringer, B.C., Kühl, T.: Validation of a 3-factor structure of spatial strategies and relations to possession and usage of navigational aids. J. Environ. Psychol. **47**, 66–78 (2016)
37. Münzer, S., Hölscher, C.: Entwicklung und Validierung eines Fragebogens zu räumlichen Strategien. Diagnostica **57**, 111–125 (2011)
38. Münzer, S., Hölscher, C.: Standardized norm data for three self-report scales on egocentric and allocentric environmental spatial strategies. Data Brief **8**, 803–811 (2016)
39. Nakagawa, S., Schielzeth, H.: A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol. Evol. **4**(2), 133–142 (2013)
40. O'Malley, M., Innes, A., Wiener, J.M.: How do we get there? Effects of cognitive aging on route memory. Mem. Cogn. **46**(2), 274–284 (2018)
41. Paranjape, A., West, R., Zia, L., Leskovec, J.: Improving website hyperlink structure using server logs. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, WSDM 2016, pp. 615–624. ACM, New York (2016)
42. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017). https://www.R-project.org/
43. Revelle, W.: psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois, r package version 1.7.8 (2017)
44. Richter, K.F., Winter, S.: Landmarks: Giscience for Intelligent Services. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-05732-3
45. Rosseel, Y.: lavaan: an R package for structural equation modeling. J. Stat. Softw. **48**(2), 1–36 (2012). http://www.jstatsoft.org/v48/i02/
46. Ruddle, R.A., Lessels, S.: The benefits of using a walking interface to navigate virtual environments. ACM Trans. Comput.-Hum. Interact. **16**(1), 5:1–5:18 (2009)
47. Ruddle, R.A., Volkova, E., Bülthoff, H.H.: Walking improves your cognitive map in environments that are large-scale and large in extent. ACM Trans. Comput.-Hum. Interact. **18**(2), 10:1–10:20 (2011)

48. Scaria, A.T., Philip, R.M., West, R., Leskovec, J.: The last click: why users give up information network navigation. In: Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM 2014, pp. 213–222. ACM, New York (2014)
49. Siegel, A.W., White, S.H.: The development of spatial representations of large-scale environments. In: Reese, H. (ed.) Advances in Child Development and Behaviour, pp. 9–55. Academic Press, New York (1975)
50. Singer, P., Helic, D., Hotho, A., Strohmaier, M.: HypTrails: a Bayesian approach for comparing hypotheses about human trails on the web. In: Proceedings of the 24th International Conference on World Wide Web, WWW 2015, pp. 1003–1013. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2015)
51. Singer, P., Helic, D., Hotho, A., Strohmaier, M.: A Bayesian method for comparing hypotheses about human trails. ACM Trans. Web **11**(3), 14:1–14:29 (2017)
52. Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M., Leskovec, J.: Why we read Wikipedia. In: Proceedings of the 26th International Conference on World Wide Web, WWW 2017, pp. 1591–1600 (2017)
53. Singer, P., Niebler, T., Strohmaier, M., Hotho, A.: Computing semantic relatedness from human navigational paths: a case study on Wikipedia. Int. J. Semant. Web Inf. Syst. **9**(4), 41–70 (2013)
54. So, W.C., Ching, T.H., Lim, P.E., Cheng, X., Ip, K.Y.: Producing gestures facilitates route learning. PLoS One **9**(11), e112543 (2014)
55. So, W.C., Shum, P.L., Wong, M.K.: Gesture is more effective than spatial language in encoding spatial information. Q. J. Exp. Psychol. **68**(12), 2384–2401 (2015)
56. Suzer, O.K., Olgunturk, N., Guvenc, D.: The effects of correlated colour temperature on wayfinding: a study in a virtual airport environment. Displays **51**, 9–19 (2018)
57. Takes, F.W., Kosters, W.A.: Mining user-generated path traversal patterns in an information network. In: Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), WI-IAT 2013, vol. 01, pp. 284–289. IEEE Computer Society, Washington, D.C. (2013)
58. Thurstone, L.L.: Primary Mental Abilities. University of Chicago Press, Chicago (1938)
59. West, R., Leskovec, J.: Human wayfinding in information networks. In: Proceedings of the 21st International Conference on World Wide Web, WWW 2012, pp. 619–628. ACM, New York (2012)
60. West, R., Paranjape, A., Leskovec, J.: Mining missing hyperlinks from human navigation traces: a case study of Wikipedia. In: Proceedings of the 24th International Conference on World Wide Web, WWW 2015, pp. 1242–1252. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2015)
61. West, R., Pineau, J., Precup, D.: Wikispeedia: an online game for inferring semantic distances between concepts. In: International Joint Conference on Artificial Intelligence (IJCAI 2009), Pasadena, California, pp. 1598–1603 (2009)
62. Wickham, H.: ggplot2: Elegant Graphics for Data Analysis. Springer, New York (2009). https://doi.org/10.1007/978-0-387-98141-3

63. Wickham, H.: stringr: simple, consistent wrappers for common string operations, r package version 1.2.0 (2017). https://CRAN.R-project.org/package=stringr
64. Wickham, H., Francois, R., Henry, L., Müller, K.: dplyr: a grammar of data manipulation, r package version 0.7.4 (2017). https://CRAN.R-project.org/package=dplyr
65. Wolbers, T., Hegarty, M.: What determines our navigational abilities? Trends Cogn. Sci. **14**(3), 138–146 (2010)