



Eyetracking metrics reveal impaired spatial anticipation in behavioural variant frontotemporal dementia



Silvia Primativo^{a,*}, Camilla Clark^a, Keir X.X. Yong^a, Nicholas.C. Firth^b, Jennifer Nicholas^a, Daniel Alexander^b, Jason D. Warren^a, Jonathan D. Rohrer^a, Sebastian J. Crutch^a

^a Dementia Research Centre, UCL Institute of Neurology, University College London, London, UK

^b Centre for Medical Image Computing, Department of Computer Science, University College London, UK

ARTICLE INFO

Keywords:

Behavioural variant frontotemporal dementia
Spatial anticipation
Anticipatory saccades
Eye movements
Machine learning

ABSTRACT

Eyetracking technology has had limited application in the dementia field to date, with most studies attempting to discriminate syndrome subgroups on the basis of basic oculomotor functions rather than higher-order cognitive abilities. Eyetracking-based tasks may also offer opportunities to reduce or ameliorate problems associated with standard paper-and-pencil cognitive tests such as the complexity and linguistic demands of verbal test instructions, and the problems of tiredness and attention associated with lengthy tasks that generate few data points at a slow rate. In the present paper we adapted the Brixton spatial anticipation test to a computerized instruction-less version where oculomotor metrics, rather than overt verbal responses, were taken into account as indicators of high level cognitive functions. Twelve bvFTD (in whom spatial anticipation deficits were expected), six SD patients (in whom deficits were predicted to be less frequent) and 38 healthy controls were presented with a 10×7 matrix of white circles. During each trial ($N = 24$) a black dot moved across seven positions on the screen, following 12 different patterns. Participants' eye movements were recorded. Frequentist statistical analysis of standard eye movement metrics were complemented by a Bayesian machine learning (ML) approach in which raw eyetracking time series datasets were examined to explore the ability to discriminate diagnostic group performance not only on the overall performance but also on individual trials. The original pen and paper Brixton test identified a spatial anticipation deficit in 7/12 (58%) of bvFTD and in 2/6 (33%) of SD patients. The eyetracking frequentist approach reported the deficit in 11/12 (92%) of bvFTD and in none (0%) of the SD patients. The machine learning approach had the main advantage of identifying significant differences from controls in 24/24 individual trials for bvFTD patients and in only 12/24 for SD patients. Results indicate that the fine grained rich datasets obtained from eyetracking metrics can inform us about high level cognitive functions in dementia, such as spatial anticipation. The ML approach can help identify conditions where subtle deficits are present and, potentially, contribute to test optimisation and the reduction of testing times. The absence of instructions also favoured a better distinction between different clinical groups of patients and can help provide valuable disease-specific markers.

1. Introduction

Eye movement investigations have provided important insights into the study of neurodegenerative conditions and in the discrimination between normal aging processes and abnormal patterns associated with dementia. Eye movement pattern changes in normal aging include a reduced ability to suppress reflexive saccades, a decline in pursuit gain, increased latency, decreased degree and velocity of vergence movements and saccadic intrusions during steady fixation tasks (for a review, see Pelak, 2010). Eyetracking investigations in people with dementia have identified abnormalities in oculomotor characteristics that are

distinct from the changes seen in normal aging (e.g., Shakespeare et al., 2015).

Patients with dementia have been shown to exhibit deficits in various eye movement measures. For example, patients with Alzheimer's disease are impaired in the pro-saccade task (fixate a target appearing on the screen; Fletcher and Sharpe, 1986; Bylisma et al., 1995; Yang et al., 2011; Yang et al., 2013) and antisaccade task (look in the opposite direction to that in which the target appeared; Abel et al., 2012; Crawford et al., 2005; Shafiq-Antonacci et al., 2003).

Patients with semantic dementia (SD), mainly characterized by anomia and a single word comprehension deficit (Gorno-Tempini et al.,

* Correspondence to: Dementia Research Centre, UCL Institute of Neurology Box 16, National Hospital, 8-11 Queen Square, London WC1N 3BG, UK.
E-mail address: s.primativo@ucl.ac.uk (S. Primativo).

2011), typically show eyetracking metrics comparable to that of controls (Garbutt et al., 2007).

Few studies and inconsistent eyetracking results have been reported for patients diagnosed with the behavioural variant of frontotemporal dementia (bvFTD). bvFTD is a clinical syndrome characterized by insidious and progressive decline in interpersonal conduct, emotional control, empathy and executive functions (Mendez and Shapira, 2011; Rascovsky et al., 2011; Mendez et al., 2013; see Harciarek and Cosentino, 2013 for a review). Meyniel (2005) reported prolonged latencies of reflexive saccades but this result has not been replicated (Boxer et al., 2006). In some studies, lower velocity during the steady-state pursuit has been described (Boxer et al., 2006; Garbutt et al., 2007). However, global basic eye movement abnormalities are not considered part of the core clinical features of frontotemporal lobar degeneration syndromes, and unimpaired performance in steady fixation, pro-saccade and smooth pursuit tasks have been reported (Pelak, 2010). Here we tested the possibility of exploiting eyetracking data in the dementia field to explore and measure complex cognitive functions and not only basic oculomotor metrics. High cognitive functions such as visual attention, memory, facial recognition and so on have been investigated in neuropsychological populations by using eyetracking experiments (e.g., Pancaroglu et al., 2016; Schuh et al., 2016; Primativo et al., 2015). The application of a similar concept in the dementia field would have multiple benefits. As compared to the standard paper-and-pencil cognitive tests, oculomotor metrics can provide a better understanding of cognitive functions by reducing the language and memory confounds associated with test instructions, diminishing the frustration and fatigue associated with request overt responses, and by removing ceiling and floor effects.

Within the broader literature about cognitive change in dementia, a small number of studies showed that eyetracking measures can be used as indicators of complex cognitive functions. For example Crutcher et al. (2009) and Richmond and colleagues (Richmond et al., 2004) used a visual paired comparison task and showed that eye movement metrics such as number of fixations and fixation duration can be indicative of short term memory difficulties in a group of patients with mild cognitive impairment as compared to aged-matched controls.

In the present paper we exploited eye movement metrics to examine a specific component of executive function in bvFTD patients, namely spatial anticipation, in an adaptation of the pen and paper Brixton spatial anticipation test (Burgess and Shallice, 1997). The term *executive function* refers to a range of functions involved in complex cognition, involving the ability to initiate, inhibit, plan, and switch behaviour in the light of new information, but also to generate strategies to accomplish complex actions. Aspects of executive dysfunction are among the earliest and most prominent features of bvFTD (Rascovsky et al., 2011; Snowden et al., 2003). Within the executive function domain, the Brixton spatial anticipation test (Burgess and Shallice, 1997) measures spatial anticipation and assesses a person's ability to detect a rule, to follow it, and to switch to a new rule. It has been shown to be very sensitive and impaired specifically in patients with frontal lesions (Burgess and Shallice, 1996; Reverberi et al., 2005; Vordenberg et al., 2014). bvFTD patients are impaired on this test (Lough et al., 2006; Kipps et al., 2007; Hornberger et al., 2010). In particular both Lough et al. (2006) and Hornberger et al. (2010) have shown that bvFTD patients are impaired in the Brixton spatial anticipation test. Conversely, SD patients generally perform in the average range on this test (Julien et al., 2008). Unfortunately, as for many currently available cognitive tests, language plays a role in the understanding of the instructions, which are quite long and complex, and so any language impairment would influence the patients' performance. Therefore, we adapted the original Brixton spatial anticipation test, developing an eyetracking instruction-less test, with the double aim of gathering fine-grained eyetracking measures that reflect complex cognitive functions, and drastically reducing the extent to which language skills can influence participants' performance. We exploited the large and rich

eyetracking data set by using a frequentist statistical analysis of standard eye movement summary metrics complemented by a Bayesian machine learning approach in order to evaluate the eyetracking adaptation of the Brixton test.

2. Materials and methods

2.1. Participants

Patients fulfilling current consensus criteria for bvFTD ([N = 12]; Rascovsky et al., 2011) or semantic dementia ([N = 6]; Gorno-Tempini et al., 2011) and 38 age-matched healthy controls took part in the study. Criteria for bvFTD involved a progressive deterioration of behaviour and/or cognition by observation or history with three of the following symptoms being present: behavioural disinhibition, apathy, loss of sympathy or empathy, perseverative, stereotyped or compulsive/ritualistic behaviour, hyperorality and dietary changes, neuropsychological profile characterized by executive deficits with relative sparing of memory and visuospatial functions. Exclusionary criteria included: pattern of deficits being better accounted for by other non-degenerative nervous system or medical disorders; behavioural disturbance being better accounted for by a psychiatric diagnosis; biomarkers strongly indicative of Alzheimer's disease or other neurodegenerative process. Criteria for SD included: both impaired confrontation naming and single-word comprehension; at least 3 of the following other diagnostic features must be present: impaired object knowledge, surface dyslexia or dysgraphia, spared repetition, spared speech production. Moreover, imaging must show predominant anterior temporal lobe atrophy, hypoperfusion or hypometabolism.

Informed consent was obtained for all participants and the study was approved by the local research ethics committee under Declaration of Helsinki guidelines. Basic demographic and genetic information is reported in Table 1. The two groups of patients did not differ in terms of age, education, disease duration and Mini-Mental Status Examination (MMSE) scores. They were also well matched with the control group, with the exception of SD patients that were younger and had a lower education level than controls (both $p < 0.05$). No other significant differences emerged. In terms of motor symptoms only one bvFTD patient showed a mild slowness and another bvFTD patient showed mild tremor. All but one bvFTD patients and all SD patients had magnetic resonance images (MRI) scan. T1-weighted volumetric MRI were acquired on a Siemens Trio TIM 3 T scanner. MRI findings were compatible with the diagnosis and no comorbidity with other neurological conditions was reported.

All participants had a general neuropsychological assessment, which included the following standard clinical tests: Wechsler Abbreviated Scale of Intelligence, Matrices and Vocabulary sub test (WASI, Wechsler, 1999); digit span forwards and backwards (Wechsler, 1981); Verbal Fluency 'F'; Trail Making Test A and B (Tombaugh, 2004) and Graded Naming Test (McKenna and Warrington, 1980). Additionally the Hayling Sentences and the standard Brixton spatial anticipation test (Burgess and Shallice, 1997) were administered. In Table 1 the discrimination statistics for the psychometric tests are reported (X^2 Mann-Whitney U tests). The results show that the raw number of errors in the Brixton test can discriminate bvFTD patients from controls but not from SD patients. Similarly, the performance in the Hayling sentences can discriminate both group of patients from controls. Among the other tests, the WASI matrices, the backward digit span and the Trial making test discriminated between bvFTD patients and controls and between the two groups of patients. The verbal fluency test could discriminate both groups of patients from controls but not from each other.

