Supplementary Material

A Preprocessing

EEG Minimal Preprocessing In this preprocessing strategy first, we identify bad channels using the EEGLAB [55] clean rawdata algorithm.[2] This algorithm consists of three steps: (1) it detects channels that have no signal variation for more than 5 seconds and filters the EEG of the remaining channels with a forward-backwards (non-causal) filter, here we use a high-pass filter with a transition band of [0.25, 0.75] Hz; (2), it searches for channels with a lower correlation to its robust estimate than a given threshold that we set to 0.85; and (3) it removes EEG channels with excessive line noise that in our case corresponds to 4 standard deviations. Subsequently, the high Variance Criterion (HVC) was applied, with a pre- specified threshold of $100\mu V$ as the upper limit for identifying bad channels. Later, the EEG data were band-pass filtered at 0.5 and 40 Hz and then, Zapline toolkit [56] was applied to remove line noise artifacts, discarding 7 power line components. Finally, all the channels marked as bad were interpolated using spherical spline interpolation.

EEG Maximal Preprocessing. In addition to the whole minimal preprocessing pipeline, in the Maximal Preprocessing Pipeline, the Independent component analysis (ICA) is used to isolate the various source generator processes underlying those recordings. Non-Brain artifactual source components are removed based on the automatic classification result as provided by Independent Component Label (ICLabel) [38]. ICA included Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments (Dimigen 2020) as implemented in Pedroni et al. [37].

Objective Quality Classification After the bad channel interpolation, all EEG datasets were rated with the objective quality criteria of the Automagic Toolbox [37]. Any data file rated as bad, meaning that the proportion of high-amplitude data points in the signal ($> 30\mu V$) was larger than 0.2, more than 20% of the time points showed a variance larger than $15\mu V$ across channels, 40% of the channels showed high variance ($15\mu V$), or the ratio of bad channels was higher than 0.4 was not included in the further data processing.

B Annotations

Saccades are detected by the velocity and acceleration of the eye movements. We used the default system (SR Research, http://www.sr-research.com/) parameters to define saccades: an acceleration threshold of 8000° per second, a velocity threshold of 30° per second, and a deflection threshold of 0.1°.

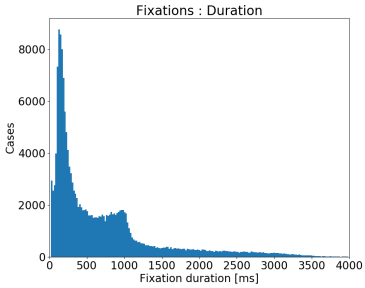
Fixations were defined as time periods without saccades. Fixation might include small saccades (i.e., microsaccades), which fall below the threshold for saccade detection. Furthermore, a blink was be regarded as a special case of a fixation, where the pupil diameter is either zero or outside a dynamically computed valid pupil, or the horizontal and vertical gaze positions are zero.

We extracted the following information about the saccades, fixations and blinks: start and end time, duration, coordinates of start positions on the computer screen in pixels, and for saccades additionally: end positions, and amplitudes.

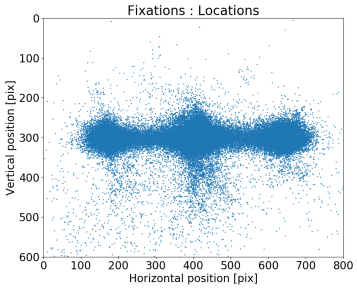
C Events Characteristics

In this appendix we show further details of each dataset. More precisely, for each experimental paradigm we present the distribution of the fixation duration, the distribution of the fixation positions, the distribution of the saccade amplitude (in pixels, where 1 pixel corresponds to 0.5 mm), and the distribution of the saccade angle.

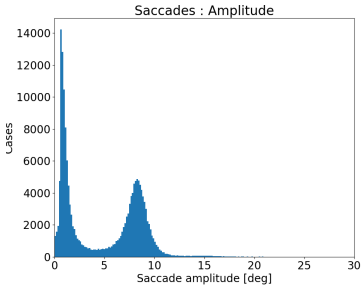
[2] `clean_rawdata()` from: `http://sccn.ucsd.edu/wiki/Plugin_list_process`



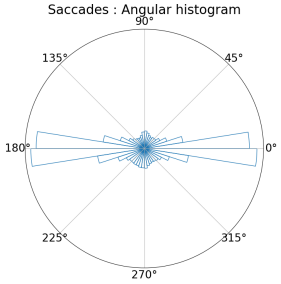
(a) Distribution of the fixation duration



