Supplemental Information on **Egocentric Network Exploration**

This supplemental material contains detailed information about the study procedure, the analysis process, and all results.

Study Procedure

Printed Material

On the following pages, we show the printed instructions users received before the study in the following order:

1. Task Description
2. System Description
3. Condition Description (Fly-Through, Jump-Through, Jump-Through Bubble[1] – before the corresponding condition, depending on the order of appearance)

[1] These were the names of the conditions presented to the participants during the experiment. In the manuscript, they were renamed to Baseline (Fly-Through), Ego-Highlight (Jump-Through), and Ego-Bubble (Jump-Through Bubble).

Task Description

In this study, we will ask you to perform multiple network analytics tasks in virtual reality. Networks (or graphs) show interconnections between a set of entities (= nodes). In our study, we will show nodes as spheres with short text labels, and interconnections as tubes / lines connecting them. Network analytics is the study of network entities and their connections. The networks used in this study were automatically generated and do not have any meaning. In this study, we compare different ways of showing networks in virtual reality, where it is possible to “immerse” into the graph.

You will have to perform a set of short tasks, such as finding a node with a given text label. Each task will be explained directly before it starts. Please, read the instructions carefully! If you do not understand the instructions or are not sure, please ask immediately. Once the instructions are understood, tell the examiner to start the task. Then we will measure the time and the correctness of your actions.

The tasks will be repeated six times:

using 3 different network interfaces and

using first a small and then a large network for each interface.

Before each network interface, you will perform a short training to get familiar with the visualization and the interaction controls. We will not obtain any performance measures during this training period. If you have any questions, please ask during the training and complete the actual study tasks as rapidly and correctly as possible.

After all three network interfaces, you will be able to take a break. We will also ask you to fill out a questionnaire in this break. After you completed all three network interfaces, we will also conduct a short interview, where you can report your experiences, difficulties, and suggestions for future improvements.

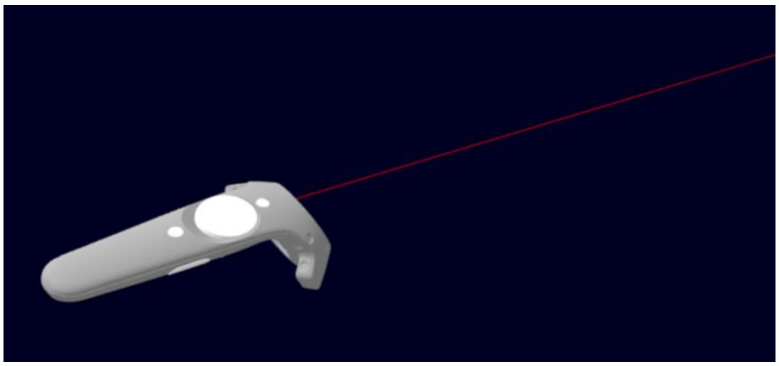
In total, the study lasts around one hour. If you feel discomfort, you are free to leave the study at any time. Upon successful completion of the study, you will receive a small monetary compensation for your time.

Thank you in advance for your participation!

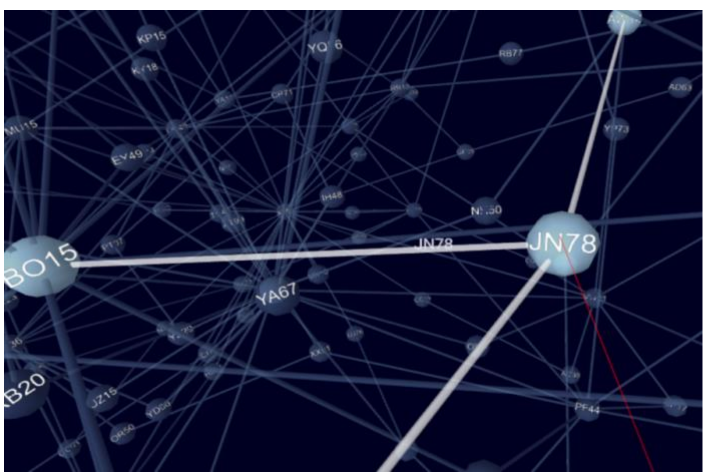
System Description

For this study, we use an HTC VIVE Virtual Reality head set. You can look around in the virtual world by moving your head and body.

For pointing and navigating, you will receive a controller to hold in your preferred hand. You will also see these controllers in the virtual world.



The controller **shoots a ray** into the scene (similar to a laser pointer). You can point to a node to see its direct connections.



For some tasks, we will ask you to **trigger a selection**. Do this by pointing the ray into the direction you want to point and pulling the trigger on the bottom of the controller with your index finger.

Before each of the three network interfaces, you will have a short training period to get to know these interaction techniques. Please, take your time to practice and ask for advice if you are unsure!

Fly-Through (Baseline)

In this network interface, you can freely fly through the network. Do this by controlling the round pad on the top of the controller using your thumb. You will fly relative to the direction in which you are looking with the headset (left/right, forward/backward)

Jump-Through (Ego-Highlight)

In this network interface, you can "jump" through the network from node to node. Do this by pointing to the node you want to select using the controller and trigger the jump using the trigger button on the controller with your index finger. Once you triggered the jump, your position will be transferred to the selected node in a short animation.

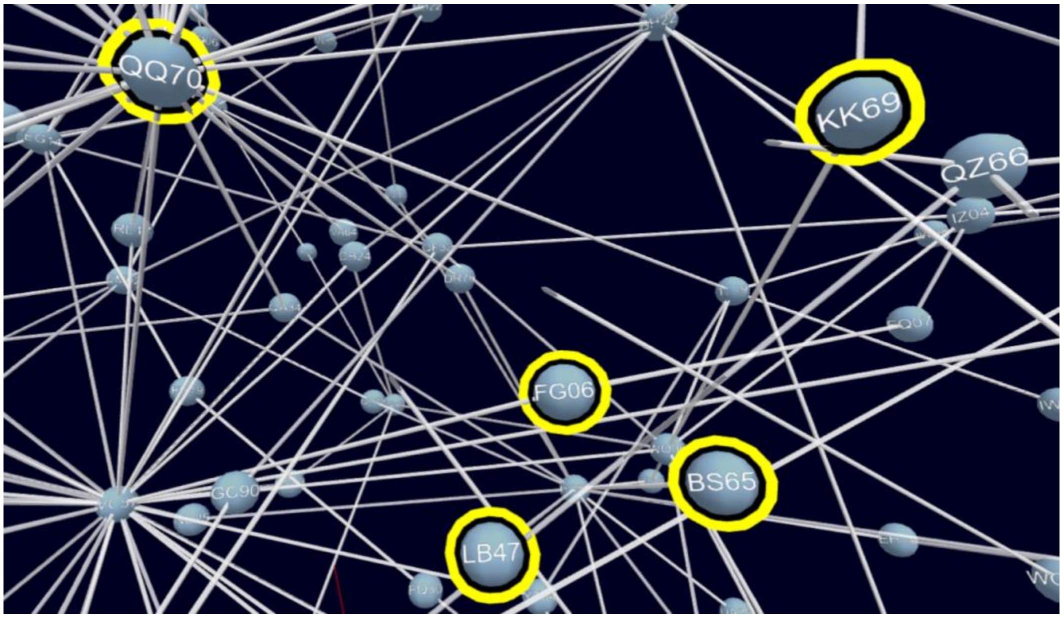
As your position is within a node, we do not show the connections from the node you are currently located at. Instead, we visualize the directly connected nodes using yellow circles around them.

Note that you can jump to any node, not just the directly connected ones.

