Do You See What I Mean? The use of eye-tracking data in readability and accessibility research

Victoria Yaneva
University of Wolverhampton

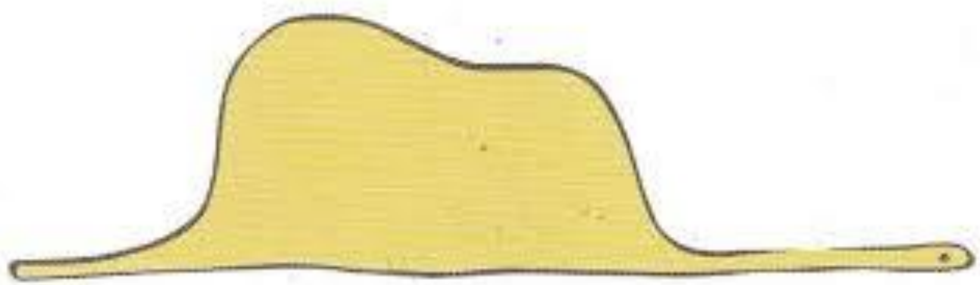
If there are any points on which you require explanation or further particulars we shall be glad to furnish such additional details as may be required by telephone.

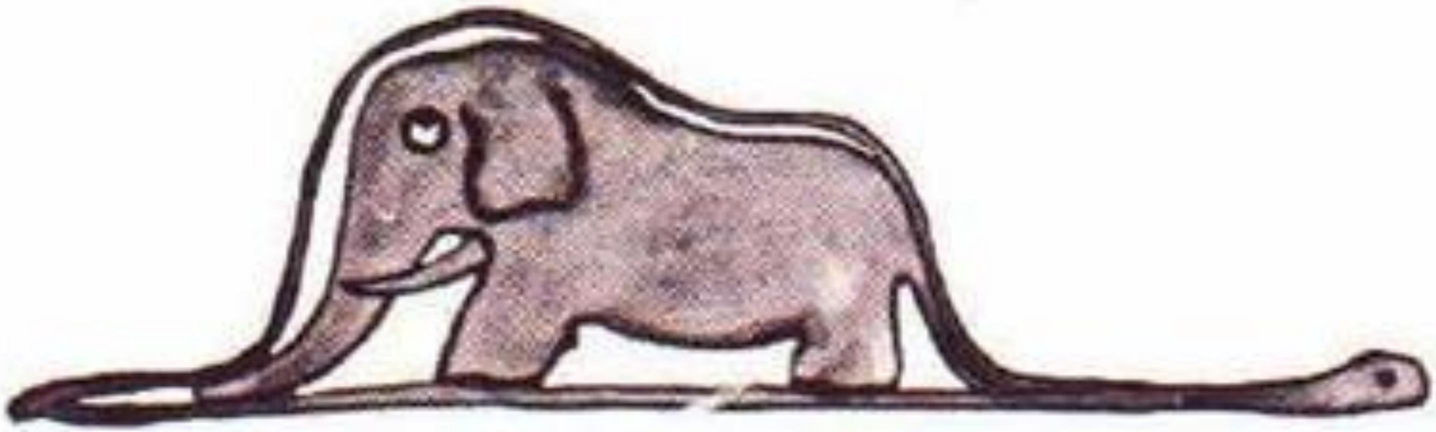
↓

If you have any questions, please phone.

Plain English Campaign

Available at: http://www.plainenglish.co.uk





Autism and Reading

Difficulties in:

- processing complex sentences
- comprehending figurative language
- comprehending long or abstract words
- making pragmatic inferences
- referring to the whole

(Frith and Snowling, 1983; Happe, 1997;
O'Connor and Klein, 2004; MacKay and Shaw, 2004)

Outline

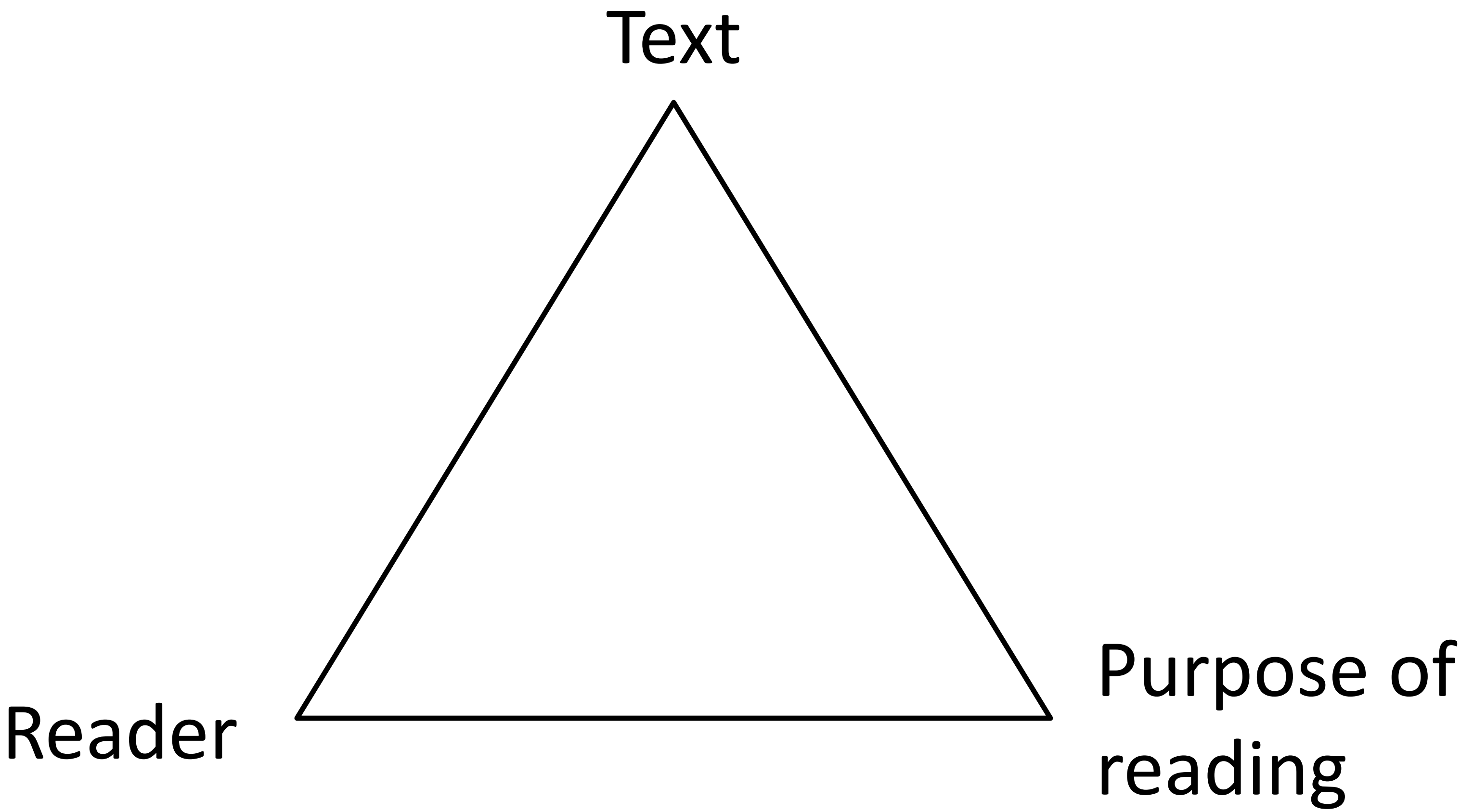
- Why I talk about readability and gaze data
- Collecting eye tracking data during a reading task
- A few experiments:
  - Predicting indivudual comprehension from gaze data
  - Sentence-level readability
  - Accessibility of web pages
- Conclusion: sharing is caring!