(b) Distribution of the fixation positions

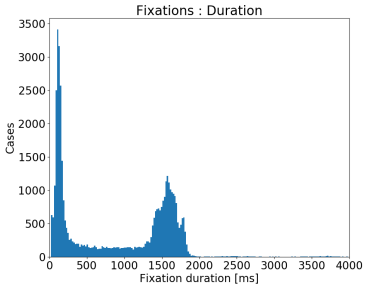


(c) Distribution of the saccade amplitude

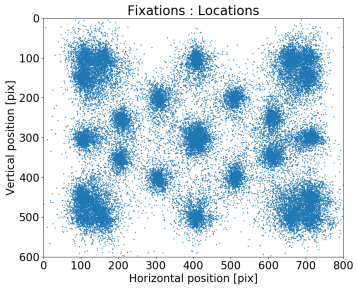


(d) Distribution of the saccade angle

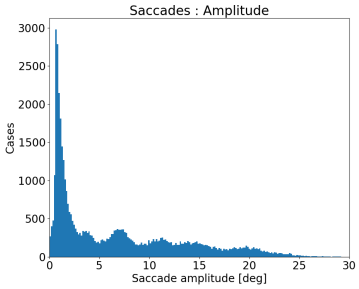
Figure 6: Antisaccade Paradigm



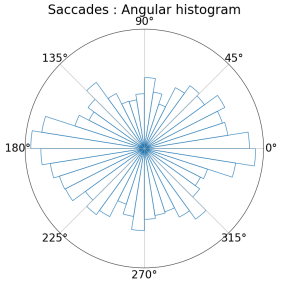
(a) Distribution of the fixation duration



(b) Distribution of the fixation positions

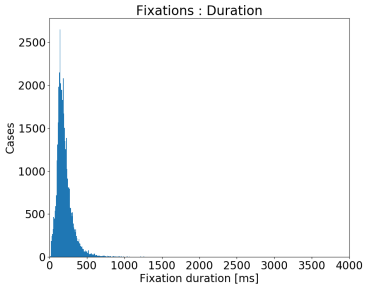


(c) Distribution of the saccade amplitude

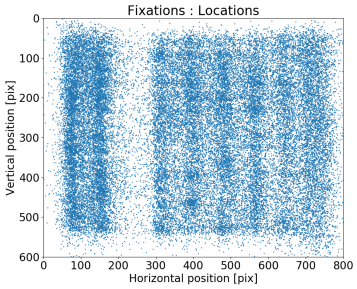


(d) Distribution of the saccade angle

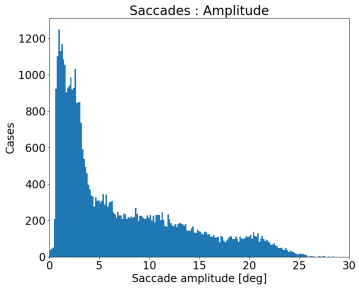
Figure 7: Large Grid Paradigm



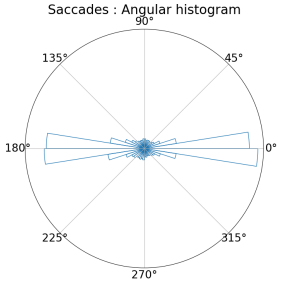
(a) Distribution of the fixation duration