2.2. Experimental design

The experiment design and the instructions were kept as short and

Table 1

Participants' demographic, genetic and neuropsychological characteristics. Data are mean (standard deviation) with the exception of gender and genetic mutations where absolute numbers are reported.

	bvFTD (N = 12)	SD (N = 6)	Controls (N = 38)
Demographics			
Age (years)	67.7 (6.3)	63.6 (8.1) [*]	70.4 (6.5)
Gender (M:F)	10:2	4:2	14:24
Education (years)	14.0 (2.6)	12.3 (2.9) [†]	15.5 (2.7)
Disease duration (years)	9.2 (6.0)	6.8 (2.8)	NA
MMSE	25.1 (3.5)	26 (2.0)	NA
Genetic mutation (number of individuals)			
MAPT	3	1	NA
C9orf72	3	0	NA
NOTCH3	1	0	NA
Neuropsychology			
WASI matrices	17.1 (6.6) ^{††}	25.5 (6.0)	24.2 (4.6)
WASI vocabulary	44.8 (19.8)	41.5 (14.7)	NT
Forward digit span (/12)	7.8 (2.6)	10.5 (1.4)	9.1 (1.9)
Backward digit span (/12)	5.5 (2.6) ^{††}	9.5 (2.7)	7.5 (2.4)
Verbal fluency (F)	8.3 (4.6) [†]	12.3 (3.6) [*]	16.7 (4.4)
Trails time (s) (Trial B - Trial A) [‡]	129.6 (78.9) ^{††}	57.8 (30.3)	49.1 (27.1)
Graded Naming Test	11.9 (9.1) [†]	1.5 (3.2)	NT
Hayling Sentences - Overall scaled score	4.0 (1.2) [†]	3.7 (1.5) [*]	6.4 (1.5)
Brixton test (raw number of errors)	28.3 (13.4) [*]	22.5 (8.3)	18.7 (6.9)

Note: Since a larger proportion of females was observed in the control group, we tested for differences in the background tests between males and females using two-tailed *t*-tests. Significant differences emerged only for the digit span test with males reporting longer series (8.6 vs. 6.9; *p* = 0.04) and for the number of errors made in the Brixton test whereby males made less errors than females (15.6 vs. 20.5; *p* = 0.04).

[‡] 1 control has not been tested and 2 bvFTD patients were unable to complete the task.

^{*} Significant difference from controls (Mann-Whitney *U* tests).

[†] Significant difference between bvFTD and SD patients (Mann-Whitney *U* tests).

simple as possible in order to maximise the possibility of obtaining a spatial anticipation measurement with no or very little linguistic influence. The participant's task was to press a button on a joypad as soon as they saw the target (a black filled-in circle) appearing in a different position within a 10 × 7 matrix of white circles globally subtending 46 × 29.7° of visual angle. The experiment was run on a Dell 2120 desktop computer with a 23-in. screen at a viewing distance of 60 cm. Each circle's diameter was 3.3°. The centre-to-centre distance between close circles (either on the horizontal and vertical plane) was 4°. The screen background colour was white (RGB: 255, 255, 255). Each trial or sequence started with one of the circles in the matrix becoming bold for

1000 ms with the aim of prompting the participant's gaze toward the first target of the sequence. The same circle then became completely black (filled in) for 1000 ms. The target moved in the matrix following 3 different patterns (straight line, zigzag, displaced zigzag) and in 4 possible directions (top, down, left, right) for a total of 12 different patterns. In Fig. 1 examples of the three different patterns are represented. Each of the 12 sequences was presented twice, first in a pseudorandomized order and then in the reverse order. The sequence order was kept constant for all the participants. Each sequence consisted of the target moving to seven different positions. The mean stimulus duration was 1000 ms (standard deviation = 163.3 ms, range =

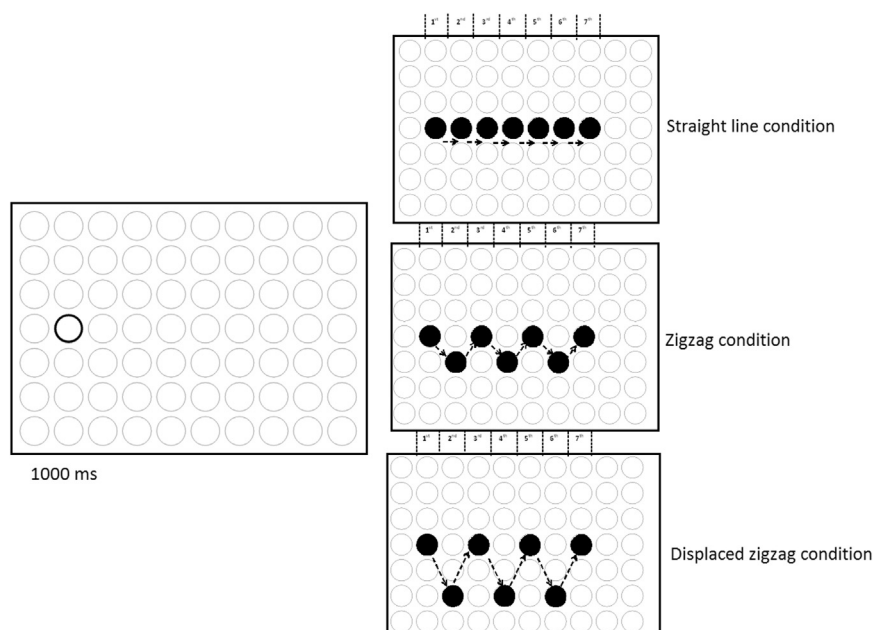


Fig. 1. On the left side a bold circle within the matrix used in the experiment is represented. The bold circle indicated the position where the first target in the sequence was going to appear in order to prompt the participant's gaze toward it before the target's onset. The bold circle duration was 1000 ms. On the right side examples of the three conditions used in the experiment are provided. In each sequence the target moved across seven spatial positions within the matrix. The target duration varied (800, 1000 and 1200 ms) and it was randomized within the sequence and kept constant for all participants. The first target in the sequence had a constant duration of 1000 ms.

800 – 1200 ms). Specifically, three different stimulus durations were chosen (800, 1000 and 1200 ms) in order to maximise the possibilities that participants actively attended the motor task rather than automatically pressing the joystick button with a fixed rhythm. For each sequence there was a pseudo-randomised pattern of durations which was kept constant for all participants. Four practice sequences were administered at the beginning of the experiment using the four possible diagonal patterns within the matrix. These sequences were not used for data analysis. The exact instruction given to participants was: “Press a button every time you see a target in a different position”.

Eye movements were recorded using a head-mounted infrared video-based eye tracker (Eyelink II; SR Research, Canada). Gaze location was recorded at 250 Hz. Participants used a chin and a head rest (wide HeadSpot; University of Houston College of Optometry) to provide stability and maintain a constant viewing distance throughout the experiment.

In order to ensure fixation stability, just before the experiment started, standard nine-point calibration and validation procedures were run. The calibration and validation target was a black dot subtending 0.5° and it was presented randomly in nine different positions on the screen. Finally, before each sequence a drift correction was performed using the same target dot used during the calibration and validation procedures, which was presented in the middle of the screen. The drift correction was used to correct for head movements but also as an inter-trial interval.

The entire procedure (calibration, validation and experiment) lasted 10–15 min for each participant with some individual differences mainly associated with the ease of set up (i.e., successfully finding the pupil position to ensure adequate calibration).

2.3. Eyetracking analysis – frequentist statistical summary metrics

Fixations and saccades were parsed by the Eyelink system, using standard velocity and acceleration thresholds ($30^\circ/s$ and $8000^\circ/s^2$). Periods during which no saccadic movement occurred were automatically identified as fixation periods. Blinks were identified and removed using Eyelink's automated blink detection. All the data were obtained from recordings with an average Cartesian prediction error of $< 1^\circ$ during the validation procedures. The frequentist statistical analysis was carried out using STATA12 software.

Basic eye movement metrics and ad-hoc metrics aiming at exploring high cognitive functions were analysed. In terms of basic oculomotor functions, we analysed the first saccade latency (i.e., time elapsed from the appearance of the target to the beginning of the saccade toward it), time to fixate the first target (i.e., the amount of time required to start fixating the first target of the sequences from the appearance of the target), mean fixation duration and the saccade velocity expressed in degrees of visual angle/sec. Time to fixate the targets (i.e., the amount of time required to start fixating each target of the sequence from the appearance of the target itself) was also measured. All these continuous measures were log transformed to reduce the skewness of the data. After the log transformation the variables of interest were normally distributed (Shapiro–Wilk test of normality, all $p > 0.05$).

Since one of the main aims of the present work was to measure spatial anticipation abilities by means of fine-grained eyetracking metrics, we measured the time to fixate the targets in the sequence, the total number of anticipatory saccades, and the proportion of correct anticipatory saccades. We classified as correct anticipatory saccades those saccades made toward the ‘forthcoming target’ according to the emerging pattern when the previous target was still displayed on the screen. Saccades with latencies lower than 80 ms were also considered anticipations (Findlay, 1981; Smit and Van Gisbergen, 1989). In the cases where the participant made a corrective saccade in the direction of the forthcoming target following a saccade of bigger or smaller amplitude, this was still considered as a correct anticipatory saccade. Finally, to further characterize the eyetracking metrics during the

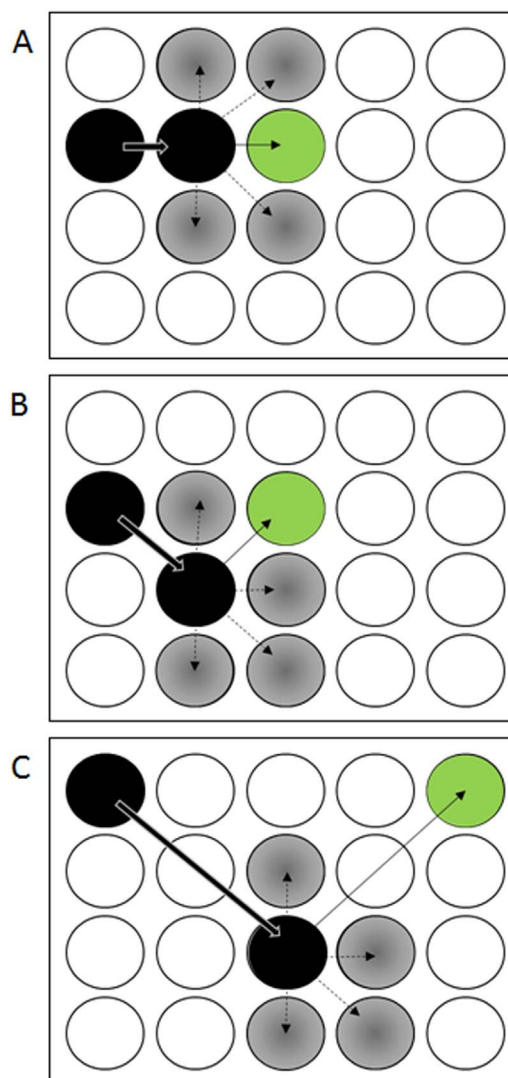


Fig. 2. Examples of correct and incorrect anticipatory saccades for the straight line (panel A), zigzag (B) and displaced zigzag (C) conditions. In each panel the ending position of correct anticipatory saccades (complete arrows) is represented by green circles, and the ending position of incorrect anticipatory saccades (dashed arrows) is represented by grey circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

spatial anticipation task, we explored all the attempts of anticipations that were not classified as “correct”, i.e., not directed toward the ‘forthcoming target’ according to the emerging pattern. Undershot and overshoot did not fall in to this category since, by definition, these would have been in the expected direction, whilst incorrect anticipatory saccades were not. We classified as *incorrect anticipatory saccades* those saccades which ended in a circle of the matrix that did not correspond to the position where the following target was going to appear made when the previous target was still displayed. In Fig. 2 we report examples of correct and incorrect anticipatory saccades for each sequence type. The sum of correct and incorrect anticipatory saccades represented the total attempts of anticipatory saccades made by participants. The classification of correct and incorrect anticipatory saccades was based on data visual inspection using the Data Viewer software (version 1.10.123).

