

Quantifying Predictability of Scan Path Data using Active Information Storage

**Patricia Wollstadt, Martina Hasenjäger, Christiane
Wiebel**

2021

Preprint:

This is an accepted article published in Entropy. The final authenticated version is available online at: <https://doi.org/10.3390/e23020167>

Quantifying the predictability of visual scanpaths using Active Information Storage

Patricia Wollstadt^{1,*}, Martina Hasenjäger¹, and Christiane B.
Wiebel-Herboth¹

¹Honda Research Institute Europe GmbH, Carl-Legien-Str. 30,
63073 Offenbach/Main, Germany

*Corresponding author: patricia.wollstadt@honda-ri.de

Abstract

Entropy-based measures are an important tool for studying human gaze behavior under various conditions. In particular, gaze transition entropy (GTE) is a popular method to quantify the predictability of a visual scanpath as the entropy of transitions between fixations and has been shown to correlate with changes in task demand or changes in observer state. Measuring scanpath predictability is thus a promising approach to identifying viewers' cognitive states in behavioral experiments or gaze-based applications. However, GTE does not account for temporal dependencies beyond two consecutive fixations and may thus underestimate the actual predictability of the current fixation given past gaze behavior. Instead, we propose to quantify scanpath predictability by estimating the active information storage (AIS), which can account for dependencies spanning multiple fixations. AIS is calculated as the mutual information between a processes' multivariate past state and its next value. It is thus able to measure how much information a sequence of past fixations provides about the next fixation, hence covering a longer temporal horizon. Applying the proposed approach, we were able to distinguish between induced observer states based on estimated AIS, providing first evidence that AIS may be used in the inference of user states to improve human-machine interaction.

Keywords: eye tracking; information theory; active information storage; scanpath.

Published in: Entropy 2021, 23(2), 167; <https://doi.org/10.3390/e23020167>

1 Introduction

The analysis of visual scanning behavior provides a rich source of information in the investigation of observers' cognitive processes or states Henderson (2003); Hayhoe and Ballard (2005); Tatler *et al.* (2011); Foulsham (2015); Shiferaw *et al.* (2019). One aspect of visual scanning behavior are *scanpaths*, which denote sequences of consecutive fixations, where a fixation is a period of little to no eye movement that allows for gathering of visual information Holmqvist *et al.* (2011). The analysis of scanpaths has gained renewed interest in recent years, for example, to study cognitive function Hayes and Henderson (2017); Raptis *et al.* (2017), personality traits Baranes *et al.* (2015); Allsop *et al.* (2017), or as marker in gaze-based applications Ebeid and Gwizdka (2018). In particular, information-theoretic measures have become a popular tool for studying cognitive function through the analysis of human gaze behavior Shiferaw *et al.* (2019); Raptis *et al.* (2017); Hao *et al.* (2019); Di Stasi *et al.* (2016); Shiferaw *et al.* (2018); Schieber and Gilland (2008); Krejtz *et al.* (2014, 2015). A commonly used measure is the (*gaze*) *transition entropy* (GTE) Krejtz *et al.* (2015), which uses a conditional Shannon Entropy Shannon (1948) to describe the regularity of transitions between fixations of pre-defined areas of interest (AOI) Shiferaw *et al.* (2019). GTE considers sequences of fixations, so-called scanpaths, under the assumption that scanpaths can be modeled as Markov chains of order one, and is calculated as the entropy of the transitions between two consecutive fixations. Low GTE—and thus a low remaining uncertainty about the location of the next fixation given the previous one—thereby is interpreted as a high predictability of the next fixation Shiferaw *et al.* (2019). GTE has been applied in various studies (see Shiferaw *et al.* (2019) for a review), which have shown that changes in GTE are associated with higher task demand Di Stasi *et al.* (2016); Chanijani *et al.* (2016); Diaz-Piedra *et al.* (2019), increased anxiety Allsop and Gray (2014); van Dijk *et al.* (2011); Gotardi *et al.* (2018), or sleep deprivation Shiferaw *et al.* (2018).