(b) Distribution of the fixation positions

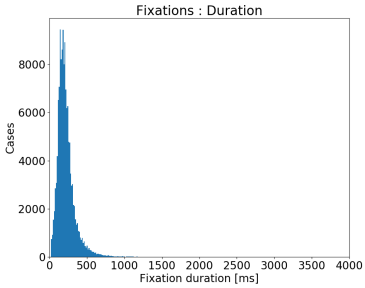


(c) Distribution of the saccade amplitude

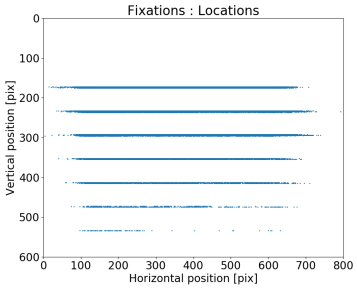


(d) Distribution of the saccade angle

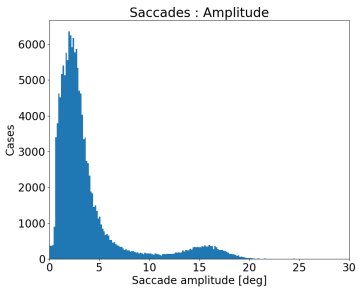
Figure 8: Visual Symbol Search Paradigm



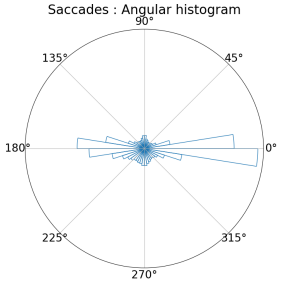
(a) Distribution of the fixation duration



(b) Distribution of the fixation positions



(c) Distribution of the saccade amplitude



(d) Distribution of the saccade angle

Figure 9: ZuCo 2.0 Paradigm

D Details of the Models

In this appendix we present the details of our models used during our experimentations.

D.1 Deep Learning Models

In the following we present five different deep learning architectures that we evaluate on the EEGEyeNet benchmark.

EEGNet. [51] is a convolutional neural network tailored to Brain-Computer Interfaces applications. This model first performs a temporal convolution to learn frequency filters, then a depthwise convolution to learn frequency-specific spatial filters followed by a separable convolution which should learn a temporal summary for each individual feature map. Finally, a pointwise convolution is applied to mix the feature maps. For details on this architecture we refer the reader to the original work [51], which we follow in our implementation.

Convolutional Neural Network (CNN). We experiment with a standard one-dimensional convolutional neural network with 12 layers and additive residual connections around blocks of three layers. Each layer consists of (1D-)convolution, batch normalization [57], ReLU activation and max pooling. In the convolutions we use 16 filters of size size 64, and for the pooling operation a kernel of size 2 and stride 1. Each residual connection performs a convolution followed by batch normalization.

Pyramidal CNN. We consider a CNN with an inverted pyramidal shape and 6 layers. Each layer consists of the same modules as the CNN, with the difference that the number of filters is a multiple of 16 that grows with depth, i.e., 16 in the first layer, 32 in the second and so on. In this case, the kernel size is 16 and there are not residual connections.

InceptionTime We re-implement the model from [58], which is an adaptation of the Inception-v4 architecture [59] to time series classification. We instantiate this model with 12 layers and skip connections every three layers. Each layer consists of an InceptionTime module, which takes as input 64 channels and performs a 1×1 bottleneck convolution that produces a feature map with 16 channels. Then, this feature map is passed through three different convolution operators with 16 filters each and kernel sizes of 64, 32, 16; and a max-pooling operator with a kernel size of 3. The four resulting feature maps are concatenated to form the layer's output, which has 64 channels.This model also uses residual connections around each block of three layers.

Xception Finally, we build a model based on the Xception architecture proposed by Chollet [60]. This model follows the same general structure as our CNN, with 12 layers and residual connections after three layers. Each layer contains a 1D depthwise separable convolution [60], with 64 filters and kernel size of 40, followed by batch normalization and ReLU activation.

Model	*Par*1	*Par*2	*Par*3	*Par*4
KNN	leaf size 10	10LR/25AN/50AP	-	-
GaussianNB	var smooth: 0.0004941713	-	-	-
LinearSVC	C:0.01	tol: 1e-05	max iter: 1200	-
RBF SVC/SVR	C:1	tol: 1e-05	max iter: 1200	gam:0.01
Linear regression	all default	-	-	-
Lasso regression	alpha: 0.01AN/1	tol: 1e-05	max iter: 1200	-
Elastic Net	alpha: 0.1AN/1	l1 ratio: 0.9LR/0.3AN/0.6AP,	tol: 1e-05	gam:0.01
Random Forest	max depth 10LR/10AN/50	est: 50AN/50AM/250	-	-
AdaBoost	lr: 0.5LR/0.5AN/0.1AM/0.01AP	est: 100LR/50	-	-
XGBoost	eta: 0.05AN/0.1	est: 100AP/250	max depth: 10AP/5	-

Table 5: Hyperparameters of classical machine learning models. LR - Left/Right task, AN - Angle task, AM - Amplitude task, and AP - Absolute Position task.