Linear regression models with robust standard errors to adjust for repeated measures were used to analyse log transformed continuous measures (first saccade latency, time to fixate the first target, mean fixation duration, saccade velocity and time to fixate the following targets in the sequence). A logistic regression model with robust

standard errors to adjust for repeated measures was used to analyse the proportion of targets where a correct anticipatory saccade was made and the proportion of targets where an incorrect anticipatory saccade was made. Wald tests were carried out to explore main effects and interactions. The Bayesian information criterion value (BIC) is reported for the logistic regression models. For all the analysed variables we explored main effects of group and sequence type (straight lines, zigzag, displaced zigzag), and their interaction. Results on comparison of groups by individual conditions are only presented where the group and condition interaction was statistically significant. All models controlled for gender, age and education. Post-hoc testing was conducted adding WASI matrices score as a covariate; the main results were unchanged, with only minor alterations to some coefficient values and degrees of freedom.

2.4. Eyetracking analysis – machine learning approach

Raw data, from all subjects ($N = 56$) and all trials ($N = 24$ per subject), recorded from the Eyelink system was split by subject and trial number to give, 24×56 datasets. Each of these datasets consisted of the subjects detected gaze, x and y pixel coordinates, recorded at a frequency of 250 Hz. Data points were removed if a blink was detected or if the Eyelink system was unable to record the subjects gaze. To enable the prediction of the next gaze location feature extraction was performed on each patient trial dataset to give all possible series of gaze locations of length n , with the corresponding next gaze location.

For each trial control subjects were split into two sets; training and test, and the training set was then subdivided into validation and holdout sets. Each training test split was constructed subject-wise (Saeb et al., 2016), the training set was constructed of 67% ($n = 26$) of the control subjects for each trial and the remaining 33% ($n = 12$) subjects were used as a test set. The training set subdivision for each trial was constructed record-wise (Saeb et al., 2016). Features were extracted for each subject resulting in $\sim 53,000$ samples for each trial, these samples were then split 8:2 into a validation set and a holdout set. The construction of a subject-wise split test set allows for a more realistic interpretation of the error in the trained models (Saeb et al., 2016). Data from SD and bvFTD individuals were used only for testing, ensuring that the trained models have been built to predict gaze locations of only control subjects.

A Bayesian ridge regression model was built for both coordinate axes for each trial (MacKay, 1992). These models were trained using the previous n time steps from the corresponding coordinate axis. Bayesian ridge regression was used due to the improved robustness to collinearity over other regression methods. The parameter n and the hyperparameters used in Bayesian ridge regression (MacKay, 1992), for each trial were selected by performing a 5-fold cross-validated grid search on the validation set, consisting only of control subjects. Models were then retrained using the best parameters from the grid search on the entire validation set. Holdout error was then reported using holdout set. Data analysis, Bayesian ridge regression, random splitting of datasets and cross-validation has been performed using Python and scikit-learn (Pedregosa et al., 2011).

At test time, error for each time step has been calculated using the Euclidean distance between the actual gaze location and the predicted gaze location. Error for a subject over a trial is reported as the median error of all the time steps. Mann-Whitney U tests were used to compare samples to the control subject samples.

2.5. Neuroimaging analysis

Voxel-based morphometry (VBM) was performed using SPM12 software (Statistical Parametric Mapping, Version 12; <http://www.fil.ion.ucl.ac.uk/spm>) running on MATLAB R2012a (<http://www.mathworks.com>). Images were rigidly orientated to standard Montreal Neurological Institute (MNI) space using the 'New segment' function in

SPM12. Rigidly-orientated scans were segmented into grey matter, white matter and CSF. The Dartel toolbox (Ashburner, 2007) was used to perform spatial normalization, first aligning grey matter and white matter segmentations to their group-wise average (Ashburner and Friston, 2009), then combining this transformation with an affine mapping to MNI space. Normalized segmentations were modulated to preserve native-space tissue volumes and smoothed with a 6 mm full-width at half-maximum Gaussian kernel. A group-wise custom template in MNI space was created by arithmetically averaging the Dartel-normalized bias-corrected MP-RAGE images of all patients. The association between regional grey matter volume and eyetracking metrics was assessed using voxel-wise linear regression models in SPM12. Total intracranial volume, age, gender and group were included as covariates. A whole-brain grey matter mask was defined to include voxels for which the intensity was > 0.1 in at least 80% of the images; this has been shown to be appropriate for participants with greater atrophy (Ridgway et al., 2009). A voxel-wise statistical threshold of $p < 0.05$, family-wise error corrected for multiple comparisons was applied. Statistical parametric maps were overlaid on the custom template.

3. Results

3.1. Frequentist statistical summary metrics

3.1.1. Basic oculomotor function

First saccade latency. This measure was defined as the time elapsed from the appearance of the bold circle indicating the position where the first target in the sequence was going to appear to the beginning of the saccade toward it (see Fig. 1). No differences emerged between the different groups (244.6, 245.8 and 248.3 ms for controls, SD and bvFTD patients respectively, all $p > 0.1$) nor between the different sequence types (245.3, 245.1 and 246.1 ms for the straight lines, zigzag and displaced zigzag conditions, respectively, all $p > 0.1$).

Time to fixate the first target. This measure represents the amount of time required by participants to start fixating the first target of the sequence after its appearance. The three groups of participants did not differ from each other (275.9, 279.2 and 307.8 ms for controls, SD and bvFTD patients, respectively, all $p > 0.1$). No difference emerged between the different sequence types (282.7, 280.5 and 285.8 ms for the straight lines, zigzag and displaced zigzag conditions, respectively, all $p > 0.1$).

Mean fixation duration. The main effect of group was not statistically significant ($F(2, 55) = 2.89$; $p > 0.05$) suggesting that the three groups of participants had similar fixation durations (337, 373 and 323 ms for controls, SD and bvFTD, respectively).

Saccade velocity (degrees of visual angle/se). The three groups of participants were also not different in terms of saccade velocity ($F(2, 55) = 0.81$, $p = 0.4$). Saccade velocities for the three groups were 116, 145 and 101 deg/s for controls, SD and bvFTD patients, respectively.

Time to fixate the targets. This measure represents the average amount of time taken to fixate each target of the sequence (with the exception of the first one) after their appearance. Negative values represent anticipations. Results are reported in Fig. 3. When taking into account the time necessary to fixate all the targets within the sequence where the participant's gaze was already prompted, a significant group by sequence type interaction emerged [$F(11, 55) = 15.5$; $p < 0.0001$; Root MSE = 0.36]. Although bvFTD patients were systematically slower than controls and SD patients in fixating the target, a formal statistically significant difference between SD and bvFTD was only observed in the displaced zigzag condition ($p = 0.009$).

A significant main effect of target position emerged [$F(7, 55) =$

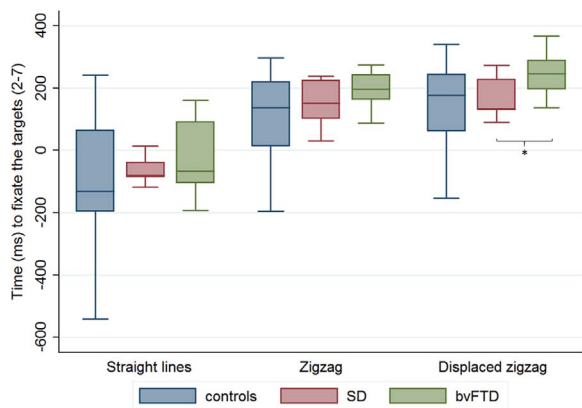


Fig. 3. Time (in milliseconds) necessary to fixate the targets in the three types of pattern. Bars represent minimum and maximum values. Asterisk (*) represent significant difference.

120.8, $p < 0.0001$] indicating that the time to fixate target 2 was longer as compared to all the other targets in the sequence (all $p < 0.0001$). Time to fixate target 3 was also shorter as compared to targets 4 and 5 (but not 6 and 7); time to fixate target 4 was shorter as compared to targets 5 and 7. Altogether, results indicate a progressive reduction of time to fixate the targets presented later in the sequence, in accordance with a clearer definition of the sequence pattern.

3.1.2. Total attempts of anticipatory saccades

The three groups of participants did not differ from each other (all $p > 0.1$) in terms of total number of anticipatory saccades (sum of correct and incorrect anticipatory saccades, see below).

3.1.3. Correct anticipatory saccades

The analysis of the proportion of correct anticipatory saccades was carried out twice: first considering all the targets in the sequence (targets 2–7) and second considering only the last part of the sequence (targets 4–7) in order to better evaluate anticipations made once the pattern had emerged clearly.

When considering correct anticipations of all targets within the sequences (Fig. 4, Panel A), a significant group by sequence type interaction emerged ($X^2_{(8)} = 214.5$, $p < 0.0001$; BIC = -755.2) suggesting that while in the straight line conditions the three groups of participants made a similar proportion of correct anticipatory saccades, in the zigzag and displaced zigzag conditions bvFTD produced a smaller proportion as compared to controls (both $p < 0.001$). None of the other pairwise comparisons reached statistical significance.

Similar results emerged when considering the last part of the sequences, when the pattern sequence was assumed to be evident (targets 4–7; see Fig. 4 Panel B). Indeed, there was a statistically significant interaction group by sequence type ($X^2_{(8)} = 135.4$, $p < 0.0001$; BIC = -675.7) suggesting that in the straight line conditions the three groups of participants made a similar proportion of correct anticipatory saccades, while in the zigzag and displaced zigzag conditions bvFTD produced a smaller proportion as compared to controls (both $p = 0.02$). None of the other pairwise comparisons reached statistical significance.

3.1.4. Incorrect anticipatory saccades

We analysed the proportion of incorrect anticipatory saccades twice: over the entire sequence (targets 2–7) and on the last 4 targets only (targets 4–7).