Jump-Through Bubble (Ego-Bubble)

In this network interface, you can "jump" through the network from node to node. Do this by pointing to the node you want to select using the controller and trigger the jump using the trigger button on the controller with your index finger. Once you triggered the jump, your position will be transferred to the selected node in a short animation.

As your position is within a node, we create a virtual "bubble" around your currently selected node. Within this bubble, there are no other nodes and connections. Directly connected nodes lie on the surface of this virtual bubble and have yellow circles around them.



Note that you can jump to any node, not just the directly connected ones.

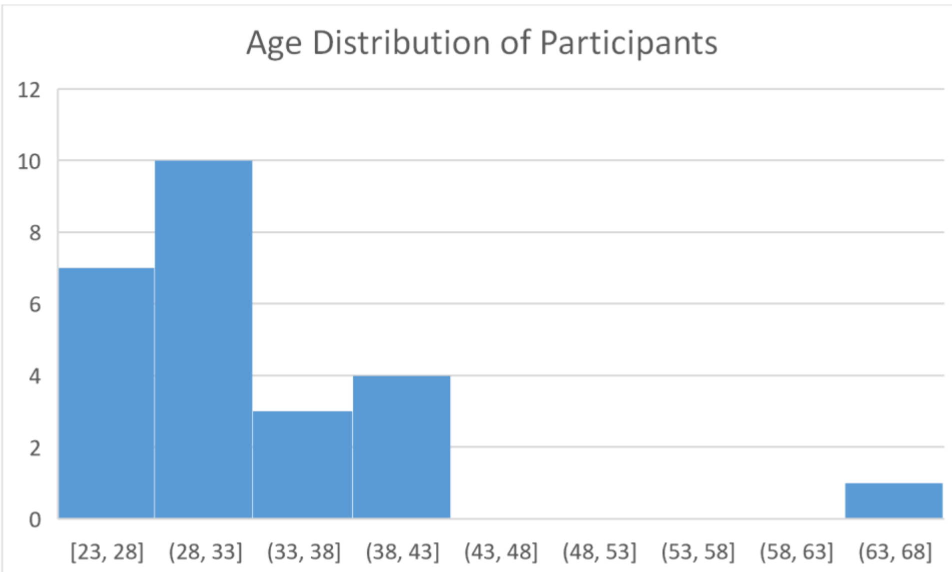
Graeko-Latin Square

The following three settings were used and repeated after every third user:

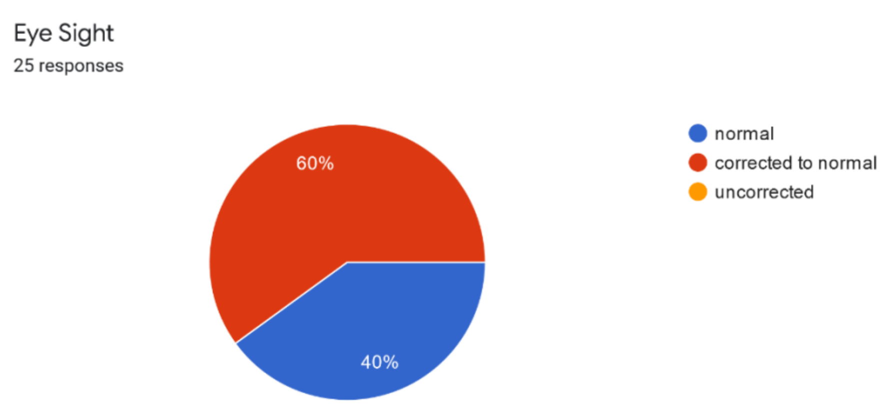
Set	Interface 1	Data Set 1	Interface 2	Data Set 2	Interface 3	Data Set 3
1	Baseline	α	Ego-Highlight	γ	Ego-Bubble	β
2	Ego-Highlight	β	Ego-Bubble	α	Baseline	γ
3	Ego-Bubble	γ	Baseline	β	Ego-Highlight	α

Participants

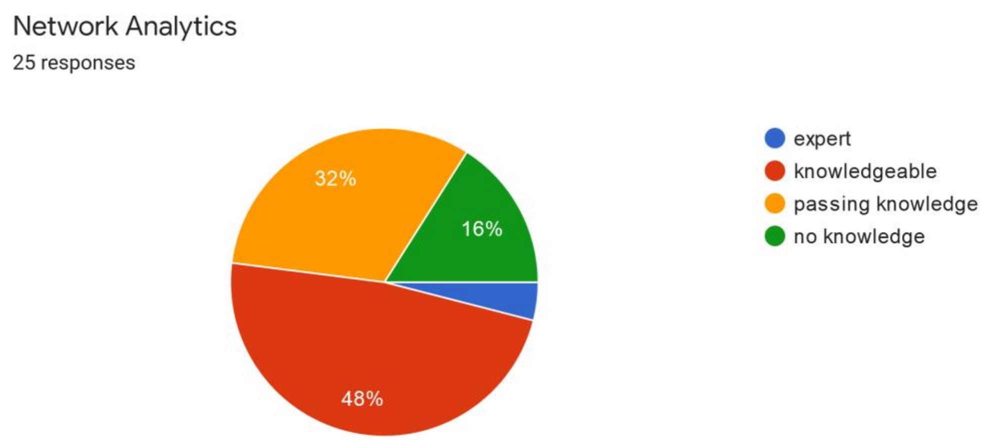
In total, 25 participants from a local university and a research institution participated. The age distribution by the participants looks as follows:



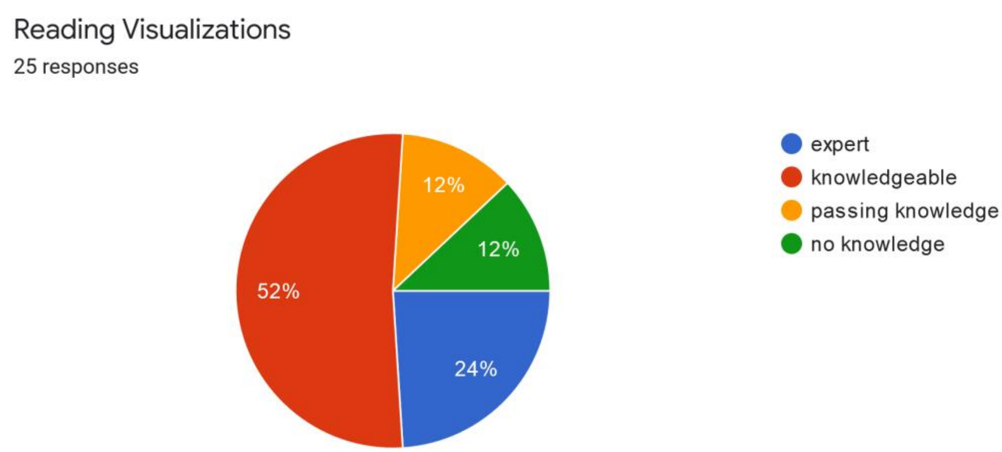
All users had corrected or corrected to normal vision:



Most users had some prior experience with network analytics:

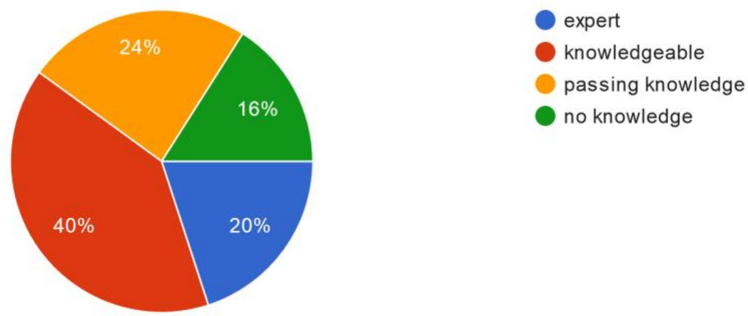


Also, three thirds of the users stated that they are knowledgeable or experts in reading visualizations:



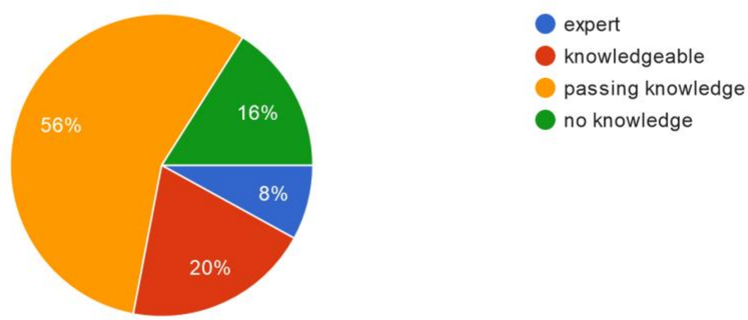
More than half of the users have experience with computer games:

Computer Games
25 responses



More than three quarters of the participants have had some VR experiences, but only around a quarter considers themselves as knowledgeable or expert:

Virtual Reality
25 responses



18 users reported their prior VR experiences:

- Played some beatsaber
- Played a couple of VR Scenarios
- I have tested a few different VR devices, and participated in VR user studies, but I do not have many hours in VR environments.
- none
- Games (Super Hot).
- played one or two vr games
- working in VR research
- CG expert
- Tried it mabe 2-3 times
- Fun for games
- I tried it once to watch an underwater scene

- VR development, VR games
- good so far
- PS4-Games
- almost none, some testing of VR networks
- Rare user.
- I have played VR Games
- Games

Questionnaires

We used two standard questionnaires for the analysis: the Simulator Sickness Questionnaire (SSQ) by Kennedy et al., 1993, and NASA Task Load Index (TLX) by Hart and Staveland, 1988.

For the SSQ, we computed the three categories nausea, oculomotor, and disorientation as described by Kennedy et al., 1993. We compute the three categories as follows:

- Nausea: sum (general discomfort, burping, stomach awareness, sweating, salivation increase, nausea, difficulty concentrating) x 9.54
- Oculomotor: sum (blurred vision, difficulty focusing, eye strain, headache, fatigue, difficulty concentrating, general discomfort) x 7.58
- Disorientation: sum (vertigo, dizziness with eyes closed, dizziness with eyes open, fullness of head, difficulty focusing, nausea, blurred vision) x 13.92

For the TLX, we computed an aggregate TLX as described by Rubio et al., in Applied Psychology, 2003: in all 15 pair-wise comparisons between the individual scores, we picked the more relevant one. The number of times a score was selected defines its weight. The following weights were obtained:

- Mental demand (MD): 5
- Physical demand (PD): 1
- Temporal demand (TD): 0
- Performance (P): 2
- Effort (E): 5
- Frustration (F): 2

The overall task load (TL) score was then computed as follows:

$$TL = \frac{5\,.\,MD + PD + 2.\,(8 - P) + 5.\,E + 2.\,F}{15}$$

As we issued the questionnaire using Google Forms, we were limited to a Likert Scale up to 10 points. We therefore used 7 points (from very low to very high), similar to the paper version[2].

[2] https://humansystems.arc.nasa.gov/groups/TLX/downloads/TLXScale.pdf

Pilot Tests

We conducted two pilot tests to get an initial estimate of mean differences and variance. For each task and layout condition, we aggregated the results of the small and the large graph so that we had N = 4 for the power analysis. Based on the initial estimates, we computed the desired sample size for $\alpha = .05$, power = $1 - \beta = .80$, and number of comparisons $\tau = 3$ to estimate the number of required participants to be able to show a statistically significant difference between the best and worst performing layout for each task.

Task	**Dependent variable**	μ_1	μ_2	σ	**Estimated sample size**
FiN	completion time (s)	8	77	42	8
FCN	completion time (s)	23	130	57	6
ED	count deviation	0.26	0.18	0.14	64
SO O→D	angle deviation	14	32	18	21
FiP	completion time (s)	45	138	52	7
FoP	completion time (s)	22	54	35	25
SO D→D	angle deviation	28	85	36	9
SO D→O	angle deviation	33	14	17	17

Questionnaire Item	μ_1	μ_2	σ	**Estimated sample size**
SSQ Nausea	71	95	22	18
SSQ Disorientation	104	146	38	18
SSQ Oculomotor	64	90	26	21
NASA TLX	3	4.2	0.65	7

The tables show that, except for the "estimate degree" task, ≤ 25 participants are sufficient for 80% power.

Analysis

We used IBM SPSS Statistics 25 for the analysis. All completion times were log-transformed. All data was checked for normal distribution. If the data was not following the normal distribution, we checked for outliers and removed them, if necessary. If the data would not follow a normal distribution after outlier removal, we performed non-parametric tests. All post-hoc comparisons were Bonferroni-corrected. For reporting effect sizes, we use the partial eta-squared effect size ranges suggested by Draper[3], namely 0.01 for small, 0.06 for medium, and 0.14 for large.

Learning Effect

Find Nearest Neighbor – Completion Time

We assessed the learning effect using the log-transformed completion time of task FiN, as we can expect a large effect here from the pilot experiment. Indeed, run (i.e., the run in which a completion time was measured, independently of the layout) has a significant main effect:

Tests of Between-Subjects Effects

Dependent Variable: logTime

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.803[a]	2	3.902	3.424	.035
Intercept	13744.100	1	13744.100	12061.531	.000
run	7.803	2	3.902	3.424	.035
Error	167.506	147	1.139		
Total	13919.409	150			
Corrected Total	175.310	149			

a. R Squared = .045 (Adjusted R Squared = .032)

On average, users took 43 seconds to complete the task in the first run, 28 seconds in the second run, and 18 seconds in the third run.

Looking at the data for the small and large graphs individually, we see that we have a close-to significant effect of run on the log-transformed completion time when users were working with the small graphs:

3 http://www.psy.gla.ac.uk/~steve/best/effect.html

Tests of Between-Subjects Effects

Dependent Variable: logTime

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.951[a]	2	1.476	2.901	.061
Intercept	6610.116	1	6610.116	12994.390	.000
run	2.951	2	1.476	2.901	.061
Error	36.626	72	.509		
Total	6649.693	75			
Corrected Total	39.577	74			

a. R Squared = .075 (Adjusted R Squared = .049)

But there is no significant difference when working with the large graphs:

Tests of Between-Subjects Effects

Dependent Variable: logTime

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.150[a]	2	2.575	1.477	.235
Intercept	7139.073	1	7139.073	4095.930	.000
run	5.150	2	2.575	1.477	.235
Error	125.494	72	1.743		
Total	7269.716	75			
Corrected Total	130.643	74			

a. R Squared = .039 (Adjusted R Squared = .013)

We therefore treat the trials using the smaller graphs as training runs and only statistically evaluated the results obtained using the large graphs.

Questionnaires

TLX scores were normally distributed between the runs, SSQ scores were not. We therefore evaluated the effect of run on the task load using a univariate ANOVA and the effect on SSQ scores using a Kruskal Wallis test.