Despite the popularity of information-theoretic measures, alternative approaches have been employed in the analysis of scanpaths (e.g., Simola *et al.* (2008); Hayes *et al.* (2011); Kübler *et al.* (2017); Coutrot *et al.* (2018)). One example is the work by Coutrot *et al.* (2018), who demonstrated an alternative approach to modeling scanpath data using Hidden Markov Models (HMM) in order to decode different task or observer characteristics. However, similar to GTE, their approach does not test or account for the influence of past fixations beyond the immediate last fixation. This view has however been challenged by a number of recent studies that found evidence for the importance of incorporating *long-range* temporal information when analyzing and modeling eye movement data Hayes and Henderson (2017); Hayes *et al.* (2011); Wiebel-Herboth *et al.* (2020). Here, long-range denotes any dependency of the location of the current fixation on a fixation further in the past than the immediate last fixation.

Hayes *et al.* (2011), for example, introduced the Successor Representation (SR model) to the analysis of scanpaths, which uses an algorithm

from reinforcement learning to represent sequential gaze data in a temporally extended fashion. More precisely, the SR model stems from temporal difference learning Sutton (1988); Dayan (1993) and incorporates among others a temporal parameter that defines the time span for which an observation (fixation) influences the model outcome. The authors found that up to 40 % of the variance in viewer intelligence, working memory capacities, and speed of processing could be explained based on differences in scan patterns that were individually modelled with the SR model Hayes and Henderson (2017); Hayes *et al.* (2011), as well as to some extent variances in ADHD scores (up to 50 %), autism quotients (up to 30 %), and dyslexia scores (up to 25 %) Hayes and Henderson (2018). Wiebel-Herboth *et al.* Wiebel-Herboth *et al.* (2020) found that an SR model had a significantly higher predictive power when classifying single participants based on their scanpaths in a visual comparison task compared to a simple transition matrix model that considered only the immediate last fixation. Moreover, Hoppe *et al.* Hoppe and Rothkopf (2019) were the first to provide quantitative evidence that humans are capable of planning eye movements beyond the next fixation by exploiting a priori knowledge about the task and stimulus in their experimental paradigm. While the authors did not provide a generic method for modeling scanpaths, their results provide a potential mechanism for the generation of long-range correlations in gaze behavior. Taken together, these results suggest that longer temporal dependencies in scanpaths may exist and be informative about their underlying cognitive processes and thus should be included in the modeling process.

However, entropy-based measures commonly applied, e.g., GTE Shiferaw *et al.* (2019); Krejtz *et al.* (2014, 2015), typically only take into account information contained within the immediate past fixation when quantifying the regularity of eye movements. When GTE was first introduced, Krejtz and colleagues Krejtz *et al.* (2014, 2015) adopted the procedure by Julian and Mondal (2013) for testing the Markov chain of order one assumption underlying the GTE computation. In their experiment, they found that in most but not all cases the assumption was valid. To our knowledge, such a validation procedure has not yet become a standard procedure in the entropy-based gaze analysis literature (for a review, see Shiferaw *et al.* (2019)). In cases where the order-one Markov chain assumption is violated, longer temporal dependencies are not accounted for in current information-theoretic approaches to scanpath analysis. As a result, if such temporal dependencies existed in a scanpath, the GTE would presumably underestimate its overall predictability.

Alternative modeling approaches, such as the SR model, also come with drawbacks. Most importantly, the model parameters have to be defined ad-hoc and cannot be learned in a data-driven fashion. This entails the risk of a circular argumentation if no external optimization criterion can be defined. Furthermore, the model parameters are not interpretable in a straightforward way, which limits the explanatory power of the approach. Thus, there is still a need for methods of scanpath modeling that can integrate both spatial and temporal information in the data Król and Król (2019).