The proportion of incorrect anticipatory saccades made by each group in each condition is shown in Fig. 5 (panel A: – targets 2–7; panel B: targets 4–7). A significant group by sequence type interaction emerged ($X^2_{(8)} = 274.9$, $p < 0.0001$; BIC = -787.6). The interaction indicated that bvFTD made more incorrect anticipatory saccades than

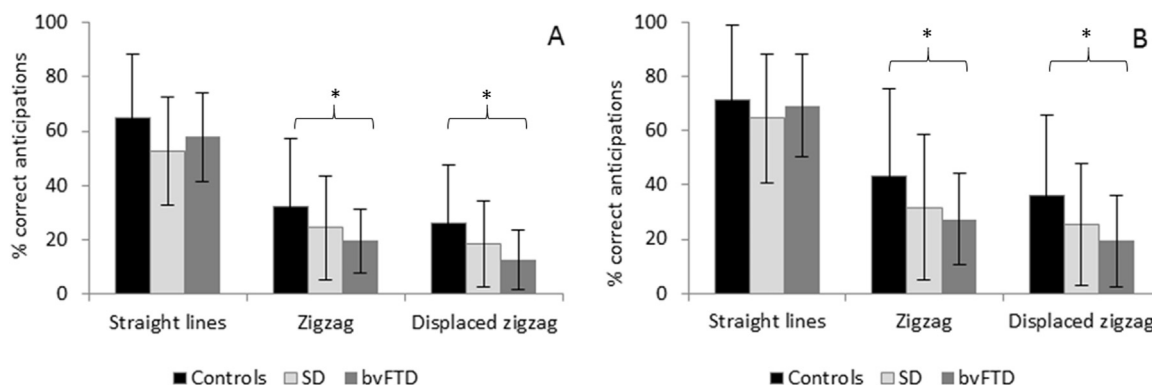


Fig. 4. Proportion of correct anticipatory saccades when considering (A) all targets in the sequence and (B) 4–7 targets only. Bars represent standard deviations. Asterisks (*) represent significant differences.

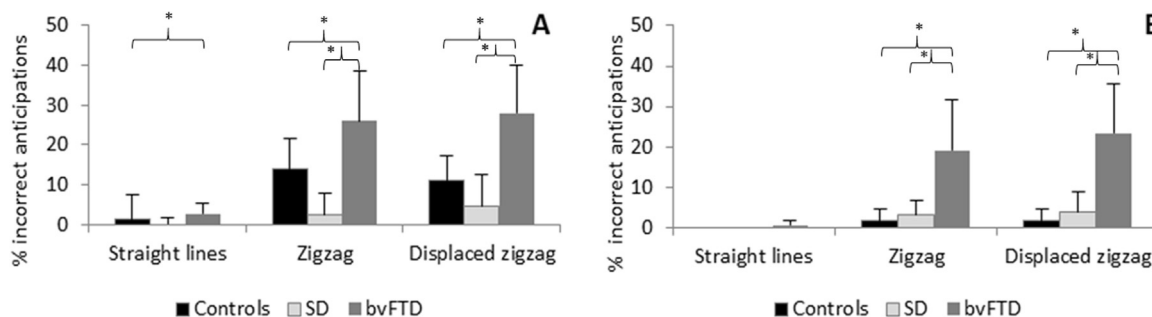


Fig. 5. Proportion of incorrect anticipatory saccades when considering (A) all targets in the sequence and (B) 4–7 targets only. Bars represent standard deviations. Asterisks (*) represent significant differences.

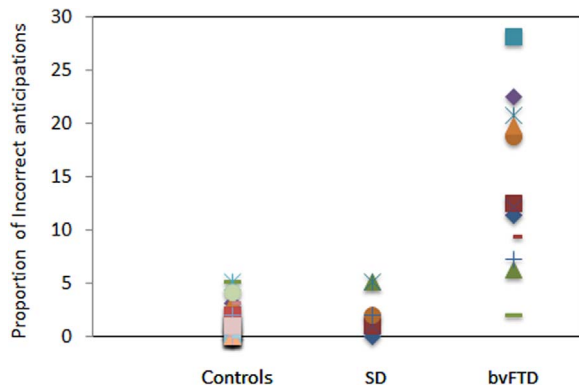


Fig. 6. Individual results for the proportion of incorrect anticipatory saccades when considering 4–7 targets only in the sequences. For each group, different symbols indicate different participants.

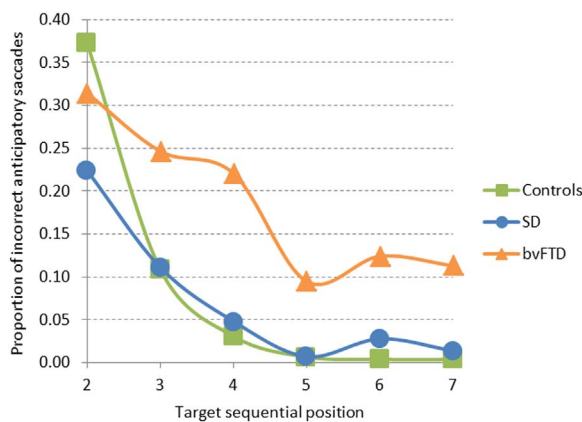


Fig. 7. Proportion of incorrect anticipatory saccades over the sequential target position for the three groups of participants across all conditions (straight lines, zigzag, displaced zigzag).

controls in the straight line condition ($z = 2.7, p = 0.008$), zigzag ($z = 4.9, p < 0.0001$) and displaced zigzag ($z = 6.4, p < 0.0001$) conditions. bvFTD patients also made more incorrect anticipatory saccades than SD patients in the zigzag ($z = -3.1, p = 0.002$) and displaced zigzag conditions ($z = -2.5, p = 0.01$). Conversely SD patients produced a similar proportion of incorrect anticipatory saccades compared with controls in all conditions (all $ps > 0.1$).

We repeated the analysis on the last part of the sequences. A significant group by sequence type interaction emerged ($\chi^2_{(8)} = 13854, p < 0.0001; BIC = -782.2$). The interaction indicated that the straight line condition did not yield differences among the three groups of participants (all $p > 0.1$) but in the zigzag and displaced zigzag conditions bvFTD patients systematically made more incorrect anticipatory saccades as compared to both controls ($z = 9.7, p < 0.0001$ for both sequence types) and SD patients ($z = -4.9, p < 0.0001$ for both sequence types). The SD patients did not differ from controls in terms of incorrect anticipations in any sequence type (all $p > 0.1$). In Fig. 6 individual results for the proportion of incorrect anticipatory saccades on the last part of the sequence are reported in order to highlight the consistency of the described pattern.

3.1.5. Sequential position and qualitative error analysis

The proportions of anticipatory saccades across sequential target positions for the 3 groups of participants are shown in Fig. 7. From the figure it emerges that while all three groups of participants made some

incorrect anticipatory saccades at the beginning of the sequences, for controls and SD patients the proportion was drastically reduced by the third position of the target. Conversely, bvFTD patients continued to generate a large number of incorrect anticipatory saccades until the target reached the 5th position and even beyond, since a small proportion of individuals kept making errors throughout the entire sequence.

In terms of a qualitative characterization of the incorrect anticipatory saccades in the bvFTD group of patients, 94.3% of patients followed one of the patterns used in the paradigm (i.e., straight lines, normal and displaced zigzag) with only a small percentage (5.7%) being on random circles within the matrix. Since the incorrect anticipatory saccades made by controls virtually always (97%) resembled one of the patterns used in the paradigm, the anticipations made by bvFTD patients on random circles in the matrix can be considered similar to the “bizarre errors” reported in the original Brixton paper. Within the non-random saccades, the large majority of them (63.1%) were directed toward the closest circle following the straight line pattern, either in the horizontal or vertical direction. A progressively smaller proportion of incorrect saccades followed a zigzag (22.4%), diagonal (13.6%) and displaced zigzag (1%) pattern (see examples represented in Fig. 2). Moreover, it is worth noting that a large proportion of such incorrect anticipatory saccades (31%) seemed to indicate a perseveration from the preceding sequence. This result is similar to what was found in the original Brixton study, where a relatively high proportion of perseverative errors (over 21%) was also reported (Burgess and Shallice, 1996).

3.2. Machine learning approach – Bayesian ridge regression

Holdout error was consistently low across each of the trials (0.57 ± 0.08 pixels), as reported in Table 2, left hand side, indicating that each model was capable of accurately predicting the subjects’ next gaze location. The number of steps selected by cross-validation in each trial’s model varied from one to 41, though for 20 of the 24 trials the number of steps used was greater than one. As the number of steps used in the model is variable across trials, no direct comparison can be made of the coefficients from the ridge regression. However, the coefficient of previous gaze point was consistently high compared to the sum of all the coefficients for each trial ($71.54 \pm 18.64\%$), highlighting that, as we might expect, the previous gaze point was the most important factor in predicting the next gaze point. The frequency of dot movements during the trial is over an order of magnitude lower than the frequency of the gaze prediction modelling during trials. This shows that though the modelling has no concept of the patterns that are being followed during the trial, it has in fact learnt the short-term behaviour of control subjects.

Analysis of the group wise differences for each of the trials was performed using the error for each subject over each trial (median of each time step), this showed a significant difference ($p < 0.05$) between controls and bvFTD for all of the 24 trials analysed. Moreover, the percentage difference in errors between bvFTD subjects and control subjects was consistently high across all the trials (average difference in errors = $40.0 \pm 9.32\%$). Conversely the difference between SD patients and controls was only significant ($p < 0.05$) in 12 of the 24 trials, and the percentage difference in errors between control and SD subjects was variable across the trials (average difference in errors = $36.2 \pm 17.2\%$).

Error maps, for three exemplar cases are shown in Fig. 8. These maps have been generated by plotting the location of subject’s gaze, colored by the \log_{10} error in the model prediction. Error maps highlight an individual’s path throughout trials and those locations where the model is making poor predictions. These maps have been generated from subjects in the test set, from trial 2, the first zigzag trial in the experiment. This trial was chosen as it best highlights the differences between groups. The example control subject exhibited a linear path

Table 2
 For trial type, codes correspond to the pattern used: straight lines (S), zigzag (Z) and displaced zigzag (DZ), before the dash and direction; top (T), bottom (B), left (L), and right (R), after the dash. Significance is highlighted; $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***), using the Mann-Whitney U test comparing samples to the control subject samples. *Left hand side*: Summary of results from gaze location modelling. Steps used corresponds to the number of previous gaze locations used in each model to predict the next gaze location. Holdout error is the median error across each of the training examples, and control, bvFTD and SD error correspond to the average median errors from each subject from the corresponding group. *Right hand side*: Summary of results from incorrect anticipation analysis. Values represent the percentage of mean incorrect anticipations (\pm Coefficient of Variation).