Parameter	*CNN*	*PyramidalCNN*	*EEGNet*	*InceptionTime*	*Xception*
Depth	12	6	2	12	18
Number of filters	16	16	16/256	16	64
Kernel size	64	16	256	64	40
Batch size	64	64	64	64	64
Epochs	50	50	50	50	50
Early stopping patience	20	20	20	20	20
Residual connections	True	False	False	True	True
Bottleneck size	-	-	-	16	-
Dropout rate	-	-	0.5	-	-

Table 6: Hyperparameters of the deep learning models. EEGNet uses two different amounts of filters in its two blocks.

D.2 Hyperparameter Tuning

In Table 5 we report our tuned hyperparameters for the classical machine learning models and in Table 6 for the deep learning models.

E Further Experiments

E.1 Experiments on the Maximally Preprocessed Data

Besides the minimally preprocessed data presented in the benchmark results at Section 4, we performed the same experiments on the maximally preprocessed data. As expected, the models perform much worse compared to the ones with minimally preprocessed data (see Table 7). However, the following results show that even after removing eye artifacts during preprocessing, it is still possible to infer gaze direction based on EEG data.

	Left-Right	Angle/Amplitude		Abs. Position
Model	Accuracy	Angle RMSE	Amp. RMSE	RMSE
KNN	60.9 ±0	1.83 ±0	72.2 ±0	123.1 ±0
GaussianNB	58.5 ±0	-	-	-
LinearSVC	63.2 ±0	-	-	-
RBF SVC/SVR	53.9 ±0	1.91 ±0	76.1 ±0	122.6 ±0
Linear Regression	-	1.76 ±0	72.2 ±0	123.5 ±0
Ridge Regression	-	1.79 ±0	71.6 ±0	122.6 ±0
Lasso Regression	-	1.77 ±0	71.4 ±0	122.6 ±0
Elastic Net	-	1.80 ±0	71.5 ±0	122.5 ±0
Random Forest	66.1 ±0.2	1.77 ±0.01	71.1 ±0	123.5 ±0.2
Gradient Boost	66.9 ±0.2	1.78 ±0.01	71 ±0	122.4 ±0
AdaBoost	65.9 ±0	1.84 ±0.01	72.8 ±0	122.7 ±0.1
XGBoost	68.3 ±0	1.75 ±0	70.5 ±0	123.2 ±0
CNN	75.5 ±1.7	**1.07** ±0.05	76.8 ±2.3	134.6 ±2.8
PyramidalCNN	**86.6** ±1.1	1.38 ±0.35	76.6 ±2.2	142.1 ±3.1
EEGNet	83.6 ±1.1	1.31 ±0.03	**67.8**±2.1	**119.9** ±0.8
InceptionTime	80.4 ±2.9	1.65 ±0.33	70.8 ±1.7	137.2 ±1.7
Xception	75.7 ±1.8	1.57 ±0.29	76.3 ±2.1	141.1 ±1.3
Naive Baseline	52.3	1.90	74.7	123.3

Table 7: Data maximally preprocessed, 5 runs per DL model, Adam optimizer with learning rate 1e-4, early stopping patience 20. *Angle* is measured in radians, *Amplitude* and *Abs. Position* in mm.

Left-Right. We see that classical machine learning models achieve a better performance than the naive baseline of 52.26%. In particular, tree-based models (RandomForest, GradientBoosting, AdaBoost and XGBoost) reach a performance of over 68%. Although it is not a high performance, it shows that the eye movement information between left and right can also be inferred from "pure" brain activity. Furthermore, the deep learning models achieve an even higher performance of over 86%, showing that one can achieve satisfactory results for this task also from the maximally preprocessed data.

Angle/Amplitude. As in the case for minimally preprocessed data (see Table 4), the results in Table 7 show that this task is more demanding than Left-Right. All classical models perform close to random. The deep learning models perform slightly better, and, interestingly, the simple CNN architecture performs best. This result is above the naive baseline but not by a big margin. Nonetheless, it shows that information about the saccade angle can be extracted from maximally preprocessed data as well. In comparison to the angle, we see that the amplitude task is strictly more difficult for maximally preprocessed data in the third column. All models except EEGNet fail at this task, with EEGNet performing slightly better than the naive baseline. This leaves the open question to what extent the angle and amplitude of saccades can be estimated from maximally preprocessed EEG data.

Absolute Position. In the last column of Table 7 we see that all models fail in this task, with performances very close to the naive baseline. Deep learning models (except EEGNet) perform even worse than the naive baseline. At this point, it is not clear whether the absolute position of fixations can be inferred from the maximally preprocessed EEG data.