Tests of Normality

	Run	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
TLX	1	.117	24	.200*	.962	24	.484
	2	.087	25	.200*	.988	25	.986
	3	.162	26	.078	.923	26	.054
SSQ_Nausea	1	.162	24	.105	.936	24	.134
	2	.224	25	.002	.813	25	.000
	3	.160	26	.084	.885	26	.007
SSQ_Oculomotor	1	.178	24	.047	.905	24	.027
	2	.166	25	.072	.920	25	.051
	3	.144	26	.178	.911	26	.027
SSQ_Disorientation	1	.179	24	.044	.908	24	.032
	2	.152	25	.137	.879	25	.007
	3	.196	26	.011	.879	26	.005

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

There is no significant main effect of run on the reported task load:

Tests of Between-Subjects Effects

Dependent Variable: TLX

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.536[a]	2	.268	.274	.761
Intercept	914.020	1	914.020	935.210	.000
Run	.536	2	.268	.274	.761
Error	70.369	72	.977		
Total	985.227	75			
Corrected Total	70.905	74			

a. R Squared = .008 (Adjusted R Squared = -.020)

Similarly, there are no significant effects of run on nausea:

Independent-Samples Kruskal-Wallis Test Summary

Total N	75
Test Statistic	.042[a,b]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.979

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

Oculomotor:

Independent-Samples Kruskal-Wallis Test Summary

Total N	75
Test Statistic	.417[a,b]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.812

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

Or disorientation:

Independent-Samples Kruskal-Wallis Test Summary

Total N	75
Test Statistic	.001[a,b]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	1.000

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

We can therefore conclude that the within-subjects design did not have a significant impact on the questionnaire results.

Find Neighbor Completion Time

The log-transformed completion time was not normally distributed for Ego-Bubble.

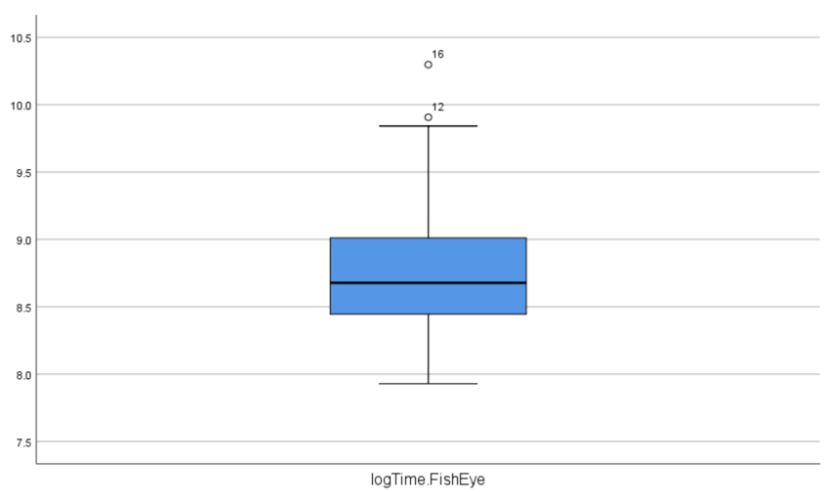
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
logTime.Ego-Bubble	.162	25	.090	.917	25	.045
logTime.Baseline	.159	25	.101	.942	25	.165
logTime.Ego-Highlight	.122	25	.200*	.962	25	.445

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

We therefore removed two outliers:



After removing these two outliers, the log-transformed completion time was normally distributed for all layout conditions:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
logTime.Ego-Bubble	.141	23	.200*	.933	23	.126
logTime.Baseline	.140	23	.200*	.940	23	.182
logTime.Ego-Highlight	.135	23	.200*	.959	23	.443

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

We therefore conducted a repeated-measures ANCOVA with layout as within-subjects factor and order as covariate.

The data is spherical:

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
layout	.951	1.011	2	.603	.953	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + order

Within Subjects Design: layout

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

There is a large and significant main effect of layout:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
layout	Sphericity Assumed	29.455	2	14.727	25.722	.000	.551	51.445	1.000
	Greenhouse-Geisser	29.455	1.906	15.453	25.722	.000	.551	49.028	1.000
	Huynh-Feldt	29.455	2.000	14.727	25.722	.000	.551	51.445	1.000
	Lower-bound	29.455	1.000	29.455	25.722	.000	.551	25.722	.998
layout * order	Sphericity Assumed	6.433	2	3.217	5.618	.007	.211	11.236	.833
	Greenhouse-Geisser	6.433	1.906	3.375	5.618	.008	.211	10.709	.819
	Huynh-Feldt	6.433	2.000	3.217	5.618	.007	.211	11.236	.833
	Lower-bound	6.433	1.000	6.433	5.618	.027	.211	5.618	.618
Error(layout)	Sphericity Assumed	24.047	42	.573					
	Greenhouse-Geisser	24.047	40.027	.601					
	Huynh-Feldt	24.047	42.000	.573					
	Lower-bound	24.047	21.000	1.145					

a. Computed using alpha = ,05

Order does not have a significant influence:

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	1025.661	1	1025.661	1299.516	.000	.984	1299.516	1.000
order	.116	1	.116	.147	.705	.007	.147	.065
Error	16.575	21	.789					

a. Computed using alpha = ,05

Bonferroni-adjusted post-hoc comparisons showed that the baseline (2) was significantly slower to complete than Ego-Highlight (3) and Ego-Bubble (1):

Pairwise Comparisons

Measure: MEASURE_1

(I) layout	(J) layout	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-2.373*	.197	.000	-2.885	-1.861
	3	-.572	.237	.075	-1.189	.044
2	1	2.373*	.197	.000	1.861	2.885
	3	1.801*	.233	.000	1.194	2.408
3	1	.572	.237	.075	-.044	1.189
	2	-1.801*	.233	.000	-2.408	-1.194

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

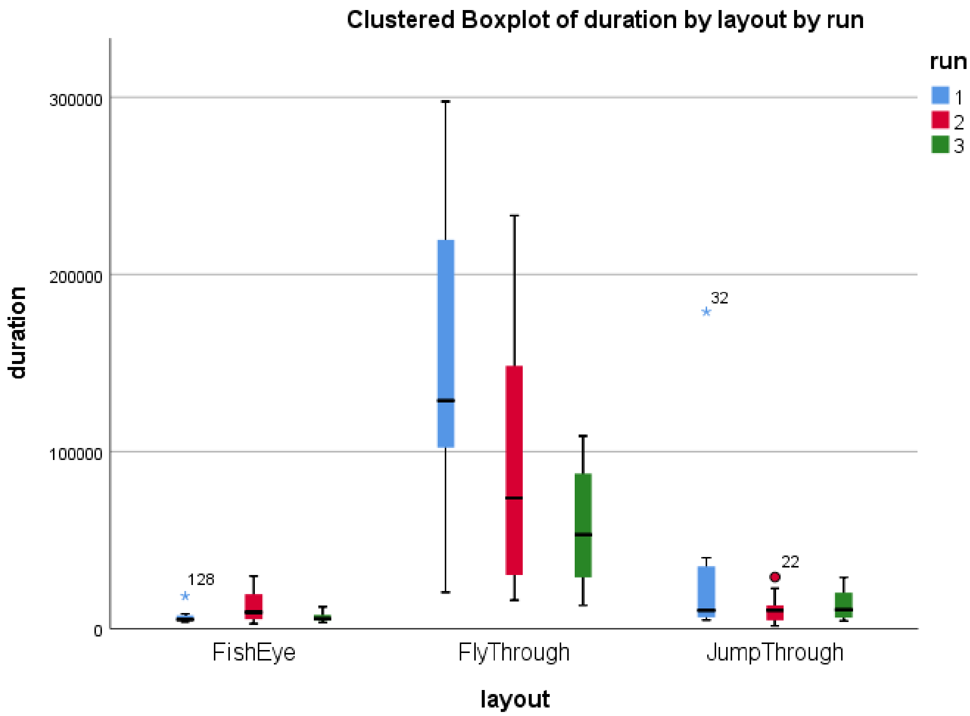
On average, Baseline took longest to complete, and Ego-Bubble was the fastest condition:

Report

seconds

layout	Mean	N	Std. Deviation
Ego-Bubble	8.6166	25	6.53234
Baseline	101.1232	25	79.95673
Ego-Highlight	20.0315	25	34.60888
Total	43.2571	75	64.76585

As there is an interaction effect between order and layout, we graphically explored the effect of order and layout:



It can be observed that the order had a noticeable effect on Baseline (FlyThrough), where a learning effect is visible. This learning effect is not visible for the other two conditions.