Trial Number	Trial type	Steps Used	Holdout Error	Machine learning approach				Summary metrics			
				Control Error (N = 38)	bvFTD Error (N = 12)	SD Error (N = 6)	Controls (N = 38)	bvFTD (N = 12)	SD (N = 6)		
1	S-R	41	0.49	0.49 ($\pm 0.17\%$)	0.75 ($\pm 0.26\%$)**	0.83 ($\pm 0.39\%$)*	0 ($\pm 0\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)	
2	Z-L	28	0.53	0.5 ($\pm 0.29\%$)	0.78 ($\pm 0.27\%$)**	0.79 ($\pm 0.34\%$)*	2.08 ($\pm 331.66\%$)	22.92 ($\pm 113.18\%$)*	8.33 ($\pm 141.42\%$)		
3	DZ-B	1	0.54	0.53 ($\pm 0.18\%$)	0.76 ($\pm 0.23\%$)**	0.67 ($\pm 0.27\%$)*	4.17 ($\pm 223.61\%$)	4.17 ($\pm 223.61\%$)	0 ($\pm 0\%$)		
4	S-T	2	0.48	0.55 ($\pm 0.34\%$)	0.7 ($\pm 0.23\%$)*	0.67 ($\pm 0.39\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)		
5	Z-R	41	0.54	0.58 ($\pm 0.31\%$)	0.78 ($\pm 0.29\%$)**	0.78 ($\pm 0.35\%$)	2.08 ($\pm 331.66\%$)	29.86 ($\pm 95.63\%$)**	4.17 ($\pm 223.61\%$)		
6	DZ-L	15	0.6	0.56 ($\pm 0.3\%$)	0.82 ($\pm 0.21\%$)**	0.79 ($\pm 0.33\%$)*	2.08 ($\pm 331.66\%$)	25 ($\pm 129.1\%$)*	8.33 ($\pm 141.42\%$)		
7	S-B	42	0.76	0.88 ($\pm 0.25\%$)	1.21 ($\pm 0.27\%$)**	1.05 ($\pm 0.34\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)		
8	Z-T	38	0.52	0.65 ($\pm 0.31\%$)	0.81 ($\pm 0.17\%$)*	0.74 ($\pm 0.33\%$)	4.17 ($\pm 223.61\%$)	30.56 ($\pm 78.99\%$)**	0 ($\pm 0\%$)		
9	DZ-R	15	0.55	0.62 ($\pm 0.29\%$)	0.83 ($\pm 0.25\%$)**	0.79 ($\pm 0.33\%$)	2.08 ($\pm 331.66\%$)	18.75 ($\pm 110.55\%$)*	4.17 ($\pm 223.61\%$)		
10	S-L	15	0.47	0.53 ($\pm 0.19\%$)	0.76 ($\pm 0.28\%$)**	0.78 ($\pm 0.40\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)		
11	Z-B	1	0.5	0.56 ($\pm 0.3\%$)	0.77 ($\pm 0.24\%$)**	0.67 ($\pm 0.28\%$)*	0 ($\pm 0\%$)	4.17 ($\pm 331.66\%$)	0 ($\pm 0\%$)		
12	DZ-T	1	0.53	0.62 ($\pm 0.24\%$)	0.79 ($\pm 0.27\%$)*	0.68 ($\pm 0.23\%$)	0 ($\pm 0\%$)	34.72 ($\pm 83.04\%$)**	0 ($\pm 0\%$)		
13	DZ-T	6	0.57	0.59 ($\pm 0.15\%$)	0.83 ($\pm 0.24\%$)**	0.76 ($\pm 0.31\%$)*	2.08 ($\pm 331.66\%$)	20.83 ($\pm 118.32\%$)*	8.33 ($\pm 141.42\%$)		
14	Z-B	1	0.52	0.5 ($\pm 0.29\%$)	0.72 ($\pm 0.17\%$)**	0.63 ($\pm 0.28\%$)	0 ($\pm 0\%$)	22.92 ($\pm 129.53\%$)**	0 ($\pm 0\%$)		
15	S-L	38	0.49	0.54 ($\pm 0.4\%$)	0.8 ($\pm 0.21\%$)**	0.71 ($\pm 0.37\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)		
16	DZ-R	6	0.58	0.52 ($\pm 0.24\%$)	0.82 ($\pm 0.27\%$)**	0.8 ($\pm 0.32\%$)*	2.08 ($\pm 331.66\%$)	26.39 ($\pm 115.19\%$)*	4.17 ($\pm 223.61\%$)		
17	Z-T	15	0.51	0.62 ($\pm 0.34\%$)	0.8 ($\pm 0.26\%$)**	0.79 ($\pm 0.38\%$)	4.17 ($\pm 223.61\%$)	16.67 ($\pm 141.42\%$)	0 ($\pm 0\%$)		
18	S-B	15	0.49	0.47 ($\pm 0.24\%$)	0.7 ($\pm 0.17\%$)**	0.68 ($\pm 0.35\%$)*	0 ($\pm 0\%$)	2.08 ($\pm 331.66\%$)	0 ($\pm 0\%$)		
19	DZ-L	42	0.58	0.62 ($\pm 0.29\%$)	0.87 ($\pm 0.21\%$)**	0.84 ($\pm 0.39\%$)	2.08 ($\pm 331.66\%$)	35.42 ($\pm 83.81\%$)**	8.33 ($\pm 223.61\%$)		
20	Z-R	15	0.54	0.57 ($\pm 0.29\%$)	0.78 ($\pm 0.25\%$)**	0.77 ($\pm 0.29\%$)*	0 ($\pm 0\%$)	10.42 ($\pm 153.62\%$)*	0 ($\pm 0\%$)		
21	S-T	15	0.53	0.55 ($\pm 0.24\%$)	0.77 ($\pm 0.16\%$)**	0.69 ($\pm 0.38\%$)	0 ($\pm 0\%$)	2.08 ($\pm 331.66\%$)	0 ($\pm 0\%$)		
22	DZ-B	15	0.59	0.54 ($\pm 0.18\%$)	0.78 ($\pm 0.21\%$)**	0.77 ($\pm 0.33\%$)*	0 ($\pm 0\%$)	20.83 ($\pm 145.6\%$)**	4.17 ($\pm 223.61\%$)		
23	Z-L	23	0.52	0.61 ($\pm 0.32\%$)	0.76 ($\pm 0.24\%$)*	0.91 ($\pm 0.58\%$)*	4.17 ($\pm 331.66\%$)	14.58 ($\pm 109.73\%$)*	8.33 ($\pm 223.61\%$)		
24	S-R	15	0.5	0.55 ($\pm 0.39\%$)	0.82 ($\pm 0.18\%$)**	1 ($\pm 0.6\%$)*	0 ($\pm 0\%$)	0 ($\pm 0\%$)	0 ($\pm 0\%$)		

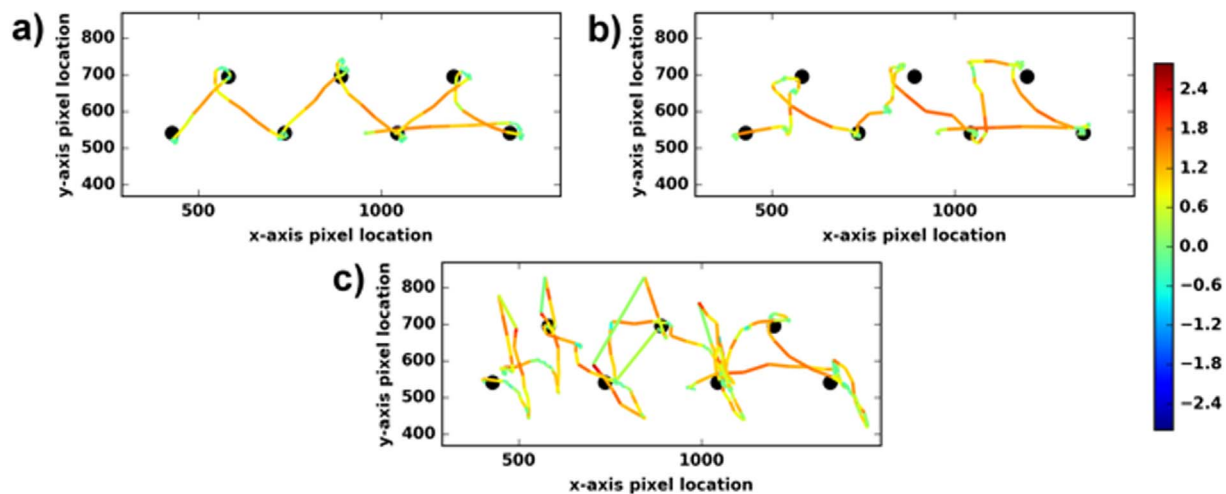


Fig. 8. Large diagonal trial (7) error maps for three exemplar subjects. Subjects: a) control, b) SD, and c) bvFTD gaze shown by a colored line, and the dot location highlighted by black dots. Line colour represents the \log_{10} error from the predicted gaze location and the actual gaze location. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between dots with low prediction error points corresponding to fixation around the dots (Fig. 8, Panel A) which led to a low median error over all points (0.62). An exemplar SD patient (Fig. 8, Panel B) showed a similar pattern to that observed in the control subject, with linear paths between dots with some fixation around the dots. However, the SD subject had a higher median error over all the points (0.94) and less precise fixation on the dots. In comparison to the control example the bvFTD subject (Fig. 8, Panel C) showed a much more erratic path, but still had fixations on the targets, which can be seen by the low prediction error on or near the targets. This suggests that either the subject was unable to maintain a prolonged fixation or the subject was making incorrect anticipations for the location of the next target. We can exclude the first possibility since we have shown that both groups of patients have fixation durations completely comparable to those of controls. The median error for this subject was the highest observed (1.28) of the exemplar subjects presented, again highlighting the differences between this bvFTD subject and the control subjects that the model was built on. The three exemplar cases shown are broadly representative of their groups, however, as we might expect, we observed intra and inter group variation.

3.3. Comparison between the original Brixton test, frequentist statistical summary metrics and machine learning approach

One of the aims of the present study was to develop and evaluate a spatial anticipation test with limited verbal instructions based on eye-tracking metrics. But what is the relationship between the scores obtained with the standard Brixton test and the metrics derived from the novel approach? The performance of the three groups of participants on the Brixton test is shown in Table 1.