E.2 Experiments on ZuCo 2.0 Dataset

For comparison, we performed the same experiments on one of these openly available datasets, namely the ZuCo 2.0. We have used the same pipeline from our infrastructure. First, we have synchronized the EEG-ET data and extracted the relevant events for annotation (fixation and saccades). Then we also performed feature extraction. Afterwards, we used our data preparation tool to extract the samples for our benchmarking tasks. Finally, we ran our benchmark with this dataset. Note that we didn't use the LR-task for this dataset due to the nature of the ZuCo 2.0 experimental paradigm. Although during reading, most of the eye movements are towards the right, the dataset contains eye movements in all other directions (see Figure 9). Other referenced datasets don't provide an interface for benchmarking purposes. Again, this is because there are differences in the recording setup, protocol and paradigms.

As expected, the models perform worse compared to the ones with minimally preprocessed data (see Table 4).

Angle/Amplitude. We can see a clear difference between the results in our dataset (see Table 4) and the results obtained in ZuCo 2.0 dataset. The models performs worse in the ZuCo 2.0 dataset. For the angle task, the best model is the EEGNet with RMSE 1.35 radians. It performs better then the naive baseline (RMSE 1.90 radians), but not by a big margin. In comparison, in our dataset, EEGNet performs much better with an RMSE of 0.70 radians. In addition, the best performing model for our dataset is the simple CNN, which has an RMSE of 0.33 radians. We can also see a clear difference for the amplitude task as well. The best performing model for ZuCo 2.0 dataset is PyramidalCNN with an RMSE of 61.9 mm, while in our dataset PyramidalCNN has a better performance with an RMSE of 30.7 mm.

Absolute Position. In this task, the results are not directly comparable to our dataset, since they depend on the distribution of the fixation positions. The difference of the distributions can be seen in Figure 7 and Figure 9. Furthermore, in Table 4 we can see that the naive baseline for our dataset is 123.3, whereas for the ZuCo 2.0 dataset (Table 8), it is 72. In other words, in ZuCo 2.0 simply predicting the mean position leads to a better performance compared to our dataset, because the fixations are not scattered across all possible positions on the screen. In contrast, our dataset is specifically tailored for eye movement prediction. However, we can still see a clear difference in the performance. In our dataset, there is a

	Angle/Amplitude		Abs. Position
Model	Angle RMSE	Amp. RMSE	RMSE
KNN	1.42 ±0	59.9±0	72.2 ±0
RBF SVC/SVR	1.47 ±0	91.4 ±0	77.5 ±0
Linear Regression	1.40 ±0	67.6 ±0	72.2 ±0
Ridge Regression	1.41 ±0	67.4 ±0	72 ±0
Lasso Regression	1.41 ±0	67.3 ±0	72 ±0
Elastic Net	1.42 ±0	67.3 ±0	72 ±0
Random Forest	1.42 ±0.01	70.6 ±0	72.9 ±0.1
Gradient Boost	1.42 ±0.01	67.4 ±0	61.9 ±0
AdaBoost	1.41 ±0.01	69 ±0	75.3 ±0.2
XGBoost	1.42 ±0	67.5 ±0	**72.1** ±0
CNN	1.54 ±0.4	67.1 ±1.1	75.8 ±0.3
PyramidalCNN	1.70 ±0.4	**61.9** ±0.6	73.8 ±1.6
EEGNet	**1.35** ±0.2	62.1 ±1.5	72.7 ±0.1
InceptionTime	1.81 ±0.02	67.3 ±1.1	80 ±1.1
Xception	1.67 ±0.4	65.9 ±1.9	76.2 ±2.4
Naive Baseline	1.41	67	72

Table 8: Results of Angle/Amplitude and Absolute Position tasks on the ZuCo dataset. Same hyperparameters as for experiments on our dataset. *Angle* is measured in radians, *Amplitude* and *Abs. Position* in mm.

considerable gap between the deep learning models and the naive baseline, whereas in the ZuCo dataset all models perform close to the baseline. One possible explanation for this (besides the size and the type of dataset) is that ZuCo dataset uses a different preprocessing method, whereas our dataset uses minimally preprocessing data which is more suitable for eye movement prediction.

F Datasheets for Datasets

For the dataset documentation we used the recommended documentation framework "Datasheets for Datasets" [61].