Why I talk about readability and gaze data

Defining Readability

“The purpose of readability assessment is to effect a ‘best match’ between intended readers and texts; thus, optimal difficulty comes from an interaction among the text, the reader, and his/her purpose for reading"

(Chall & Dale 1995)


Text
Reader
Purpose of reading

We need more information about the **reader** and about the **process.**

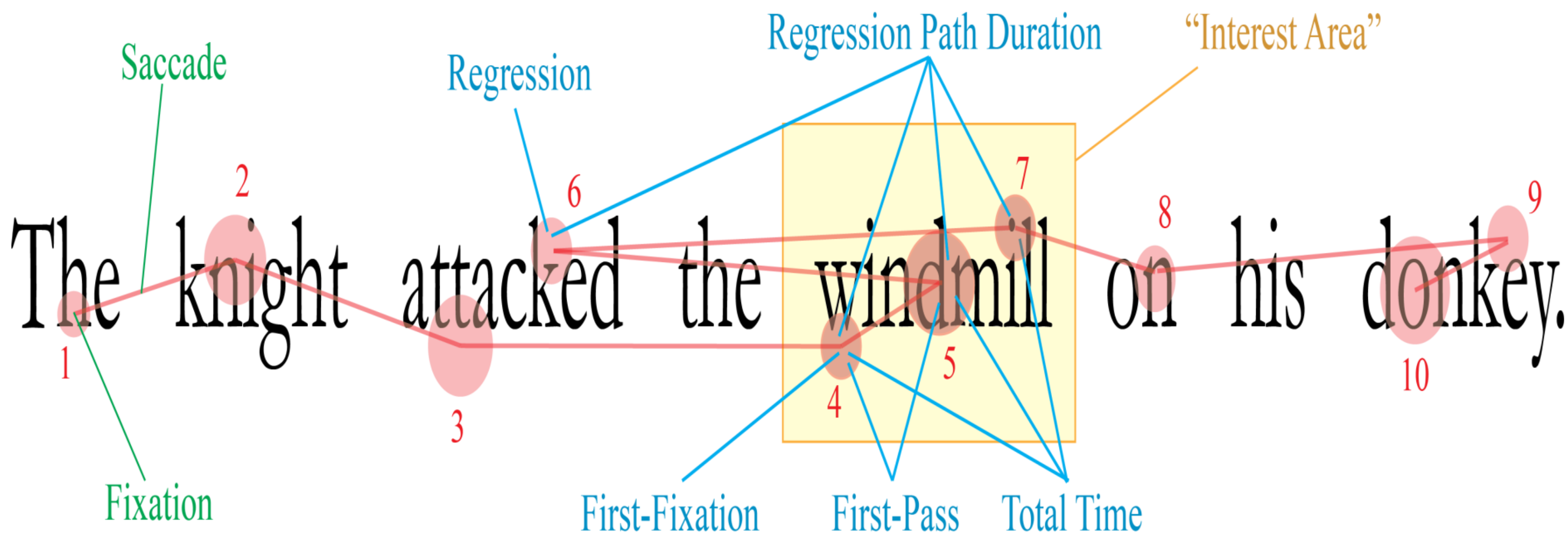
This is especially important if we investigate **neurodiverse reader groups.**

The Strong Eye–Mind Hypothesis

"There is no appreciable lag between what is fixated and what is processed"

(Just and Carpenter, 1980)

Eye tracking



(Cusimano, 2012)

Eye tracking data provides:

- Insight into the online cognitive processing of the text → Experiment on using gaze data to predict comprehension
- User data about the processing of smaller units (e.g. words or sentences) → Experiment on sentence readability classification using gaze data
- Insight into the cognitive load imposed on the reader/user → Experiments on the processing of information on web pages

Data Collection

Participants

27 adults with a confirmed diagnosis of autism and 31 participants without autism

Texts	Group	Participants	Age in years	Years of schooling
1 - 9	ASD	9 (5 male, 4 female)	$\mu = 33$, SD = 9.18	$\mu = 15.66$, SD = 2.12
1 - 9	Control	9 (5 male, 4 female)	$\mu = 31.33$, SD = 7.48	$\mu = 16.88$, SD = 1.83
10 - 17	ASD	14 (8 male, 6 female)	$\mu = 37.9$, SD = 9.6	$\mu = 16$, SD = 3.77
10 - 17	Control	14 (10 male, 4 female)	$\mu = 33.42$, SD = 8.77	$\mu = 14$, SD = 17.71
18 - 20	ASD	8 (7 male, 1 female)	$\mu = 36.5$, SD = 9.78	$\mu = 15.63$, SD = 3.74
18 - 20	Control	10 (6 male, 4 female)	$\mu = 31.3$, SD = 6.4	$\mu = 18.1$, SD = 2.6