Overall, as reported in Table 1, in the Brixton spatial anticipation test bvFTD patients produced a larger proportion of errors (28.3) than controls and SD patients (18.7 and 22.5, $p = 0.002$ and 0.05 , respectively) and this metric significantly correlated ($p < 0.05$) with the WASI matrix reasoning scores ($r^2 = -0.6$) and performance on the digit span backward test and Trails time ($r^2 = -0.48$ and 0.55 , respectively). However, when looking at the individual results on the Brixton test, a spatial anticipation deficit emerged in only 7 out of 12 bvFTD patients (58%), whose classification scores were ‘poor’, ‘abnormal’ or ‘impaired’. The other 5 bvFTD patients were classified as ‘moderate average’ or above according to the Brixton test classification system. We used the frequentist statistical summary metrics to calculate how many individual patients failed in the new eye-tracking test, by

falling over 2.5 standard deviations from controls’ performance in terms of proportion of incorrect anticipatory saccades. The proportions were log transformed to reduce the skewness of the data. Results revealed that bvFTD patients were impaired not only as a group but also on an individual basis. Indeed, 11 out of 12 bvFTD patients (91.7%) failed (vs. 0/38 controls) the new spatial anticipation eye-tracking paradigm, including the 5 patients who were not impaired in the standard Brixton spatial anticipation test ($X^2_{(1)} = 3.6$; $p = 0.06$). All the basic eye movement measures were within ± 2.5 standard deviations from controls’ performance for each individual. It is also worth noting that 2 out of 6 SD patients (33%) failed the Brixton spatial anticipation test. The new eye-tracking metrics revealed that none of the SD patients failed in the new eye-tracking instruction-less paradigm. This apparent discrepancy suggests that the poor performance on the Brixton test for the 2 SD patients was mainly driven by a poor understanding of the task instructions which, instead, played a marginal role in the metrics derived from the new eye-tracking paradigm. Moreover, correlations between the proportion of incorrect anticipations on last part of the trials and the other neuropsychological tests have been run. Results indicated significant correlations ($p < 0.05$) with the WASI matrix reasoning scores ($r^2 = -0.53$), Trails time ($r^2 = -0.43$), and raw number of errors produced on the Brixton Test ($r^2 = 0.46$).

What advantages can the ML approach bring to the data analysis? As the resultant data from applying machine learning to these data is 24 dimensional, the definition of a simple metric to diagnose individual participants is not possible. However, the machine learning approach, being based on $\sim 53,000$ samples for each trial, can provide important information about the significant difference on each individual trial between patients and controls. In Table 2 we compared the significant differences highlighted by the ML approach (left hand side) and the frequentist summary metrics (right hand side). As can be seen in the Table, the ML approach described a significant difference in 24/24 trials for bvFTD while the frequentist summary metrics can only identify differences in 13/24 trials ($X^2 = 11.79$, $p = 0.0006$). The ML approach also identified trials where a difference between SD patients and controls was detectable (12/24) while the frequentist summary metrics approach did not (0/24) ($X^2 = 13.44$, $p = 0.0002$). The identification of trials where SD patients were impaired in the new eye-tracking spatial anticipation test has been confirmed by a visual inspection of the eye-tracking data. The fact that the ML approach can detect a deficit in bvFTD patients and, in some cases also in SD patients, even in those trials where the frequentist summary metrics approach did not find any difference with controls might be useful for optimising the task and reducing testing times.

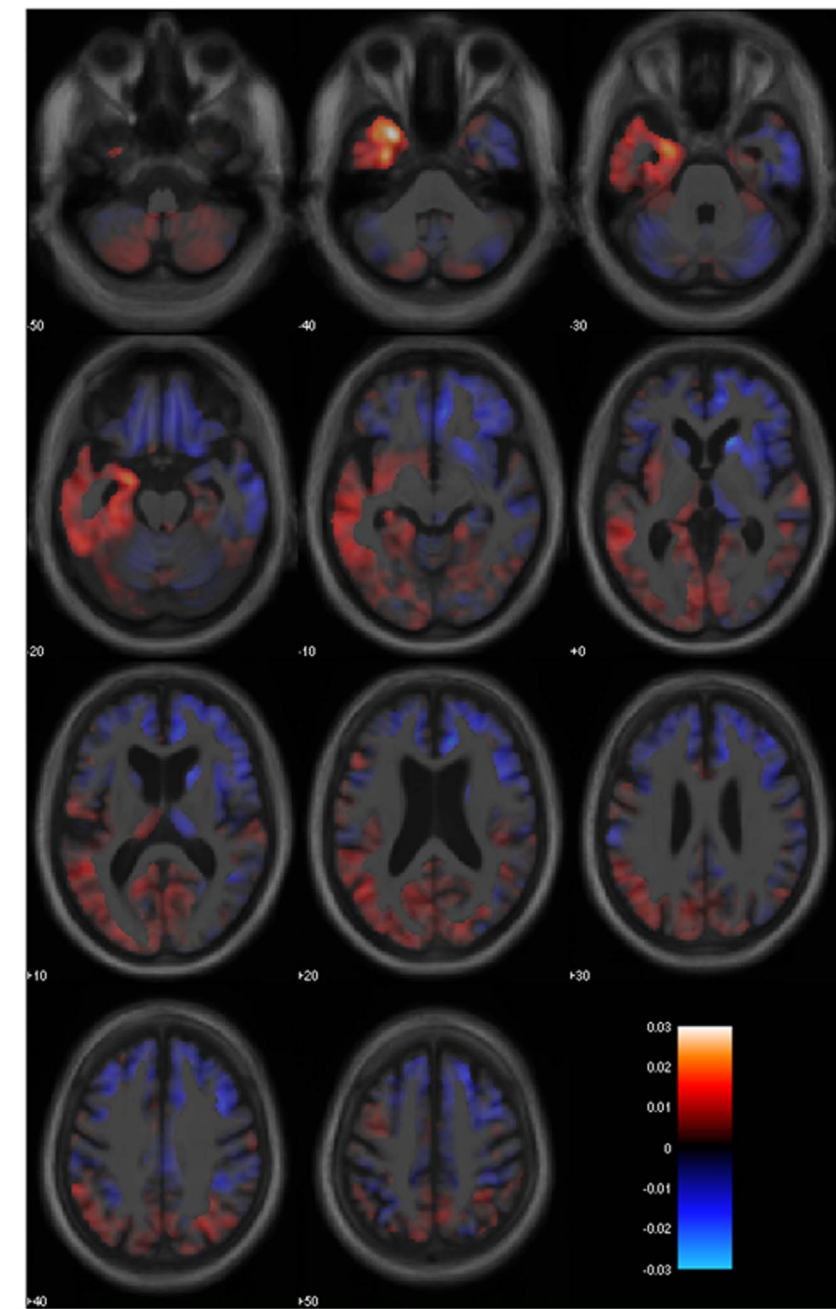


Fig. 9. Uncorrected t-values effect maps for the associations between the portion of incorrect anticipatory saccades and grey matter volume displayed on axial sections. Warmer colours indicate stronger positive associations between a larger proportion of incorrect anticipatory saccades and lower grey matter volume, with cooler colours representing the reverse contrast. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Button press reaction times

The participants' task in the present study was to press a button every time they saw a target in a different position. A linear regression model on log transformed reaction times revealed shorter RTs for the straight line condition (408.8 ms) as compared to both the zigzag (433.1 ms) and displaced zigzag conditions (439.9 ms, both $p < 0.0005$), while the difference between the zigzag and displaced zigzag conditions was not significant ($p = 0.54$). Results also indicated that bvFTD patients were slower (456.4 ms) than controls (405.9 ms, $p = 0.005$) and SD patients (412.8 ms, $p = 0.02$). There was no significant difference between controls and SD patients ($p = 0.5$).

We repeated the analysis on individual variance (i.e. the difference between each reaction time data point and the group average RT within the same condition) in order to reduce the focus on absolute reaction times and emphasise eventual individual differences. In this case the

three groups did not differ to each other [$F(2, 55) = 0.29$, $p = 0.7$]. This result suggests that the main group effect revealed by the main analysis might represent the effect of large individual variability.

3.5. Imaging results

Uncorrected p-values whole brain effect maps showing neuroanatomical associations between the proportion of incorrect spatial anticipations and grey matter volume are displayed in Fig. 9. When correcting for multiple comparisons no significant association was found. This null result might be related to the small sample size and the consequent lack of power in the analysis. A tendency toward associations with grey matter reductions in the right superior and middle frontal gyrus and the frontal pole can however be observed. The strongest association emerging from the VBM analysis, although not statistically significant, corresponds to the anatomical location of the frontal eye fields (FEF; Paus, 1996).

4. Discussion

In the present study we developed a new eyetracking spatial anticipation test adapted from the Brixton spatial anticipation test (Burgess and Shallice, 1997). The task instructions were kept as simple as possible, whereby the participant was only asked to press a button when seeing a target, in order to reduce as much as possible the confounds related to language load. The aim of the study was to evaluate whether such an adapted eyetracking test can provide a sensitive measure of a spatial anticipation deficit in bvFTD patients. New possibilities in eyetracking data analysis were explored by complementing a frequentist statistical analysis with a Bayesian machine learning approach.

Results indicated that both bvFTD and SD patients showed normal performance in terms of primary visually guided saccades, as revealed by saccade latencies, time to fixate the first target in the sequence after its onset, fixation duration and saccade velocity. Following our expectations, the test implicitly elicited anticipatory saccades to the upcoming target in controls once the pattern became clear, attesting to the possibility of gathering informative eyetracking metrics about specific executive process from this paradigm. In this context, bvFTD but not SD patients made a smaller number of correct anticipatory saccades than controls in zigzag and displaced zigzag conditions. Even more relevant is the larger number of incorrect anticipatory saccades made by bvFTD patients as compared to both controls and SD patients in the zigzag and displaced zigzag conditions. These results have also been confirmed and enhanced by a machine learning and completely data driven approach.

How can we explain the present results? Although a spatial working memory deficit might limit patients' ability to appreciate the emerging pattern, this does not satisfactorily explain the whole set of results. The qualitative analysis of the incorrect anticipatory saccades suggests an answer to this question. It could be hypothesised that bvFTD patients have a general impairment in the control of their saccadic eye movements. In this circumstance, however, randomly directed square wave jerks (Lemos and Eggenberger, 2013) would have been expected. This was not the case: although incorrect, in 94% of the cases these anticipatory saccades followed one of the patterns used in the experiment and the proportion of randomly directed saccades was very low.

Two other hypotheses may be considered which are not mutually exclusive: a rule switching deficit and an impairment of the inhibition of internal source of error. In over 30% of the cases, when bvFTD patients made an incorrect anticipatory saccade, they were perseverating following the immediately previous sequence pattern. This is compatible with the well described difficulty shown by bvFTD patients in switching from a previous rule to a new one (Thompson et al., 2005). Moreover, it is also worth considering that the largest proportion of incorrect saccades was characterized by following the simplest pattern (i.e., horizontal or vertical straight lines), regardless of whether it was consistent with the previous spatial pattern or not. This suggests that saccades were triggered on the basis of an internal representation of future target behaviour regardless of the ongoing (or immediately previous) actual spatial pattern. This interpretation (i.e., failure to inhibit an internal source of error) has also been proposed as an explanation for the incorrect anticipatory saccades observed in schizophrenic patients (Hommer et al., 1991) and in a single case report of a patient with a lesion involving the left supplementary eye field in the medial frontal cortex (Husain et al., 2003).

Our interpretation of the data is that a rule switching deficit and an impairment related to the inhibition of an internal source of error are reflected in the spatial anticipation deficit observed in our sample of bvFTD patients. Indeed, together the two mechanisms may be responsible for the production of a large number of incorrect anticipatory saccades in the bvFTD patients, whereby a rule switching deficit may have caused the expectation of a previously seen spatial pattern and the difficulty in the inhibition of an internal source of error may have led patients to systematically expect a horizontal or vertical pattern even in

those circumstances where a more complex pattern was emerging.