F.1 Motivation

For what purpose was the dataset created? The primary purpose of the project is to advance research that studies the combination of brain activities and gaze position.

Who created the dataset The dataset was collected by the Methods of Plasticity Research Lab at the University of Zurich. All persons participating in the data collection are acknowledged on our website: www.eegeye.net

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number This work was supported by the Velux Stiftung Project No. 1126 and by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF) Grant $100014_1 75875$.

Any other comments? [N/A]

F.2 Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings;

people and interactions between them; nodes and edges)? the EEGEyeNet dataset. contains electroencephalography and eye-tracking recordings following three different experimental paradigms. Together with the raw data, we release two sets of preprocessed data: minimally and maximally preprocessed; as well as the preprocessing code.

How many instances are there in total (of each type, if appropriate)? The number of instances is presented in the Table 3 (Benchmark statistic).

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? The dataset used in this project was recorded in our laboratory in the context of a larger project that aims to quantify age effects on eye movement behaviour and electroencephalography (EEG) recordings of resting-state and task-related activity. Therefore, for the EEGEyeNet dataset, we have used three paradigms that can help in the gaze position prediction.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? For each experimental paradigm, as described in the section 3.4, we release raw data, two sets of preprocessed data: minimally and maximally preprocessed; as well as the preprocessing code.

Is there a label or target associated with each instance? If so, please provide a description [Yes] All labels and targets are described in the Chapter 4.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text. [N/A]

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? [N/A]

Are there recommended data splits (e.g., training, development/validation, testing)? [Yes] All recommended splits with rationales behind them are reported in the chapter 4.

Are there any errors, sources of noise, or redundancies in the dataset? Raw data is often contaminated by artifacts influenced by technical and environmental factors and the recorded participant's specific.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? The dataset is self contained, there are no restrictions associated with any of the external resources that might apply to a future user.

Does the dataset contain data that might be considered confidential [No] The dataset was anonymized.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? [N/A] .

Does the dataset relate to people? [Yes]

Does the dataset identify any subpopulations (e.g., by age, gender)? [Yes] Subpopulations are described in the Subsection 3.1.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? [No] .

Does the dataset contain data that might be considered sensitive in any way [No]

Any other comments? [N/A]

F.3 Collection Process

How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? We recorded ET and EEG brain activity, all procedures and mechanims are described in the Section 3.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? How were these mechanisms or procedures validated? All procedures and mechanims are described in the Section 3.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)? [N/A]

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)? Research Assistants and PhD Students were involved in the data collection process. All obtained fixed salary for Research Assistants/PhD Students.

Over what timeframe was the data collected? The collection of data used in the project began in the first quarter of 2018 and lasted until May 15, 2021.

Were any ethical review processes conducted (e.g., by an institutional review board)? This study was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Institutional Review Board of Canton Zurich (BASEC-Nr. 2017-00226).

Does the dataset relate to people? [Yes]

Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)? We collected the data from the individuals in question directly.

Were the individuals in question notified about the data collection? [Yes] All participants gave their written informed consent before participation in the study, as provided in the Appendix "Formal consent of Participants".

Did the individuals in question consent to the collection and use of their data? [Yes] The template of the formal consent can be found in the Appendix J.

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? [Yes] The Participants signed the following agreement:"I decide voluntarily and can withdraw this decision at any time. If I no longer want to participate, my data will be irrevocably deleted. I only inform the project management and do not have to justify this decision". The German version of the agreement can be found in the Appendix J.

Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis)been conducted? [No]

Any other comments? [Yes] Prior to visiting the laboratory, participants completed a 10 min. pre-screening interview over the phone with a research assistant to confirm their eligibility and safety to participate in the study. This brief interview obtains information

regarding an individual's psychiatric history, including past or present diagnoses and/or treatment, as well as current medications and any neurological disorders (see H). If a participant demonstrates no contraindications for EEG (e.g., history of seizures or epilepsy), he or she is then scheduled for a research study appointment.

F.4 Preprocessing/cleaning/labeling

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? [Yes] All preprocessing steps are described in the Dataset section remainder of the questions in this section.

Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses) [Yes] Together with the raw data, we release two sets of preprocessed data: minimally and maximally preprocessed; as well as the preprocessing code. This way, we give a user the freedom to manipulate raw data while easing the experimentation barrier by additionally providing ready-to-use clean data.