Find Common Neighbors Completion Time

We identified three outlier cases for Ego-Highlight. We removed the three subjects from the task analysis. The remaining 22 log-transformed completion times are normally distributed:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
logTime.Ego-Bubble	.076	22	.200*	.981	22	.929
logTime.Baseline	.104	22	.200*	.955	22	.398
logTime.Ego-Highlight	.117	22	.200*	.977	22	.860

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The data is not spherical:

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
layout	.625	8.916	2	.012	.728	.809	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + order
Within Subjects Design: layout

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

There is a large and significant main effect of layout:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
layout	Sphericity Assumed	7.710	2	3.855	20.853	.000	.510	41.705	1.000
	Greenhouse-Geisser	7.710	1.455	5.299	20.853	.000	.510	30.341	.999
	Huynh-Feldt	7.710	1.618	4.764	20.853	.000	.510	33.745	1.000
	Lower-bound	7.710	1.000	7.710	20.853	.000	.510	20.853	.991
layout * order	Sphericity Assumed	3.421	2	1.711	9.254	.000	.316	18.507	.968
	Greenhouse-Geisser	3.421	1.455	2.351	9.254	.002	.316	13.464	.917
	Huynh-Feldt	3.421	1.618	2.114	9.254	.001	.316	14.975	.937
	Lower-bound	3.421	1.000	3.421	9.254	.006	.316	9.254	.825
Error(layout)	Sphericity Assumed	7.394	40	.185					
	Greenhouse-Geisser	7.394	29.101	.254					
	Huynh-Feldt	7.394	32.365	.228					
	Lower-bound	7.394	20.000	.370					

a. Computed using alpha = ,05

Order does not have a significant influence:

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	1135.059	1	1135.059	6177.347	.000	.997	6177.347	1.000
order	.047	1	.047	.253	.620	.013	.253	.077
Error	3.675	20	.184					

a. Computed using alpha = ,05

Bonferroni-adjusted post-hoc comparisons showed that Baseline (2) was significantly slower to complete than Ego-Highlight (3) and Ego-Bubble (1):

Pairwise Comparisons

Measure: MEASURE_1

(I) layout	(J) layout	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-.826*	.163	.000	-1.253	-.400
	3	-.116	.096	.724	-.369	.136
2	1	.826*	.163	.000	.400	1.253
	3	.710*	.120	.000	.396	1.024
3	1	.116	.096	.724	-.136	.369
	2	-.710*	.120	.000	-1.024	-.396

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

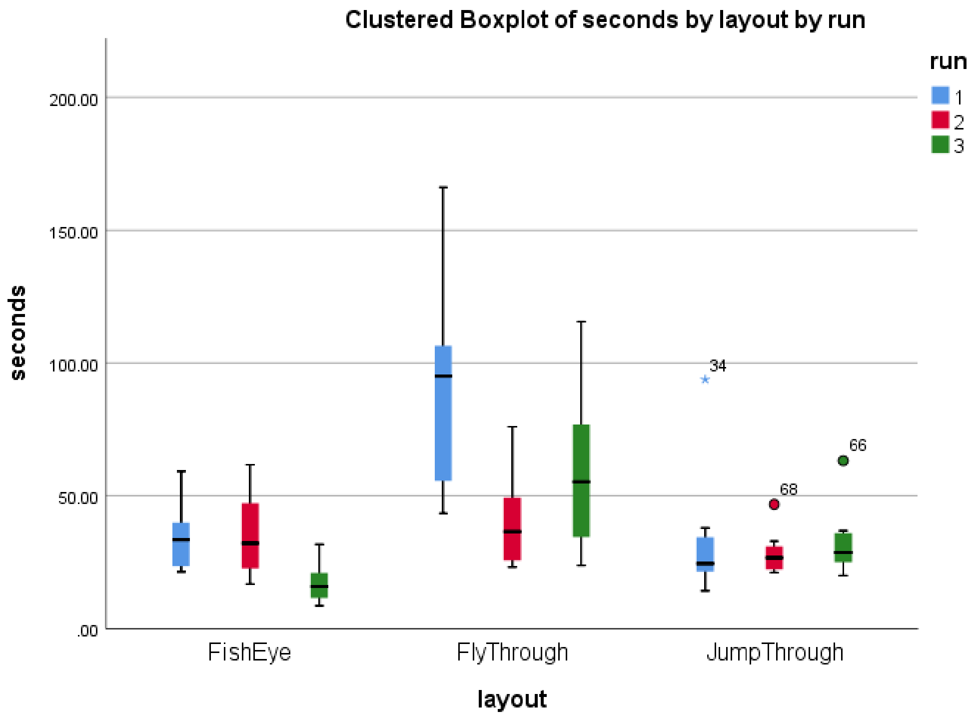
On average, Baseline was completed in 65 seconds, Ego-Highlight within 32 seconds, and Ego-Bubble within 28 seconds.

Report

seconds

layout	Mean	N	Std. Deviation
Ego-Bubble	28.3871	25	14.56900
Baseline	64.5030	25	37.44565
Ego-Highlight	31.6128	25	16.32020
Total	41.5010	75	29.66246

As there is an interaction effect between order and layout, we graphically explored the effect of the two variables:



It can be seen that only in the Baseline condition (FlyThrough), users were performing noticeably slower in the first run.

Find Common Neighbors Correctness

Unsurprisingly, none of the three correctness values are normally distributed:

Tests of Normality

	layout	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
correctRate	FisEye	.478	25	.000	.508	25	.000
	FlyThrou	.390	25	.000	.674	25	.000
	JumpThro	.534	25	.000	.308	25	.000
missRate	Ego-Bubble	.478	25	.000	.508	25	.000
	FlyThrou	.367	25	.000	.704	25	.000
	JumpThro	.534	25	.000	.308	25	.000
falsePositiveRate	Ego-Bubble	.404	25	.000	.557	25	.000
	FlyThrou	.506	25	.000	.445	25	.000
	JumpThro	.496	25	.000	.456	25	.000

a. Lilliefors Significance Correction

We therefore performed a non-parametric Friedman Test. There is no significant differences between the layout conditions for correctness rate.

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	4.348[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.114

a. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

There is also no significant difference for miss rate:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	5.429[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.066

a. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

And there is also no significant difference for false positive rate:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	1.351[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.509

a. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

The average correctness rate is highest for Ego-Highlight (92%) and lowest for Baseline (74%). The average miss rate is highest for Baseline (28%) and lowest for Ego-Highlight (8%), while the false positive rate is highest for Ego-Bubble (18%) and lowest for Ego-Highlight (11%).