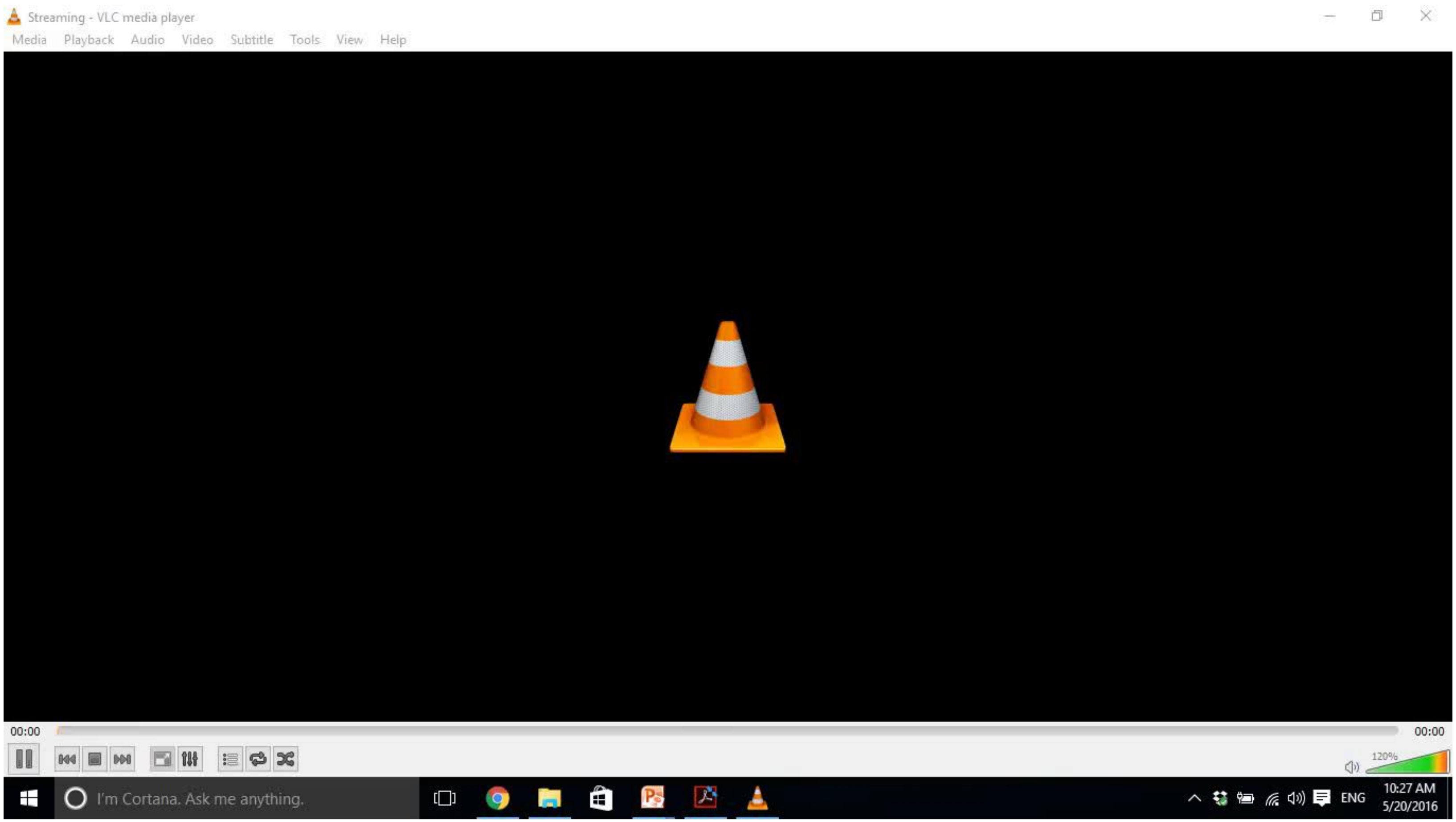
Materials

- 20 text passages with varying complexity
- Miscellaneous registers : educational (7), news (10) and general informational articles (3)
- Average number of words per text was 156, SD = 49.94
- A range of readability levels covered; mean Flesch Reading Ease score was 65.07, SD = 13.71

Questions

- Literal MCQs
- Reorganisation MCQs
- Gap Inference MCQs

Recordings


Streaming - VLC media player
Media Playback Audio Video Subtitle Tools View Help
00:00
00:00
120%
I'm Cortana. Ask me anything.
ENG
10:27 AM
5/20/2016

The Corpus

Item	AOI name	POS	Coref	ASD ATV	ASD AF	ASD AR	Con ATV	Con AF	Con AR
13	< /s >< s >								
14	Your	prp$	set 11	0.225	2.229	2.618	0.221	2.234	2.505
15	team	nn		0.22	2.219	2.447	0.213	2.075	2.076
16	is	vbz		0.112	1.704	1.959	0.108	1.859	2.024
17	losing	vbg		0.255	2.155	2.438	0.297	2.4	72.89
18	by	in							
19	just	rb		0.198	1.833	2.094	0.194	1.788	2.067
20	one	cd		0.159	1.945	1.945	0.149	1.762	2.051
21	goal	nn		0.188	1.903	1.852	0.184	1.966	2.789
22	.	.							
23	< /s >< s >								

Finally, some experiments….

Experiment 1: Using gaze data to predict comprehension

(Yaneva and Søgaard (under review))

Related work

- Best predictor of comprehension: average duration of gaze fixations
- Bad predictors of comprehension: number of fixations per sentence and overall reading time per sentence

Underwood et al. (1990)

- Best predictors: fixation durations and the distance between consecutive fixations

Martinez-Gomez and Aizawa (2014)

Our approach

- ASD -> ASD
- ASD -> Control
- Control -> ASD
- Control -> Control

Our Approach

- Random forests algorithm
- Instance weighting (Shi modaira, 2000) was used to facilitate adaptation between groups

- Each instance weight is computed by training a random forest classifier to distinguish between target and test data (our two groups of subjects), ignoring the comprehension scores.

- We then weight each training data point by the probability in our model that this data point belongs to the test data.