Is the software used to preprocess/clean/label the instances available? If so, please provide a link or other access point. [Yes]

Any other comments? [N/A]

F.5 Uses

Has the dataset been used for any tasks already? [Yes] A subset of the Antisaccade Paradigm was already used in the publication [42].

Is there a repository that links to any or all papers or systems that use the dataset? [No] This repository will be created and maintained in the future.

What (other) tasks could the dataset be used for? The dataset includes experimental paradigms assessing key cognitive functions, like inhibitory control and processing speed. Additionally, the dataset can be used for a development of new segmentation algorithms for fixations, saccades, and blinks.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? [Yes] As described in the preprocessing section of the manuscript, depending on how the data is processed and whether the eye movements component is retained, we may receive different results.

Are there tasks for which the dataset should not be used? [No]

Any other comments? [N/A]

F.6 Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? [No]

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)? The dataset will be made available along with the manuscript on our website. Raw and preprocessed EEG and eye-tracking data are available online and have the digital object identifier: DOI 10.17605/OSF.IO/KTV7M

When will the dataset be distributed? The dataset will be distributed along with the manuscript submission.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? The dataset will distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? [No]

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? [No]

Any other comments [N/A]

F.7 Maintenance

Who is supporting/hosting/maintaining the dataset? The dataset will be maintanted by the EEGEyeNet team.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)? Email addres: akastrati@ethz.ch, martyna.plomecka@uzh.ch

Is there an erratum? If so, please provide a link or other access point [No]

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? [Yes] The dataset will be regularly updated, information about each update will be published in our repository, available on the website www.eegeye.net.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)? [No] .

Will older versions of the dataset continue to be supported/hosted/maintained? [Yes] Yes, all published versions of the dataset will be on our webpage www.eegeye.net.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? [Yes] We release our complete infrastructure and provide a simple and easy-to-use interface to evaluate new methods.

Any other comments? [N/A]

G Risk categorisation

Our study was classified as "Risk category A" by the Institutional Review Board of Canton Zurich (BASEC-Nr. 2017-00226). Applied methods of EEG and eye-tracking are non-invasive and do not pose risk to participants. The involved tasks do not involve deception and completion of these do not pose any harm to participants.

H Inclusion and Exclusion criteria

Inclusion criteria for participation in the study were left and right-handedness, healthy male and female participants, with age between 20 and 80 years old. Before each recording, we assured that the participant did not suffer from psychiatric symptoms, had no severe neurologic disorders, prior head injuries, a stroke, a transient circulatory disorder of the brain, diagnosis of dementia Huntington's disease, Parkinson's disease, sensory and/or motor problems that interfere with computer tasks. Moreover, exclusion criteria were current use of psychotropic drugs, intake of recreational synthetic or natural drugs.

I Potential negative societal impacts of the work

In the specific case of our dataset, one risk of EEG-based eye-tracking models is that they could allow the tracking of the gaze of a subject without their consent, i.e., if the subject agrees to get their EEG data analyzed but not their gaze patterns. Furthermore, a major ethical concern is data privacy. A large-scale dataset such as the one we release in this work could be used in the future to develop methods capable of inferring sensitive information from EEG recordings, such as medical conditions. For this reason, we do not release any privacy-sensitive data. Especially, the anonymity of the participants will be guaranteed when presenting the data at scientific meetings or publishing them in scientific journals. Interested third parties can access the data (but under no circumstances the personal data) via the research repository, solely for scientific purposes (e.g. replication or further analysis). Individual participant medical information obtained from this research project is considered confidential, and disclosure to third parties is prohibited. Participant confidentiality will be further ensured by utilizing identification code numbers to correspond to medical information in the computer files.

J Formal consent of Participants

In the following pages we present the template of the formal consent given to each participant. Every participant signed the formal consent before taking part in this study.

Universität Zürich[UZH]

Psychologisches Institut
Methoden der Plastizitätsforschung

Einwilligungserklärung

Schriftliche Einwilligungserklärung zur Teilnahme an einem Studienprojekt
Bitte lesen Sie dieses Formular sorgfältig durch. Bitte fragen Sie den Versuchsleiter/die Versuchsleiterin, wenn Sie etwas nicht verstehen oder wissen möchten.