Report

layout		correctRate	missRate	falsePositiveRate
Ego-Bubble	Mean	.8400	.1600	.1800
	N	25	25	25
	Std. Deviation	.34521	.34521	.33993
Baseline	Mean	.7400	.2800	.1600
	N	25	25	25
	Std. Deviation	.38514	.38406	.37417
Ego-Highlight	Mean	.9200	.0800	.1067
	N	25	25	25
	Std. Deviation	.27689	.27689	.28415
Total	Mean	.8333	.1733	.1489
	N	75	75	75
	Std. Deviation	.34222	.34380	.33172

Estimate Degree Error

We computed the degree estimation error as follows:

$err_{ED} = abs(\frac{ed-d}{d})$,

where *ed* is the user's node degree estimation, and *d* is the actual node degree.

The estimation error is not normally distributed for Ego-Highlight and Ego-Bubble:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
relDegreeDevAbs.Ego-Bubble	.204	25	.009	.868	25	.004
relDegreeDevAbs.Baseline	.113	25	.200*	.956	25	.335
relDegreeDevAbs.Ego-Highlight	.157	25	.114	.892	25	.013

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

We therefore performed a non-parametric Friedman test, which showed a significant effect of layout:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	6.720
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.035

Bonferroni-adjusted post-hoc comparisons showed a significant difference between Ego-Bubble and Baseline:

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
relDegreeDevAbs.Ego-Bubble-relDegreeDevAbs.Ego-Highlight	-.240	.283	-.849	.396	1.000
relDegreeDevAbs.Ego-Bubble-relDegreeDevAbs.Baseline	-.720	.283	-2.546	.011	.033
relDegreeDevAbs.Ego-Highlight-relDegreeDevAbs.Baseline	.480	.283	1.697	.090	.269

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

On average, the error was highest using Baseline and lowest using Ego-Bubble:

Report

relDegreeDevAbs

layout	Mean	N	Std. Deviation
Ego-Bubble	.1454	25	.13747
Baseline	.2550	25	.15618
Ego-Highlight	.1541	25	.13646
Total	.1849	75	.15028

The average raw deviation $(ed - d)$ shows that the errors were mainly underestimations, with Baseline having the most severe underestimation of neighboring nodes and Ego-Bubble the lowest:

Report

degreeDev

layout	Mean	N	Std. Deviation
Ego-Bubble	-4.0400	25	8.37894
Baseline	-6.5600	25	11.28450
Ego-Highlight	-4.8400	25	8.47880
Total	-5.1467	75	9.40772

In relation to the node degree d, this leads to a more than 10% underestimation using Baseline:

Report

relDegreeDev

layout	Mean	N	Std. Deviation
Ego-Bubble	-.0725	25	.18825
Baseline	-.1184	25	.27845
Ego-Highlight	-.0982	25	.18258
Total	-.0964	75	.21865

Spatial Orientation Overview → Detail

Angle deviation is not normally distributed for Ego-Bubble and Ego-Highlight:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
estimateAngle_OD_deviation .Ego-Bubble	.097	25	.200*	.963	25	.469
estimateAngle_OD_deviation .Baseline	.242	25	.001	.787	25	.000
estimateAngle_OD_deviation .Ego-Highlight	.168	25	.065	.853	25	.002

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

We therefore performed a non-parametric Friedman test. There is no significantly significant difference between the three layout conditions:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	.960[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.619

a. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

On average, the angle deviation was highest for Baseline and lowest for Ego-Bubble:

Report

estimateAngle_OD_deviation

layout	Mean	N	Std. Deviation
Ego-Bubble	16.238254561040340	25	10.801222429932475
Baseline	20.265427589149617	25	19.578469177190737
Ego-Highlight	19.393224857073590	25	18.890260929837353
Total	18.632302335754520	75	16.760730653669228

Find Shortest Path Correctness

We compared the user-reported paths to determine 1) whether the user-reported path is, in fact, a path, and 2) how much the length of the user-reported path deviates from the actual shortest path.

For all conditions, only around 12% of reported paths were no paths (one Ego-Highlight condition result is missing in the raw data):

layout * isPath Crosstabulation

Count

			isPath		
			FALSE	TRUE	Total
layout	Ego-Bubble	0	3	22	25
	Baseline	0	3	22	25
	Ego-Highlight	1	3	21	25
Total		1	9	65	75

There was one user who had all three paths incorrect, two users who had two incorrect paths, and two users with one incorrect path each. We filtered out these six cases with missing or incorrect path reports. The path deviation values were not normally distributed. Therefore, we performed a Friedman test, which did not show any significant differences:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	19
Test Statistic	.250[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.882

a. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

Only few of the reported paths were marginally longer than the ground truth shortest paths. On average, the reported paths were less than 0.32 nodes longer than the ground truth path with 5 nodes using Ego-Bubble, and 0.16 nodes longer using Baseline.

Report

	pathDeviation.Ego-Bubble	pathDeviation.Baseline	pathDeviation.Ego-Highlight
Mean	.32	.16	.21
N	19	19	19
Std. Deviation	.820	.501	.535

The highest rate of correctly reported shortest paths was obtained using Baseline (89.5%), and 84.2% with Ego-Highlight and Ego-Bubble. The longest path was reported by one user with Ego-Bubble (8 nodes).

pathDeviation.Ego-Bubble

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	16	84.2	84.2	84.2
	1	1	5.3	5.3	89.5
	2	1	5.3	5.3	94.7
	3	1	5.3	5.3	100.0
	Total	19	100.0	100.0	

pathDeviation.Baseline

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	17	89.5	89.5	89.5
	1	1	5.3	5.3	94.7
	2	1	5.3	5.3	100.0
	Total	19	100.0	100.0	

pathDeviation.Ego-Highlight

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	16	84.2	84.2	84.2
	1	2	10.5	10.5	94.7
	2	1	5.3	5.3	100.0
	Total	19	100.0	100.0	

Find Path Completion Time

We analyzed the log-transformed completion times only for those cases, where the reported path was actually a path, but not necessarily the shortest one. That means, that we removed 6 cases. The remaining 19 cases were normally distributed with respect to the log-transformed completion time with the exception of Ego-Bubble. We removed one additional outlier case, so that we ended up with a normal distribution.

Log-transformed completion times for Ego-Bubble were not normally distributed. We removed the two outlier cases, then the completion times followed a normal distribution:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
logTime.Ego-Bubble	.114	18	.200*	.973	18	.858
logTime.Baseline	.126	18	.200*	.933	18	.222
logTime.Ego-Highlight	.126	18	.200*	.962	18	.632

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The data is spherical:

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
layout	.970	.464	2	.793	.970	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + order

Within Subjects Design: layout

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

There is a moderately large, yet insignificant main effect of layout on completion time, and also no interaction with order:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
layout	Sphericity Assumed	1.497	2	.749	2.956	.066	.156	5.912	.535
	Greenhouse-Geisser	1.497	1.941	.771	2.956	.068	.156	5.737	.526
	Huynh-Feldt	1.497	2.000	.749	2.956	.066	.156	5.912	.535
	Lower-bound	1.497	1.000	1.497	2.956	.105	.156	2.956	.366
layout * order	Sphericity Assumed	.486	2	.243	.960	.394	.057	1.920	.202
	Greenhouse-Geisser	.486	1.941	.251	.960	.392	.057	1.863	.199
	Huynh-Feldt	.486	2.000	.243	.960	.394	.057	1.920	.202
	Lower-bound	.486	1.000	.486	.960	.342	.057	.960	.152
Error(layout)	Sphericity Assumed	8.104	32	.253					
	Greenhouse-Geisser	8.104	31.053	.261					
	Huynh-Feldt	8.104	32.000	.253					
	Lower-bound	8.104	16.000	.507					

a. Computed using alpha = ,05

Also, the order does not have an effect:

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	964.527	1	964.527	3559.135	.000	.996	3559.135	1.000
order	.017	1	.017	.064	.804	.004	.064	.057
Error	4.336	16	.271					

a. Computed using alpha = ,05

Spatial Orientation Detail → Detail

Angle deviations were not normally distributed for Ego-Bubble and Ego-Highlight:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
estimateAngle_DD_deviation .Ego-Bubble	.207	25	.007	.866	25	.004
estimateAngle_DD_deviation .Baseline	.146	25	.176	.962	25	.461
estimateAngle_DD_deviation .Ego-Highlight	.238	25	.001	.791	25	.000

a. Lilliefors Significance Correction

We therefore performed a non-parametric Friedman test. No statistically significant differences concerning angle deviations were found:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	.080[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.961

a. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

On average, the angle deviations were very similar with high standard deviations, with Ego-Bubble having slightly lower average errors than Baseline and Ego-Highlight:

Report

estimateAngle_DD_deviation

layout	Mean	N	Std. Deviation
Ego-Bubble	31.765462356652420	25	22.060983516263110
Baseline	33.851428040600340	25	18.662564575875530
Ego-Highlight	39.903043737323110	25	34.722526771507454
Total	35.173311378191990	75	25.959545368098550

Follow Path Time

The log-transformed completion times were normally distributed for all three conditions:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
logFPTime.Ego-Bubble	.116	25	.200*	.969	25	.618
logFPTime.Baseline	.145	25	.185	.944	25	.185
logFPTime.Ego-Highlight	.116	25	.200*	.953	25	.295

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The data is spherical:

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

					Epsilon[b]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
layout	.922	1.699	2	.428	.928	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + run.Baseline

Within Subjects Design: layout

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

There is a large and significant main effect of layout, but no interaction between order and layout:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
layout	Sphericity Assumed	7.152	2	3.576	49.243	.000	.691	98.486	1.000

	Greenhouse-Geisser	7.152	1.856	3.854	49.243	.000	.691	91.383	1.000
	Huynh-Feldt	7.152	2.000	3.576	49.243	.000	.691	98.486	1.000
	Lower-bound	7.152	1.000	7.152	49.243	.000	.691	49.243	1.000
layout * order	Sphericity Assumed	.611	4	.153	2.104	.096	.161	8.415	.577
	Greenhouse-Geisser	.611	3.712	.165	2.104	.102	.161	7.808	.553
	Huynh-Feldt	.611	4.000	.153	2.104	.096	.161	8.415	.577
	Lower-bound	.611	2.000	.306	2.104	.146	.161	4.207	.385
Error(layout)	Sphericity Assumed	3.195	44	.073					
	Greenhouse-Geisser	3.195	40.827	.078					
	Huynh-Feldt	3.195	44.000	.073					
	Lower-bound	3.195	22.000	.145					

a. Computed using alpha = ,05

There is also a large and significant effect of the order:

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	7642.822	1	7642.822	31059.064	.000	.999	31059.064	1.000
order	2.513	2	1.257	5.107	.015	.317	10.214	.765
Error	5.414	22	.246					

a. Computed using alpha = ,05

Bonferroni-adjusted post-hoc comparisons confirmed that Baseline (2) was significantly slower than both, Ego-Highlight (3) and Ego-Bubble (1):

Pairwise Comparisons

Measure: MEASURE_1

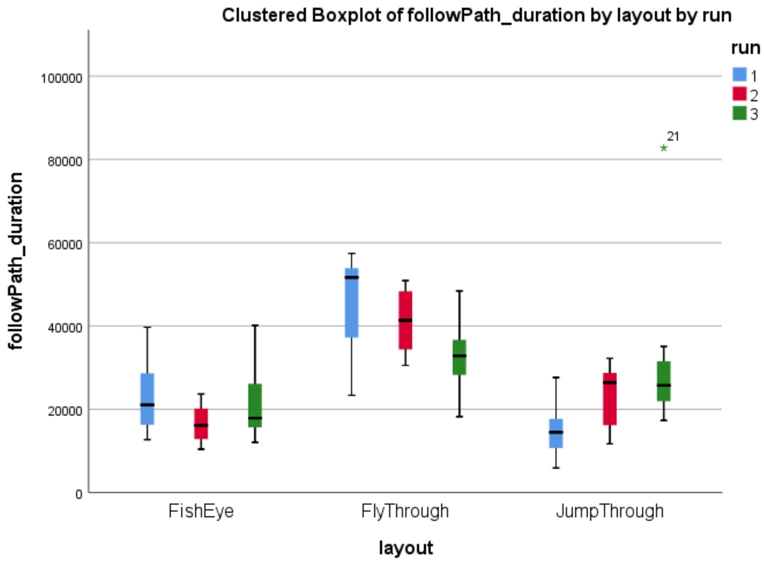
(I) layout	(J) layout	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-.686*	.065	.000	-.855	-.517
	3	-.065	.084	1.000	-.282	.152
2	1	.686*	.065	.000	.517	.855
	3	.621*	.079	.000	.417	.825
3	1	.065	.084	1.000	-.152	.282
	2	-.621*	.079	.000	-.825	-.417

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

We explored the effect of the order on the completion time by investigating the influence of the order of each individual condition. For Baseline (FlyThrough), we can see a learning effect, i.e., the later Baseline appeared in the sequence of conditions, the faster users could follow the given path.



On average, Baseline required almost 40 seconds to follow the path, and Ego-Bubble and Ego-Highlight required less than 25 seconds:

Report

followPath_duration

layout	Mean	N	Std. Deviation
Ego-Bubble	20509.84	25	8039.306
Baseline	39714.24	25	10750.550
Ego-Highlight	23459.20	25	14549.926
Total	27894.43	75	14119.905

Spatial Orientation Detail → Overview

Angle deviations were not normally distributed:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
estimateAngle_DO_deviation .Ego-Bubble	.156	25	.120	.830	25	.001
estimateAngle_DO_deviation .Baseline	.239	25	.001	.723	25	.000
estimateAngle_DO_deviation .Ego-Highlight	.205	25	.008	.818	25	.000

a. Lilliefors Significance Correction

We therefore performed a non-parametric Friedman test. No statistically significant differences concerning angle deviations were found:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	3.440[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.179

a. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

On average, angle deviations were similar with a high standard deviation. Ego-Bubble had a slightly higher average error than Ego-Highlight and Baseline:

Report

estimateAngle_DO_deviation

layout	Mean	N	Std. Deviation
Ego-Bubble	23.495634375983904	25	16.307442743896296
Baseline	18.470322689324266	25	18.474293153009560
Ego-Highlight	20.398679617929012	25	17.061664949213263
Total	20.788212227745724	75	17.195750822118278

Simulator Sickness

Nausea

The computed nausea score is not normally distributed:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SSQ_Nausea.Ego-Bubble	.218	25	.003	.878	25	.006
SSQ_Nausea.Baseline	.187	25	.024	.873	25	.005
SSQ_Nausea.Ego-Highlight	.168	25	.066	.868	25	.004

a. Lilliefors Significance Correction

We therefore performed a Friedman test, which did not show any significant differences between the layout conditions:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	2.741[a]
Degree Of Freedom	2