- We train our random forest classifier on this weighted sample in order to better predict comprehension scores

Results

From	**To**	Baseline	Instance weighting
Control	Control	86.8	**89.9**
Control	Autists	93.0	**97.8**
Autists	Control	88.9	88.9
Autists	Autists	93.6	93.6

Table 2: F_1 scores for random forests vs. random forests with instance weighting.

Feature analysis

- Most predictive feature for the control group: time viewed (sum of fixation and revisit length).
- Most predictive feature for the ASD group: number of regressions.

Predicting Comprehension: Conclusions

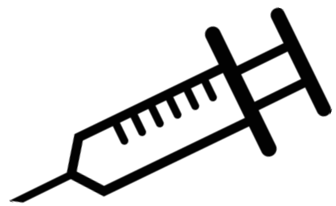
- Different reading strategies for the two groups
- Nevertheless, there are systematic signals of comprehension that transfer between groups
- Adapting to the slightly different reading patterns by using instance weighing leads to improved performance

Experiment 2: Sentence-level Readability Classification for Readers with Autism

(Yaneva, 2016)

The data

157 sentences from the ASD corpus



100 Sentences with controlled length

(Laufer and Nation, 1999)

The Data: Sentences from the ASD corpus

- Ranked based on the average number of fixations
- Split into two classes using median split (M = 10.66)

Examples

Easy: "Stretching helps loosen tight muscles and tissues".

Difficult: "Their album "Yesterday and Today" (also known as the "Butcher Album") is highly collectible and if you have an original it is highly priced and is one of the holy grails of record collecting.".

The Data: 100 Controlled-length sentences (1)

Item example:

The story is very <didactic>.

a) tries hard to teach something

b) is very difficult to believe

c) deals with exciting actions

d) is written with unclear meaning

(Laufer and Nation, 1999)

The Data: 100 Controlled-length sentences (2)

Threshold for “easy” sentences (65 sentences) to have a minimum of 60% correct answers from all participants.

Examples of difficult sentences:

“That was an excellent soliloquy!”

“It was very bawdy.”

“He rode roughshod.”

“Whose sloop is that?”

The Data: Overall

- 257 sentences in total
- ASD corpus sentences: 97 easy and 98 difficult
- Controlled length sentences: 65 easy and 35 difficult
- Total = 162 easy sentences and 133 difficult sentences

Features: Shallow Descriptors

Table 5.2: Sentence classification: Shallow descriptors

Label	Feature	Description
DESWC	Word count	Number of words in the sentence
DESWLsy	Word length in syllables, m	Average number of syllables for all words
DESWLsyd	Word length in syllables, SD	SD of the mean number of syllables for all words
DESWLlt	Word length in letters, m	Average number of letters for all words
DESWLltd	Word length in letters, SD	SD of the mean number of letters measure
DESSL	Sentence length in words, m	Average number of words for all sentences
DESSLd	Sentence length in words, SD	SD of the mean number of words for all sentences

Cognitively-motivated Features

Table 5.4: Sentence classification: Cognitively-motivated features

Label	**Feature**	**Description**
WRDFRQc	CELEX word freq., m	Average freq. for words in CELEX database
WRDFRQa	CELEX Log freq. (all), m	Log freq. for all words in CELEX database
WRDFRQmc	CELEX Log min freq., m	Log min. freq. for words in CELEX database
WRDAOAc	Age of acquisition, m	Age of acquisition norms from MRC
WRDFAMc	Familiarity, m	Familiarity norms from MRC
WRDCNCc	Concreteness, m	Concreteness norms from MRC
WRDIMGc	Imagability, m	Imagability norms from MRC
WRDMEAc	Meaningfulness,m	Meaningfulness norms (Nickerson & Cartwright 1984)
WRDPOLc	Polysemy, m	Number of core meanings of the word (Miller 1995)
WRDHYPn	Hypernymy for Ns, m	Sub- and superordinate WordNet relations (nouns)
WRDHYPv	Hypernymy for Vs, m	Sub- and superordinate WordNet relations (verbs)
WRDHYPnv	Hypernymy for Ns and Vs, m	WordNet relations (nouns and verbs)
SYNLE	Left embeddedness, m	Number of words before the main verb
SYNNP	Modifiers per NP, m	Number of modifiers per noun phrase

Modelling

- Algorithm: Best performance achieved by the SPegasos classifier

(Shalev-Shwartz et al. 2011)

- Baseline: Sentence length in words
- Feature selection: Best First attribute selection filter for supervised learning built in Weka

(Frank & Witten 1998)

- Training and evaluation: 10-fold cross-validation

Results

Table 5.7: Sentence-classifier results for 10-fold cross-validation

	Baseline	All features	Selected features
Precision	0.816	0.745	0.841
Recall	0.79	0.743	0.817
F	0.787	0.743	0.815
Accuracy	0.78	0.74	0.82

Sentence Readability: Conclusion

- Comparison to other sentence-level classifiers:
- READIT (Dell'Orletta et al. 2011) report 78.2% accuracy for sentences in Italian;
- (Vajjala & Meurers 2014) - 80% accuracy by using pairs of original and manually simplified sentences from news articles;
- (Pilan et al. 2014) report 71% accuracy for classifying Swedish sentences for foreign language learners.
- (Inui et al. (2001) - 95% precision and 89% recall

Experiment 3: Accessibility of Web Pages

Eraslan et al. (in preparation)