To this end, we here propose a novel approach to the information-theoretic

analysis of scanpath data, which is able to measure predictability in a scanpath while accounting for temporal dependencies of arbitrary order: we propose to estimate active information storage (AIS) Lizier *et al.* (2012) from scanpaths, which measures the predictability of a sequence as the mutual information between the sequences' past and its next state. In particular, the relevant past is modeled as the collection of all past variables that provide information about the next value and can be identified using novel estimation procedures that optimize the past state in a data-driven fashion Faes *et al.* (2011); Wollstadt *et al.* (2019); Novelli *et al.* (2019).

AIS has been successfully applied in a variety of disciplines to measure predictability of time series Gómez *et al.* (2014); Brodski-Guerniero *et al.* (2017); Wollstadt *et al.* (2017); Faes *et al.* (2013); Wang *et al.* (2012); Lizier *et al.* (2012). In the context of scanpath analysis, we believe AIS can provide several benefits at once: (1) It implicitly tests the order-one Markov chain assumption, as it provides the optimized past state for a given data sample. As such, it can provide direct evidence for whether fixations beyond the last fixation have a predictive value. (2) The length of the optimized past state is directly interpretable. That is, the optimization finds the temporal horizon over which past fixation(s) are informative about and thus has the potential to support the generation of explanatory hypothesis. (3) AIS allows for an individually optimized computation of predictability that may be of greater usefulness in gaze-based applications, e.g., driver assistance Ebeid and Gwizdka (2018); Shiferaw *et al.* (2018, 2019). In sum, we argue that AIS may be applied to quantify the predictability of a scanpath, in particular, while information provided by fixations beyond the immediate fixation can be detected and accounted for.

In the following, we introduce AIS together with the necessary information-theoretic background and describe its estimation from scanpath data. As a proof of concept, we estimate AIS from scanpath data recorded in a visual comparison task under two different experimental conditions that induced different observer states. The contributions of our paper are the following: we provide the first application of AIS in the analysis of scanpath data. The AIS estimation revealed that most participants showed temporal correlations beyond the last fixation in the data. We demonstrate that changes in predictability quantified by AIS reflect changes in experimental condition that induced different observer states. Moreover, we show that the experimental condition can be successfully decoded from AIS estimates, indicating that AIS is a promising feature for identifying user states. The decoding based on AIS estimates resulted in a significantly higher accuracy than decoding from GTE, which indicates that long-term information captured by AIS but not GTE adds relevant information about the observer state in our experiment.

2 Materials and Methods

2.1 Information-Theoretic Preliminaries

Formally, we consider a scanpath as realizations $(x_1, \dots, x_t, \dots, x_N)$, $x_t \in \mathcal{A}_{X_t}$ of a stationary random process $\mathcal{X} = (X_1, X_2, \dots, X_t, \dots, X_N)$, where a random process is a collection of random variables, X , ordered by an integer $t \in \{1, \dots, N\} \subseteq \mathbb{N}$. As a shorthand, we write $p(x_t)$ for the probability, $p(X_t = x_t)$, of variable X_t taking on the value $x_t \in \mathcal{A}_{X_t}$, and $p(x_t|y)$ for the conditional probability of X_t taking on the value x_t if the outcome y of a second variable, Y , is known.

The Shannon entropy Shannon (1948) is then defined as

$$H(X) = - \sum_{x \in \mathcal{A}_X} p(x) \log p(x) \quad (1)$$

and quantifies the expected uncertainty associated with the random variable X or the amount of information to be gained when observing outcomes of X . The conditional entropy is then the average information or uncertainty remaining in X if the outcome of Y is known:

$$H(X|Y) = - \sum_{x \in \mathcal{A}_X, y \in \mathcal{A}_Y} p(x, y) \log p(x|y). \quad (2)$$

Based on these definitions, we define the mutual information (MI) as the average amount of information one variable, X , provides about a second variable, Y ,

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= \sum_{x, y} p(x, y) \log \frac{p(x|y)}{p(x)}. \end{aligned} \quad (3)$$

The MI quantifies the information X provides about Y and vice versa; it is zero for independent variables ($p(x, y) = p(x)p(y)$) or if either $H(X)$ or $H(Y)$ is zero, i.e., there is no information to share. The MI is bound from above by the entropy of both variables involved, $0 \leq I(X; Y) \leq H(X), H(Y)$ (Figure 1A).