Asymptotic Sig.(2-sided test)	.254

a. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

Oculomotor

The computed oculomotor scores were not normally distributed for the three layout conditions:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SSQ_Oculomotor.Ego-Bubble	.122	25	.200*	.945	25	.196
SSQ_Oculomotor.Baseline	.141	25	.200*	.936	25	.120
SSQ_Oculomotor.Ego-Highlight	.216	25	.004	.832	25	.001

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The Friedman test showed a significant effect of layout:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	8.195
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.017

Bonferroni-adjusted post-hoc comparisons showed a significant difference between Ego-Bubble and Baseline:

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
SSQ_Oculomotor.Ego-Bubble-SSQ_Oculomotor.Ego-Highlight	-.480	.283	-1.697	.090	.269
SSQ_Oculomotor.Ego-Bubble-SSQ_Oculomotor.Baseline	-.720	.283	-2.546	.011	.033
SSQ_Oculomotor.Ego-Highlight-SSQ_Oculomotor.Baseline	.240	.283	.849	.396	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Disorientation

Values for disorientation are not normally distributed:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SSQ_Disorientation.Ego-Bubble	.207	25	.007	.867	25	.004
SSQ_Disorientation.Baseline	.181	25	.034	.918	25	.046
SSQ_Disorientation.Ego-Highlight	.170	25	.061	.852	25	.002

a. Lilliefors Significance Correction

According to a Friedman test, there is a significant difference between the layout conditions with respect to disorientation:

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	25
Test Statistic	8.617
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.013

Bonferroni-adjusted post-hoc comparisons showed a significant difference between Ego-Bubble and Baseline:

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
SSQ_Disorientation.Ego-Bubble-SSQ_Disorientation.Ego-Highlight	-.460	.283	-1.626	.104	.312
SSQ_Disorientation.Ego-Bubble-SSQ_Disorientation.Baseline	-.740	.283	-2.616	.009	.027
SSQ_Disorientation.Ego-Highlight-SSQ_Disorientation.Baseline	.280	.283	.990	.322	.967

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Task Load Index

The computed task load index was normally distributed across all three conditions:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TLX.Ego-Bubble	.137	25	.200*	.972	25	.696
TLX.Baseline	.088	25	.200*	.962	25	.463
TLX.Ego-Highlight	.131	25	.200*	.946	25	.207

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The data is spherical:

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

					Epsilon[b]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
condition	.838	4.072	2	.131	.860	.921	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept

Within Subjects Design: condition

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Layout has a large and significant effect on the reported task load:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
condition	Sphericity Assumed	4.393	2	2.197	6.546	.003	.214	13.092	.891
	Greenhouse-Geisser	4.393	1.721	2.553	6.546	.005	.214	11.264	.853
	Huynh-Feldt	4.393	1.841	2.386	6.546	.004	.214	12.052	.870
	Lower-bound	4.393	1.000	4.393	6.546	.017	.214	6.546	.690
Error(condition)	Sphericity Assumed	16.107	48	.336					
	Greenhouse-Geisser	16.107	41.299	.390					
	Huynh-Feldt	16.107	44.188	.365					
	Lower-bound	16.107	24.000	.671					

a. Computed using alpha = ,05

Post-hoc comparisons reveal that the task load was assessed significantly higher for Baseline (2) than for the other two conditions:

Pairwise Comparisons

Measure: MEASURE_1

(I) condition	(J) condition	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-.541*	.194	.030	-1.040	-.042
	3	-.061	.143	1.000	-.429	.306
2	1	.541*	.194	.030	.042	1.040
	3	.480*	.150	.012	.094	.866
3	1	.061	.143	1.000	-.306	.429
	2	-.480*	.150	.012	-.866	-.094

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

However, the average reported task load was also medium for Ego-Highlight and Ego-Bubble, while it was tending towards moderately strong for Baseline:

Descriptive Statistics

	Mean	Std. Deviation	N
TLX.Ego-Bubble	3.2907	.77719	25
TLX.Baseline	3.8320	1.14166	25
TLX.Ego-Highlight	3.3520	.92946	25

User Rankings

The following ranks (3 = lowest, 1 = highest) were given by the users:

	Baseline	Ego-Highlight	Ego-Bubble
3	15	5	5
2	6	9	9
1	4	11	11

Mind that one user assigned rank 1 to both Ego-Highlight and Ego-Bubble.

The following **positive** aspects were mentioned by the users:

Baseline

- flying is good to get an overview
- flying helps to orient
- handy to be able to look anywhere
- unchanged; simple; free movement
- free navigation, more dynamic
- free movement, wider perspective, easier to find neighbors, flying is fun
- good for counting neighbors
- best for node degree estimation
- free movement
- not necessarily harder to use

- free navigation
- estimate nodes degree is easier
- freedom to look

Ego-Highlight:

- yellow neighbors
- yellow neighbors
- nothing moves
- did not have to fly, more straight-forward; more controlled
- neighbors visible; counting easier
- better overview compared to the "pipes" [FT]
- very similar to Ego-Bubble
- yellow neighbors; no distortion;
- common neighbors nicely visible
- good for orientation
- highlighting of direct neighbors is more important than seeing the links
- neighbor highlighting
- neighbor highlighting
- better orientation than in Ego-Bubble

Ego-Bubble:

- yellow neighbors
- felt easier; no need to move around; direct neighbors in yellow
- less visual clutter
- neighbors better visible
- neighbors visible; counting easier
- less feeling of vertigo; somehow the movement does not get lost
- easiest to handle
- less visual clutter
- highlighted neighbors
- direct neighbor highlight + removes unnecessary elements that clutter the view, thus "tidier"
- less motion sickness, tidier overview (since less link clutter)
- tidier overview, especially compared to Baseline; direct neighbor highlighting; not so much diff to Ego-Highlight noticed - causing less nausea
- neighbor highlighting

The following **negative** aspects were mentioned:

Baseline:

- standing in the node: areas very confusing; have to go out first; lines disturb the view

- difficult to see with all the connections; neck pain because you have to look up
- getting sick from the flying
- for some tasks harder; less differentiation between neighbors and non-neighbors
- visual clutter; flickering during highlighting; occlusion by edges; dizziness; not accurate; movement direction inaccurate
- slowly to move around; less efficient; exhausting; seeing direct neighbors is difficult
- neighbor information was missing; context was missing
- occlusions due to edges; overview is destroyed; one has to move to look around --> loss of orientation
- everything is completely occluded; nothing to see; slow; motion sickness
- direct neighbor edges caused clutter; orientation harder (also because of hariball layout)
- very hard to find common neighbors and shortest path
- links cause too much clutter
- visual clutter
- visual clutter
- dizziness / motion sickness when freely navigating
- no direct neighbor highlighting; does not offer good overview; cause more nausea when looking around because there is no "fix point" (anchor for orientation)
- no direct neighbor highlighting
- free flying in viewing direction is hard; would prefer flying in direction of laser

Ego-Highlight:

- would be better with flying
- sometimes, nodes are very far away
- one has to abstract, but this is easy to learn
- very rigid; as if one would "fall" into it; feeling of relation got lost; there is nothing one can do
- confusing (because it was the first condition?)
- restriction to moving only to node positions = stiff, rotating graph in immersive perspective would be a good substitute, lot of occlusion but no way to deal with it since movement is not possible freely
- visual clutter, no free movement
- complicated to do tasks (FoP)
- some nodes are occluded

Ego-Bubble:

- cut edges; zero advantage compared to Ego-Highlight
- hard to orient; changes the graph; finding the path difficult; nodes moving towards oneself; difficult to know how far away a node actually was
- no reference (due to distortion?) - no free movement
- Ego-Highlight and Baseline have their advantages depending on task; Ego-Bubble has no special advantage compared to Ego-Highlight