Design

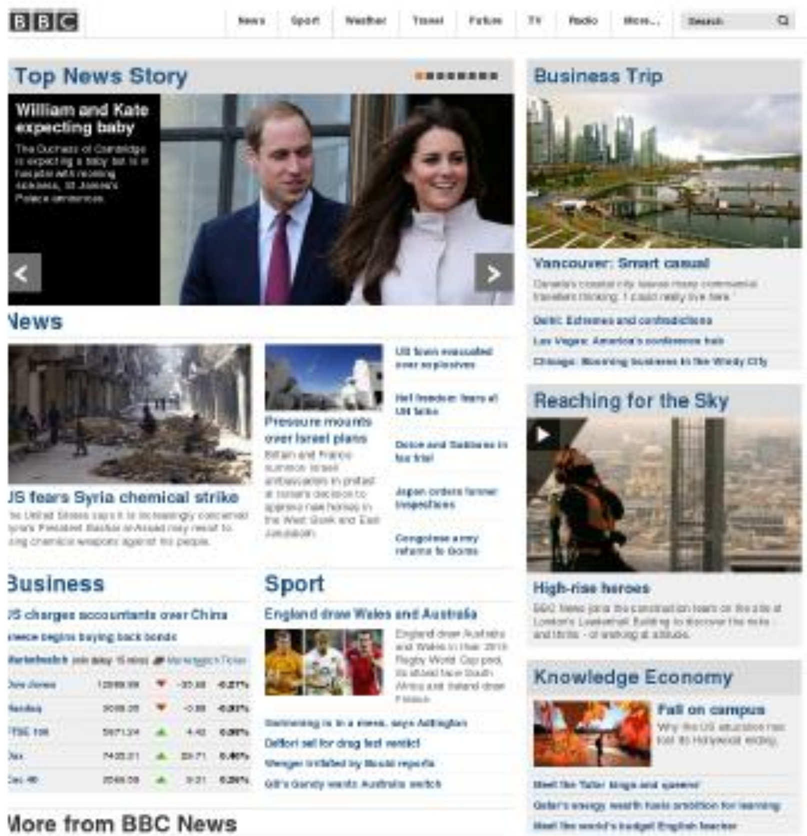
- 18 participants with ASD and 18 control participants
- 6 web pages with varying visual complexity
- 2 tasks per page, 30 seconds time limit

Example: “Can you locate the link that allows watching the TV ads relating to iPad mini?”


iPad mini
Every inch an iPad.
iPad
The new iMac
The new iTunes is here.
iPhone 5


AVG
For Home
For Business
For Mobile
Support
FREE Mobile Protection
NEW! Ultimate Protection
for your PC and online identity
• Award-winning virus protection
• Faster scan and video streaming
• Protects your identity when shopping and banking
FREE TRIAL
Protection
Backup & PC Performance
Small business solutions
AVG AntiVirus FREE 2013
Essential Protection. Easy to use.
AVG AntiVirus 2013
Advanced Protection. Easy to use.
AVG Premium Security 2013
Ultimate Protection. Privacy. Performance.
About AVG
How can we help
Partner with AVG
Media Center
Customer area
Connect with AVG


Get a fast, free translation solution
Babylon invites you to download its famous translation software – Free of Charge!
One click translation
Over 75 languages
Over 2,000 dictionaries
Download
Free version
Other products
Full Text Translation
Human Translation
Learn a Language
Corporate Customers


BBC
News
Sport
Weather
Travel
Future
TV
Radio
More...
Search
Top News Story
William and Kate expecting baby
Business Trip
Vancouver: Smart casual
News
US fears Syria chemical strike
Pressure mounts over Israel plans
Reaching for the Sky
High-rise heroes
Business
Sport
England draw Wales and Australia
Knowledge Economy
Fall on campus
More from BBC News

Get started with a domain
Search for a new domain
.com
Search
Go Daddy Website Builder
Fast Track to Mobile
50% off
Website Builder.
50% off
Online Store.
FREE online
bookkeeping.
outright.
SMALL BUSINESS


YAHOO!
UK
Search
'We are no longer together but it is fine'
Amazing photo of Rooney's painful isolation
giffgaff
For only £10 a month:
Unlimited texts
1GB Internet
Free calls to giffgaff
Order a Free SIM
DAILY OFFERS

Areas of Interest

Results (1/6)

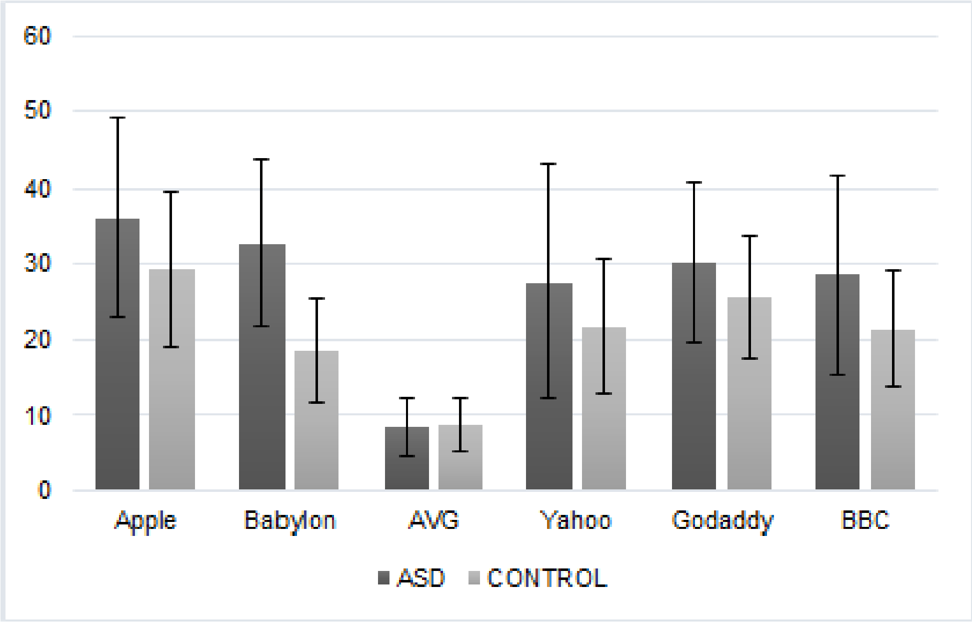
- RQ 1: Are people with autism less successful in locating the correct information or items on web pages under limited time constraints?

- Result: The ASD group was significantly less successful compared to the control group ($U = 3295.5$, $z = -3.009$, $p < 0.01$, $r = 0.22$).