Our data are also compatible with the pattern of brain atrophy normally described for the behavioural variant of frontotemporal dementia. Occipital and parietal areas, usually spared in bvFTD (Seeley et al., 2009), are involved in reflexive and visually guided saccades (Johnston and Everling, 2008; Helie et al., 2013) and the present results show, indeed, that such eye movement metrics are comparable in bvFTD, SD patients and healthy controls. On the other hand frontal cortical brain areas, by definition the primary locus of atrophy in bvFTD (Seeley et al., 2009), have a determinant role in the control and execution of voluntary eye movements. Key components of the eye movement circuit are the frontal eye fields. The FEF trigger internally generated saccades, such as saccades toward targets that are not yet present (predictive saccades), no longer visible (memory-guided saccade) or located in the opposite direction (antisaccade), rather than purely reflexive saccades (Schall, 1995; Pouget, 2015). In our case, the impairment shown by bvFTD patients is characterized by both a limitation in the number of correct anticipatory [predictive] saccades and by a reduced ability to inhibit voluntary internally-guided saccades.

Coherently, it has been shown that bvFTD patients are impaired in both pro-saccade (Burrell et al., 2012) and anti-saccadic tasks, where they manifest a reflexive saccade inhibition deficit (Meyniel, 2005; Boxer et al., 2006; Burrell et al., 2012). This result has also been interpreted as an inhibition deficit and it has been linked to damage at the level of the FEF described in bvFTD patients (Burrell et al., 2012) but which is usually described as spared in SD (Chan et al., 2001; Galton et al., 2001). Altogether anatomical data coming from bvFTD and SD patients are in agreement with the differences observed in terms of oculomotor behaviour in the two groups of patients studied in the present study.

Our results cannot simply be interpreted in terms of general slowness of bvFTD patients relative to the other two groups of participants. Indeed, basic eye movement metrics (saccade latencies, time to fixate the target, mean fixation duration and saccade velocity) were similar for the three groups of participants, suggesting that basic motor eye movement functions are spared in bvFTD patients and comparable across groups. The reduced number of anticipations and, even more importantly, the large number of incorrect anticipations, which took the patients' eyes away from the upcoming target and required a saccadic adjustment to finally fixate the actual target within the matrix can together justify the longer times required to fixate the targets in the sequence (Fig. 3) and the longer reaction times observed in button pressing.

The present study also provides important methodological inputs about the potential ways eyetracking data can be analysed and can provide valuable information to cognitive psychology. Using a linear regression model, we have shown that it is possible to train a model which can predict the next gaze location of a control subject with an average error on an external test set of 0.57 pixels. Applying these models, trained only using a proportion of the control subjects, we have shown that the models have a substantially higher error on bvFTD patients than when applied to controls and SD patients. This increase in error suggests that the model has learnt a trend that is specific to control subjects. Using models trained on individual trials we have demonstrated that it is possible to show a significant difference between groups, using a small cohort and a small amount of trials. We have also shown that the ML can detect a spatial anticipation deficit on certain trials where the frequentist summary metrics could not. This is the case for a larger proportion of trials in bvFTD and SD patients as compared to controls. Although false positive results are possible, this is quite unlikely given the lack of pathological scores found in the controls used as test set ($N = 12$). Assessment and analysis techniques which enable the detection of relevant differences between clinical populations will have broad applicability beyond the domain of spatial anticipation to other areas of cognitive testing in dementia, and may offer the opportunity to reduce testing time and associated tiredness and distress for participants.

Using linear regression models, as well as other machine learning methods has the advantage of being well supported, open source and time efficient. ML models can also be applied to many different types of data simultaneously, so it is possible to collect data from other sources and include this in the modelling process, such as brain atrophy location or gait analysis.

Important strengths of the frequentist statistical summary metrics were also detected. In particular, it allowed the identification of a spatial anticipation deficit in 92% in bvFTD patients as compared to the 58% obtained by using the pen and paper Brixton spatial anticipation test (albeit with greater testing time). Importantly, the metrics also highlighted unimpaired performance in SD patients whereas the original test detected a deficit in 33% of individuals, possibly due to the complexity of the original test instructions. The nature of the machine learning approach adopted in the present study did not allow a direct comparison on an individual basis.

Some limitations of the current paper are noteworthy. Sample sizes are limited, especially for the groups of patients. This might have contributed to the null effect obtained in the neuroimaging analysis. A future study looking at an extended sample size is desirable. Demographically, although the two groups of patients were balanced in terms of gender, a larger proportion of females were observed in the control group. In all the statistical analyses we co-varied, among other variables, for gender. Moreover, Bielak et al. (2006) in reporting the norms for the Brixton test on 457 individuals, demonstrated that female participants made a larger proportion of errors than males. The same result emerges in our data suggesting that, if anything, our data may represent an underestimate of the effect reported in bvFTD patients. In terms of the machine learning approach, one limitation of applying modelling in the way we did is that the results are not as interpretable compared to traditional eyetracking measures. Psychology research has long used measures such as saccades and fixations which can be more easily interpreted as an indication of some high order function. However, predictive error of a model is less interpretable, as it is more an indication of how different a subject belonging to a clinical population is to the control (training) population. We believe, however, that the combination of the two methodologies can reduce the overall limitation and enhance the advantages associated with each one.

In the present study we have developed an eyetracking paradigm, based on an existing pen and paper test, which can provide fine-grained eyetracking metrics indicative of high level cognitive impairments in frontotemporal dementia. It has been shown that the presence of a spatial anticipation deficit in bvFTD patients and its absence in SD patients is well detected by eyetracking metrics, namely correct and incorrect anticipatory saccades. In particular, we believe that the proportion of incorrect attempts of anticipation may constitute a particularly valuable marker of anticipation deficits, one of the core symptoms of bvFTD.

One future application of the paradigm is to evaluate subtle executive dysfunction in a population of individuals at risk of developing frontotemporal dementia. Frontotemporal dementia has a substantial genetic component, with an autosomal dominant inheritance pattern in around 10–20% of cases across large published series (Rohrer et al., 2009; Rohrer and Warren, 2011). We aim to explore whether this fine-grained eyetracking measure of a specific element of executive dysfunction may add value in clinical or research settings to an early diagnosis of frontotemporal dementia. Although the test in its current form is brief enough to be used in a clinical context, further improvements will need to focus on adapting the test for use on cheaper eyetracking systems which also have easier training requirements for the person running the test. In this context, the application of a ML approach to the results can speed up data analysis, which is normally extremely time consuming when frequentist summary metrics need to be extracted. Finally the present test could potentially represent a sensitive outcome measures for clinical trials, whose increased number needs to be paralleled by appropriate cognitive endpoints in terms of sensitivity and specificity (for a review see Miller et al., 2014).

Acknowledgments

This work was undertaken at UCLH/UCL which received a proportion of funding from the Department of Health's NIHR Biomedical Research Centre funding scheme. The Dementia Research Centre is supported by Alzheimer's Research UK, Brain Research Trust, and The Wolfson Foundation. This work was supported by the NIHR Queen Square Dementia Biomedical Research Unit and by an Alzheimer's Research UK Senior Research Fellowship and ESRC/NIHR (ES/K006711/1) and EPSRC (EP/M006093/1) grants to SC. JDW was funded by the Alzheimer's Society and also supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. JDR is supported by an MRC Clinician Scientist Fellowship (MR/M008525/1) and has received funding from the NIHR Rare Disease Translational Research Collaboration (BRC149/NS/MH).

Conflict of interest

Nothing to report.

References

- Abel, L.A., Bowman, E.A., Velakoulis, D., Fahey, M.C., Desmond, P., Macfarlane, M.D., Walterfang, M., 2012. Saccadic eye movement characteristics in adult niemann-pick type c disease: relationships with disease severity and brain structural measures. *PLoS One* 7 (11), e50947. <http://dx.doi.org/10.1371/journal.pone.0050947>.
- Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. *Neuroimage* 38 (1), 95–113.
- Ashburner, J., Friston, K.J., 2009. Computing average shaped tissue probability templates. *Neuroimage* 45 (2), 333–341.
- Bielak, A., Mansueti, L., Strauss, E., Dixon, R., 2006. Performance on the Hayling and Brixton tests in older adults: norms and correlates. *Arch. Clin. Neuropsychol.* 21 (2), 141–149. <http://dx.doi.org/10.1016/j.acn.2005.08.006>.
- Boxer, A.L., Garbutt, S., Rankin, K.P., Hellmuth, J., Neuhaus, J., Miller, B.L., Lisberger, S.G., 2006. Medial versus lateral frontal lobe contributions to voluntary saccade control as revealed by the study of patients with frontal lobe degeneration. *J. Neurosci.* 26 (23), 6354–6363. <http://dx.doi.org/10.1523/JNEUROSCI.0549-06.2006>.
- Burgess, P.W., Shallice, T., 1996. Bizarre responses, rule detection and frontal lobe lesions. *Cortex J. Devoted Study Nerv. Syst. Behav.* 32 (2), 241–259.
- Burgess, P.W., Shallice, T., 1997. *The Hayling Island and Brixton Test manual*. Bury St Edmunds: Thames Valley Test Co.
- Burrell, J.R., Hornberger, M., Carpenter, R.H.S., Kiernan, M.C., Hodges, J.R., 2012. Saccadic abnormalities in frontotemporal dementia. *Neurology* 78 (23), 1816–1823. <http://dx.doi.org/10.1212/WNL.0b013e318258f75c>.
- Bylsma, F.W., Rasmussen, D.X., Rebok, G.W., Keyl, P.M., Tune, L., Brandt, J., 1995. Changes in visual fixation and saccadic eye movements in Alzheimer's disease. *Int. J. Psychophysiol.: Off. J. Int. Organ. Psychophysiol.* 19 (1), 33–40.
- Chan, D., Fox, N.C., Scahill, R.I., Crum, W.R., Whitwell, J.L., Leschziner, G., Rossor, M.N., 2001. Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann. Neurol.* 49 (4), 433–442.
- Crawford, T.J., Higham, S., Renvoize, T., Patel, J., Dale, M., Suriya, A., Tetley, S., 2005. Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's disease. *Biol. Psychiatry* 57 (9), 1052–1060. <http://dx.doi.org/10.1016/j.biopsych.2005.01.017>.
- Crutcher, M.D., Calhoun-Haney, R., Manzanares, C.M., Lah, J.J., Levey, A.I., Zola, S.M., 2009. Eye tracking during a visual paired comparison task as a predictor of early dementia. *Am. J. Alzheimer's Dis. Other Dement.* 24 (3), 258–266. <http://dx.doi.org/10.1177/1533317509332093>.
- Findlay, J.M., 1981. Spatial and temporal factors in the predictive generation of saccadic eye movements. *Vision. Res.* 21 (3), 347–354.
- Fletcher, W.A., Sharpe, J.A., 1986. Saccadic eye movement dysfunction in Alzheimer's disease. *Ann. Neurol.* 20 (4), 464–471. <http://dx.doi.org/10.1002/ana.410200405>.
- Galton, C.J., Patterson, K., Graham, K., Lambon-Ralph, M.A., Williams, G., Antoun, N., Hodges, J.R., 2001. Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. *Neurology* 57 (2), 216–225.
- Garbutt, S., Matlin, A., Hellmuth, J., Schenk, A.K., Johnson, J.K., Rosen, H., Boxer, A.L., 2007. Oculomotor function in frontotemporal lobar degeneration, related disorders and Alzheimer's disease. *Brain* 131 (5), 1268–1281. <http://dx.doi.org/10.1093/brain/awn047>.
- Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S., et al., 2011. Classification of primary progressive aphasia and its variants. *Neurology* 76 (11), 1006–1014.
- Harciarek, M., Cosentino, S., 2013. Language, executive function and social cognition in the diagnosis of frontotemporal dementia syndromes. *Int. Rev. Psychiatry* 25 (2), 178–196. <http://dx.doi.org/10.3109/09540261.2013.763340>.
- Helie, S., Chakravarthy, S., Moustafa, A.A., 2013. Exploring the cognitive and motor functions of the basal ganglia: an integrative review of computational cognitive