BASEC-Nummer (nach Einreichung):	2017-00226
Titel der Studie (wissenschaftlich und Laiensprache):	Prediction of eye-motion with EEG
verantwortliche Institution (Projektleitung mit Adresse):	Psychologisches Institut Universität Zürich Andreasstrasse 15 8050 Zürich
Ort der Durchführung:	Binzmühlstrasse 14 8050 Zürich
Leiter / Leiterin der Studie am Studienort: Name und Vorname in Druckbuchstaben:	
Teilnehmerin/Teilnehmer: Name und Vorname in Druckbuchstaben: Geburtsdatum:	☐ weiblich ☐ männlich

- Ich wurde über den Zweck, den Ablauf des Projekts, über mögliche Vor- und Nachteile sowie über eventuelle Risiken informiert.
- Ich nehme an diesem Projekt freiwillig teil und akzeptiere den Inhalt der zum oben genannten Projekt abgegebenen schriftlichen Information. Ich hatte genügend Zeit, meine Entscheidung zu treffen.
- Meine Fragen im Zusammenhang mit der Teilnahme an diesem Projekt sind mir beantwortet worden. Ich behalte die schriftliche Information und erhalte eine Kopie meiner schriftlichen Einwilligungserklärung.
- Ich weiss, dass meine Daten in verschlüsselter Form zu Forschungszwecken weitergegeben werden können (auch ins Ausland).
- Ich kann jederzeit und ohne Angabe von Gründen von der Teilnahme zurücktreten, ohne dass ich deswegen Nachteile habe.
- Ich bin mir bewusst, dass die in der Teilnehmerinformation genannten Pflichten einzuhalten sind.

Ort, Datum	Unterschrift Teilnehmerin/Teilnehmer

Bestätigung der Prüfperson: Hiermit bestätige ich, dass ich dieser Teilnehmerin/ diesem Teilnehmer Wesen, Bedeutung und Tragweite des Projekts erläutert habe. Ich versichere, alle im Zusammenhang mit diesem Projekt stehenden Verpflichtungen gemäss des geltenden Rechts zu erfüllen. Sollte ich zu irgendeinem Zeitpunkt während der Durchführung des Projekts von Aspekten erfahren, welche die Bereitschaft der Teilnehmerin/ des Teilnehmers zur Teilnahme an der Studie beeinflussen könnten, werde ich sie/ ihn umgehend darüber informieren.

Ort, Datum	Name und Vorname der informierenden der informierenden Prüfperson in Druckbuchstaben Unterschrift der Prüfperson

Universität Zürich[UZH]

Psychologisches Institut
Methoden der Plastizitätsforschung

Einwilligungserklärung für Weiterverwendung von Daten in verschlüsselter Form

Teilnehmerin/Teilnehmer: Name und Vorname in Druckbuchstaben:	
Geburtsdatum:	☐ weiblich ☐ männlich

Ich erlaube, dass meine Daten in verschlüsselter Form für die Forschung (nicht kommerziell) weiter verwendet werden dürfen. Dies bedeutet, dass die Daten in verschlüsselter Form auf einem Server gespeichert sind und für zukünftige, noch nicht näher definierte Forschungsprojekte auf unbestimmte Zeitdauer verwendet werden dürfen.

Ich entscheide freiwillig und kann diesen Entscheid zu jedem Zeitpunkt wieder zurücknehmen. Wenn ich nicht mehr mitmachen möchte, werden meine Daten unwiderruflich gelöscht. Ich informiere lediglich die Projektleitung und muss diesen Entscheid nicht begründen.

Ich habe verstanden, dass die Daten verschlüsselt sind und der Schüssel sicher aufbewahrt wird. Die Daten können im In- und Ausland an andere (nicht kommerzielle) Forschungsgruppen zur Analyse gesendet werden, wenn diese dieselben Standards wie in der Schweiz einhalten. Alle rechtlichen Vorgaben zum Datenschutz werden eingehalten.

Normalerweise werden alle Daten gesamthaft ausgewertet und die Ergebnisse zusammenfassend publiziert. Die Ergebnisse aus den Daten werden nicht kommerziell genutzt werden.

Ort, Datum	Unterschrift Teilnehmerin/ Teilnehmer

Bestätigung des Prüfperson: Hiermit bestätige ich, dass ich dieser Teilnehmerin/ diesem Teilnehmer Wesen, Bedeutung und Tragweite der Weiterverwendung von Daten erläutert habe.

Ort, Datum	Name und Vorname der der informierenden Prüfperson in Druckbuchstaben Unterschrift der Prüfperson