Results (2/6)

- RQ 2: Do people with autism get more distracted by irrelevant elements compared to neurotypical people?
- Result: Yes

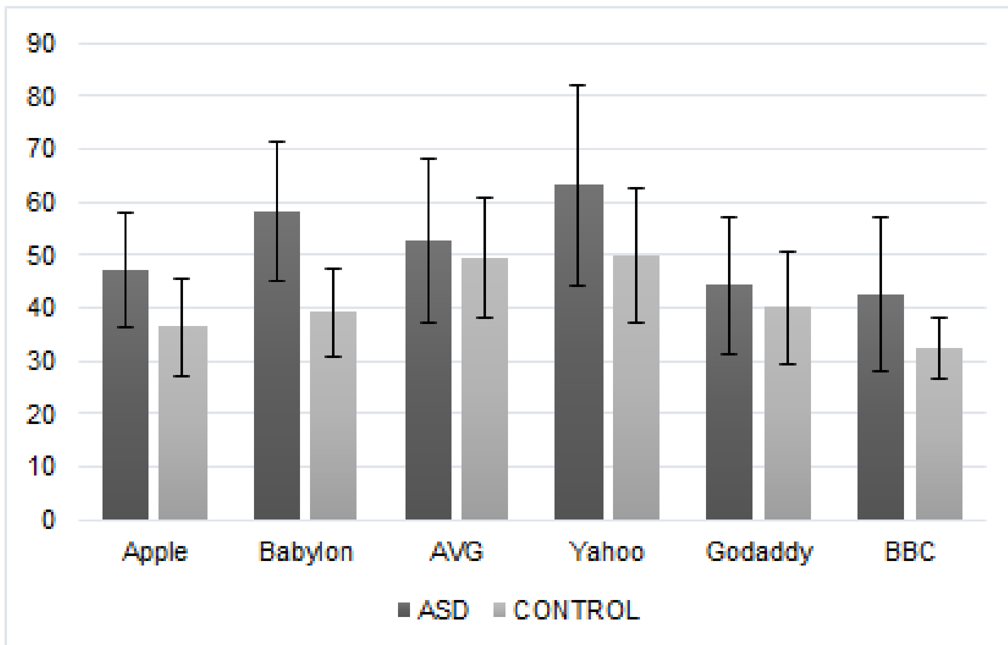
The number of irrelevant elements in the individual scanpaths of the ASD and control groups



Results (3/6)

- RQ 3: Do people with autism have longer scanpaths compared to neurotypical people?
- Result: Yes

The lengths of the individual scanpaths of the ASD and control groups

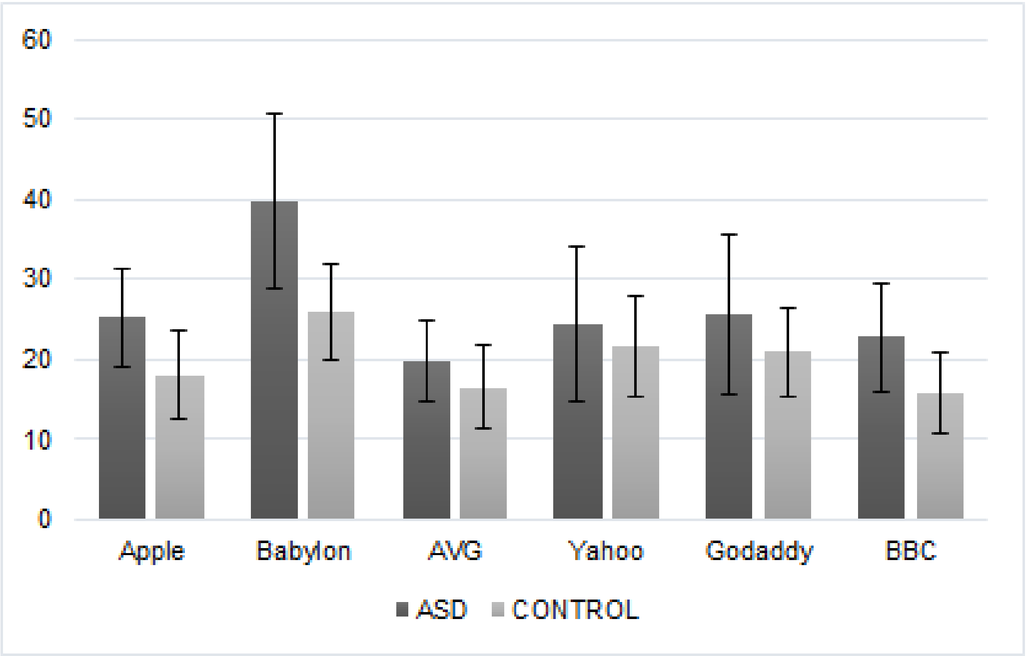


Results (4/6)

- RQ 4: Do people with autism make more transitions between the elements of web pages compared to neurotypical people?