- neuroscience models. *Front. Comput. Neurosci.* 7.
- Hommer, D.W., Clem, T., Litman, R., Pickar, D., 1991. Maladaptive anticipatory saccades in schizophrenia. *Biol. Psychiatry* 30 (8), 779–794.
- Hornberger, M., Savage, S., Hsieh, S., Mioshi, E., Piguet, O., Hodges, J.R., 2010. Orbitofrontal dysfunction discriminates behavioral variant frontotemporal dementia from Alzheimer's disease. *Dement. Geriatr. Cogn. Disord.* 30 (6), 547–552. <http://dx.doi.org/10.1159/000321670>.
- Husain, M., Parton, A., Hodgson, T.L., Mort, D., Rees, G., 2003. Self-control during response conflict by human supplementary eye field. *Nat. Neurosci.* 6 (2), 117–118.
- Johnston, K., Everling, S., 2008. Neurophysiology and neuroanatomy of reflexive and voluntary saccades in non-human primates. *Brain Cogn.* 68 (3), 271–283. <http://dx.doi.org/10.1016/j.bandc.2008.08.017>.
- Julien, C.L., Thompson, J.C., Neary, D., Snowden, J.S., 2008. Arithmetic knowledge in semantic dementia: is it invariably preserved? *Neuropsychologia* 46 (11), 2732–2744. <http://dx.doi.org/10.1016/j.neuropsychologia.2008.05.010>.
- Kipps, C.M., Nestor, P.J., Fryer, T.D., Hodges, J.R., 2007. Behavioural variant frontotemporal dementia: not all it seems? *Neurocase* 13 (4), 237–247. <http://dx.doi.org/10.1080/13554790701594870>.
- Lemos, J., Eggenberger, E., 2013. Saccadic intrusions: review and update. *Curr. Opin. Neurol.* 26 (1), 59–66. <http://dx.doi.org/10.1097/WCO.0b013e32835c5e1d>.
- Lough, S., Kipps, C.M., Treise, C., Watson, P., Blair, J.R., Hodges, J.R., 2006. Social reasoning, emotion and empathy in frontotemporal dementia. *Neuropsychologia* 44 (6), 950–958. <http://dx.doi.org/10.1016/j.neuropsychologia.2005.08.009>.
- MacKay, D.J.C., 1992. Bayesian interpolation. *Neural Comput.* 4 (3), 415–447. <http://dx.doi.org/10.1162/neco.1992.4.3.415>.
- McKenna, P., Warrington, E.K., 1980. Testing for nominal dysphasia. *J. Neurol. Neurosurg. Psychiatry* 43 (9), 781–788. <http://dx.doi.org/10.1136/jnnp.43.9.781>.
- Mendez, M.F., Joshi, A., Tassniyom, K., Teng, E., Shapira, J.S., 2013. Clinicopathologic differences among patients with behavioral variant frontotemporal dementia. *Neurology* 80 (6), 561–568. <http://dx.doi.org/10.1212/WNL.0b013e3182815547>.
- Mendez, M.F., Shapira, J.S., 2011. Loss of emotional insight in behavioral variant frontotemporal dementia or frontal anosodiaphoria. *Conscious. Cogn.* 20 (4), 1690–1696. <http://dx.doi.org/10.1016/j.concog.2011.09.005>.
- Meyniel, C., 2005. Saccade impairments in patients with fronto-temporal dementia. *J. Neurol. Neurosurg. Psychiatry* 76 (11), 1581–1584. <http://dx.doi.org/10.1136/jnnp.2004.060392>.
- Miller, J.B., Banks, S.J., Léger, G.C., Cummings, J.L., 2014. Randomized controlled trials in frontotemporal dementia: cognitive and behavioral outcomes. *Transl. Neurodegener.* 3 (1), 1.
- Pancaroglu, R., Hills, C.S., Sekunova, A., Viswanathan, J., Duchaine, B., Barton, J.J.S., 2016. Seeing the eyes in acquired prosopagnosia. *CortexJ. Devoted Study Nerv. Syst. Behav.* 81, 251–265. <http://dx.doi.org/10.1016/j.cortex.2016.04.024>.
- Paus, T., 1996. Location and function of the human frontal eye-field: a selective review. *Neuropsychologia* 34 (6), 475–483.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12 (Oct), 2825–2830.
- Pelak, V.S., 2010. Ocular motility of aging and dementia. *Curr. Neurol. Neurosci. Rep.* 10 (6), 440–447. <http://dx.doi.org/10.1007/s11910-010-0137-z>.
- Pouget, P., 2015. The cortex is in overall control of “voluntary” eye movement. *Eye* 29 (2), 241–245.
- Primativo, S., Arduino, L.S., Daini, R., De Luca, M., Toneatto, C., Martelli, M., 2015. Impaired oculo-motor behaviour affects both reading and scene perception in neglect patients. *Neuropsychologia* 70, 90–106.
- Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., Miller, B.L., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 134 (9), 2456–2477. <http://dx.doi.org/10.1093/brain/awr179>.
- Reverberi, C., Lavaroni, A., Gigli, G.L., Skrap, M., Shallice, T., 2005. Specific impairments of rule induction in different frontal lobe subgroups. *Neuropsychologia* 43 (3), 460–472. <http://dx.doi.org/10.1016/j.neuropsychologia.2004.06.008>.
- Richmond, J., Sowerby, P., Colombo, M., Hayne, H., 2004. The effect of familiarization time, retention interval, and context change on adults' performance in the visual paired-comparison task. *Dev. Psychobiol.* 44 (2), 146–155. <http://dx.doi.org/10.1002/dev.10161>.
- Ridgway, G.R., Omar, R., Ourselin, S., Hill, D.L., Warren, J.D., Fox, N.C., 2009. Issues with threshold masking in voxel-based morphometry of atrophied brains. *Neuroimage* 44 (1), 99–111.
- Rohrer, J.D., Guerreiro, R., Vandrovцова, J., Uphill, J., Reiman, D., Beck, J., et al., 2009. The heritability and genetics of frontotemporal lobar degeneration. *Neurology* 73 (18), 1451–1456.
- Rohrer, Jonathan D., Warren, J.D., 2011. Phenotypic signatures of genetic frontotemporal dementia. *Curr. Opin. Neurol.* 24 (6), 542–549.
- Saeb, S., Lonini, L., Jayaraman, A., Mohr, D.C., Kording, K.P., 2016. Voodoo Machine Learning for Clinical Predictions (No. biorxiv:059774v1). Retrieved from <<http://dx.doi.org/http://biorxiv.org/lookup/doi/10.1101/059774>>.
- Schall, J.D., 1995. Neural basis of saccade target selection. *Rev. Neurosci.* 6 (1), 63–85.
- Schuh, J.M., Eigsti, I.-M., Mirman, D., 2016. Discourse comprehension in autism spectrum disorder: effects of working memory load and common ground. *Autism Res.: Off. J. Int. Soc. Autism Res.* <http://dx.doi.org/10.1002/aur.1632>.
- Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L., Greicius, M.D., 2009. Neurodegenerative diseases target large-scale human brain networks. *Neuron* 62 (1), 42–52. <http://dx.doi.org/10.1016/j.neuron.2009.03.024>.
- Shafiq-Antonacci, R., Maruff, P., Masters, C., Currie, J., 2003. Spectrum of saccade system function in Alzheimer disease. *Arch. Neurol.* 60 (9), 1272–1278. <http://dx.doi.org/10.1001/archneur.60.9.1272>.
- Shakespeare, T.J., Kaski, D., Yong, K.X.X., Paterson, R.W., Slattery, C.F., Ryan, N.S., Crutch, S.J., 2015. Abnormalities of fixation, saccade and pursuit in posterior cortical atrophy. *Brain: J. Neurol.* 138 (Pt7), 1976–1991. <http://dx.doi.org/10.1093/brain/awv103>.
- Smit, A.C., Van Gisbergen, J.A.M., 1989. A short-latency transition in saccade dynamics during square-wave tracking and its significance for the differentiation of visually-guided and predictive saccades. *Exp. Brain Res.* 76 (1), 64–74.
- Snowden, J.S., Gibbons, Z.C., Blackshaw, A., Doubleday, E., Thompson, J., Craufurd, D., Neary, D., 2003. Social cognition in frontotemporal dementia and Huntington's disease. *Neuropsychologia* 41 (6), 688–701.
- Thompson, J.C., Stopford, C.L., Snowden, J.S., Neary, D., 2005. Qualitative neuropsychological performance characteristics in frontotemporal dementia and Alzheimer's disease. *J. Neurol., Neurosurg. Psychiatry* 76 (7), 920–927.
- Tombaugh, T., 2004. Trail making test A and B: normative data stratified by age and education. *Arch. Clin. Neuropsychol.* 19 (2), 203–214. [http://dx.doi.org/10.1016/S0887-6177\(03\)00039-8](http://dx.doi.org/10.1016/S0887-6177(03)00039-8).
- Vordenberg, J.A., Barrett, J.J., Doninger, N.A., Contardo, C.P., Ozoude, K.A., 2014. Application of the Brixton spatial anticipation test in stroke: ecological validity and performance characteristics. *Clin. Neuropsychol.* 28 (2), 300–316. <http://dx.doi.org/10.1080/13854046.2014.881555>.
- Yang, Q., Wang, T., Su, N., Liu, Y., Xiao, S., Kapoula, Z., 2011. Long latency and high variability in accuracy-speed of prosaccades in Alzheimer's disease at mild to moderate stage. *Dement. Geriatr. Cogn. Disord. Extra* 1 (1), 318–329. <http://dx.doi.org/10.1159/000333080>.
- Yang, Q., Wang, T., Su, N., Xiao, S., Kapoula, Z., 2013. Specific saccade deficits in patients with Alzheimer's disease at mild to moderate stage and in patients with amnesic mild cognitive impairment. *Age (Dordr. Neth.)* 35 (4), 1287–1298. <http://dx.doi.org/10.1007/s11357-012-9420-z>.