- Result: Yes

The number of transitions made by the ASD and control groups

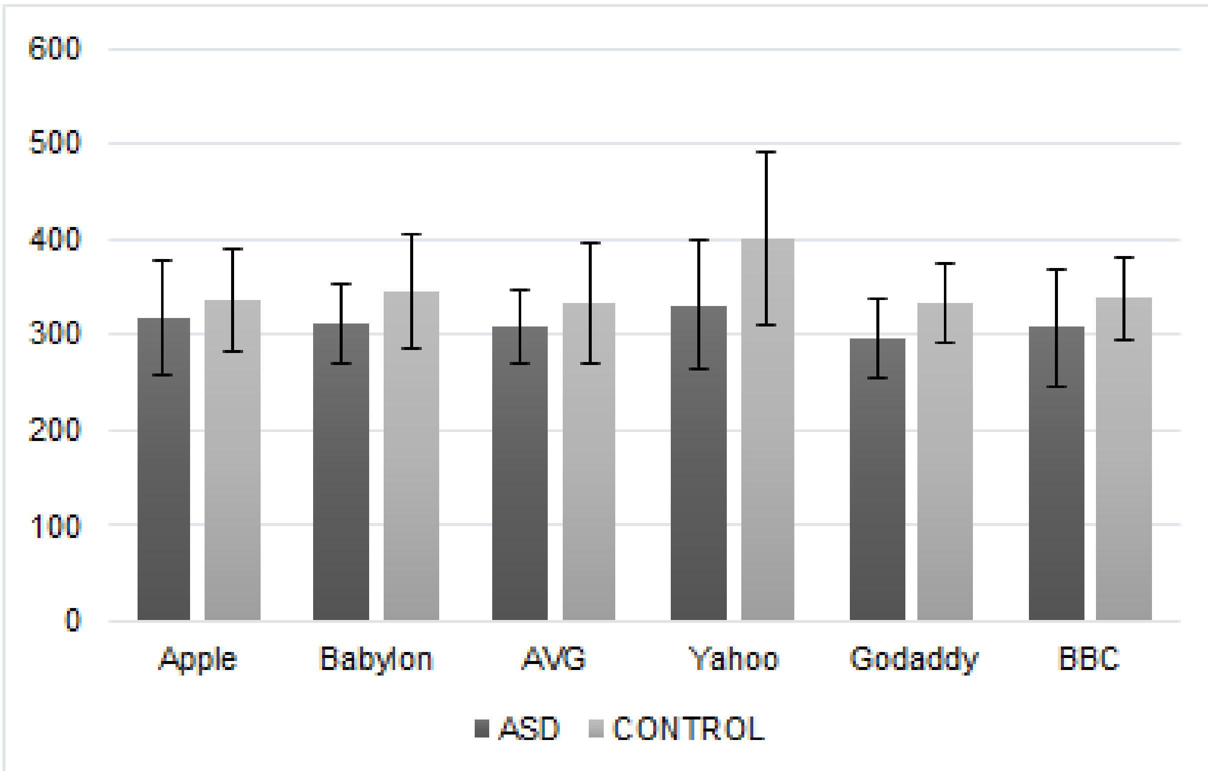


Results (5/6)

- RQ 5: Do people with autism make shorter fixations compared to neurotypical people?

- Result: Yes

The fixation durations of the ASD and control groups in milliseconds



Results (6/6)

- RQ 6: Is there a difference between the trending scanpaths of people with autism and neurotypical people on web pages?

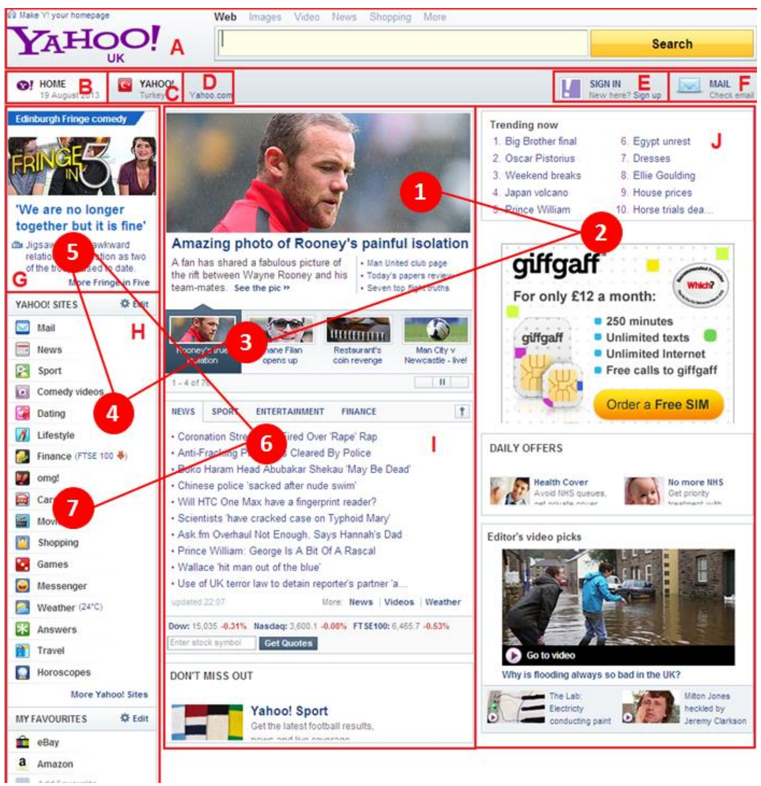
- Result:

  - 45% similarity between the trending scanpaths

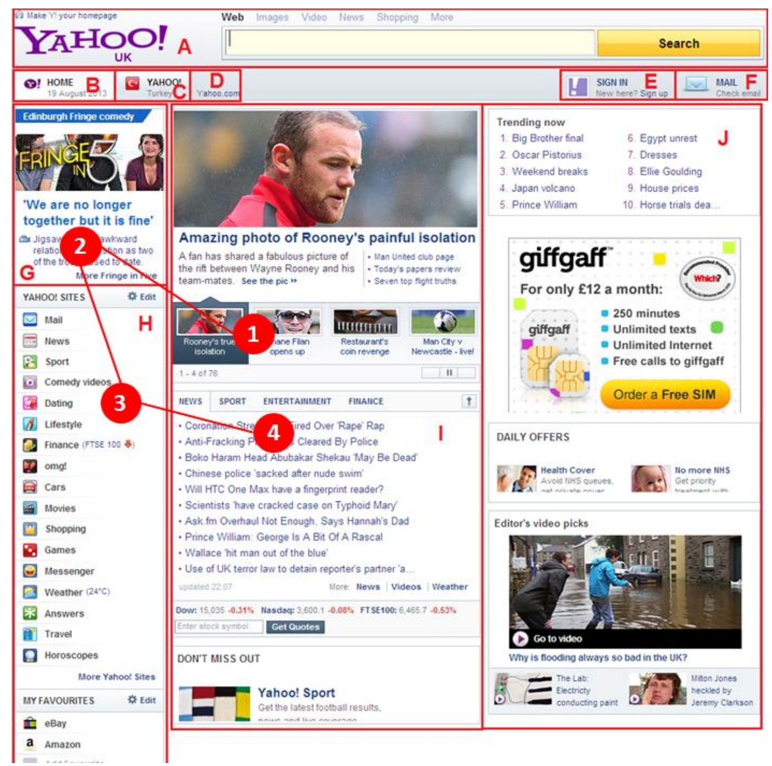
  - Greater variance within the ASD group

Trending scanpaths for the Yahoo! page

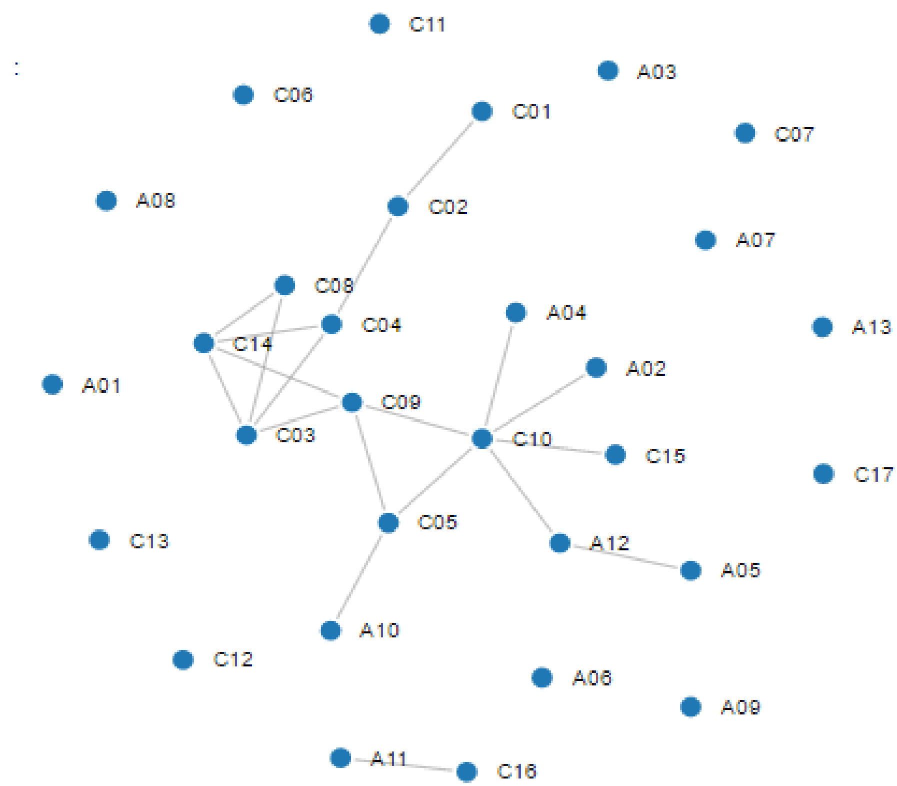
ASD group



Control group



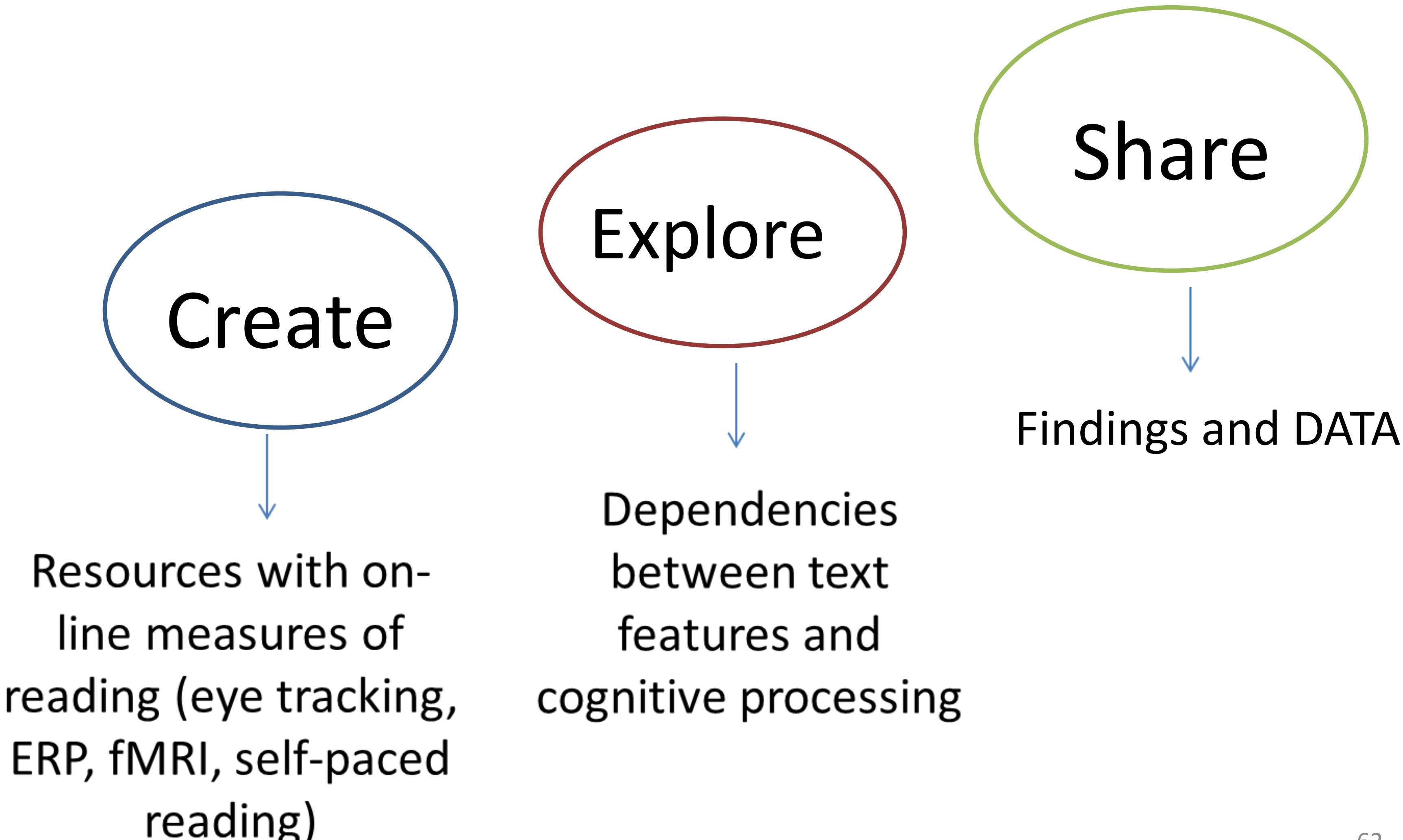
Scanpath similarity between people with autism (A) and control group participants (C)



Impact

- First empirical proof that adults with high-functioning autism have barriers to accessing information on web pages.
- Tested the WC3 assumption that ASD users: “may not pay attention to primary content because distracted by secondary content”.
- We propose improvements to the WC3 Cognitive Accessibility User Research and existing web accessibility guidelines.

Conclusion:
Sharing is Caring


Create
Explore
Share
Resources with on-line measures of reading (eye tracking, ERP, fMRI, self-paced reading)
Dependencies between text features and cognitive processing
Findings and DATA

Thank you!

victoriayaneva.info

AUTOR: autor4autism.com

v.yaneva@wlv.ac.uk

@victoria_yaneva

TEDx “Why we all need access to meaning”
https://www.youtube.com/watch?v=5jNwceqD06g