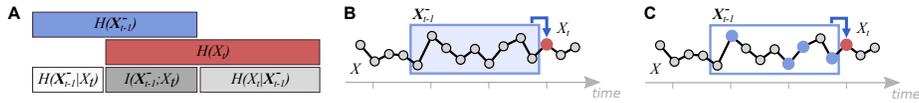


Figure 1: **(A)** Relationship between entropy and conditional entropy, H , and mutual information I of two non-independent random variables X_t and X_{t-1}^- (adapted from MacKay (2005)). Here, the conditional entropy corresponds to the gaze transition entropy (GTE). **(B)** Active information storage (AIS) quantifies the predictability of the value of time series X at time t , X_t (red marker), from its immediate past state, X_{t-1}^- (blue box). **(C)** Non-uniform embedding representing the past state of time series, X , as a selection of past variables (blue markers) up to a maximum lag $k_{max} < t$, which carry significant information about the next value, X_t (red marker).

2.2 Active Information Storage (AIS)

AIS Lizer *et al.* (2012) quantifies how much information a processes' past state \mathbf{X}_{t-1}^- contains about its next value X_t and thus measures the average predictability of X_t from its immediate past Lizer *et al.* (2012); Wibral *et al.* (2014) (Figure 1B). AIS is calculated as the MI between \mathbf{X}_{t-1}^- and X_t ,

$$\begin{aligned} AIS(X_t) &= I(\mathbf{X}_{t-1}^-; X_t) = H(X_t) - H(X_t | \mathbf{X}_{t-1}^-) \\ &= \sum_{x_t, \mathbf{x}_{t-1}^-} p(x_t, \mathbf{x}_{t-1}^-) \log \frac{p(x_t | \mathbf{x}_{t-1}^-)}{p(x_t)}, \end{aligned} \quad (4)$$

where the past state \mathbf{X}_{t-1}^- is defined as a collection of random variables up to a maximum lag k_{max} (see also the next section),

$$\mathbf{X}_{t-1}^- = \{X_{t-1}, \dots, X_{t-t_l}, \dots, X_{t-k_{max}}\}. \quad (5)$$

AIS is low for processes with highly random transitions and high for processes that visit many different states in a regular fashion Lizer *et al.* (2012); Wibral *et al.* (2014). Formally, $0 \leq AIS(X_t) \leq H(X_t), H(\mathbf{X}_{t-1}^-)$, i.e., AIS is zero for processes with no memory such that they are completely random, and the AIS is upper bounded by the entropy of the past state and entropy of the next value of a process.

Relationship between GTE and AIS

GTE measures the remaining uncertainty in a fixation, given knowledge of the previous fixation as a conditional entropy, $H(X_t | X_{t-1})$. Hence, for past states of length one, AIS and GTE are complementary, i.e., $H(X_t) = I(X_t; X_{t-1}) + H(X_t | X_{t-1})$ (Equation (4) and Figure 1A). However, for processes that do not fulfill the Markov condition, $p(X_t | X_{t-1}, \dots, X_{t-l}) = p(X_t | X_{t-1})$, i.e., processes that are not sufficiently described by a Markov chain of order one, GTE may underestimate the actual predictability of the next state from the whole *relevant* past of X (see also the next section). Furthermore, both measures differ in their interpretation—while GTE measures the remaining uncertainty in the next fixation, AIS, as a MI, measures how much information the past provides about the next fixation. The latter thus provides a more direct measure of predictability Wibral *et al.* (2014); Crutchfield and Feldman (2003).

2.3 Estimating AIS from Scanpath Data

2.3.1 Optimization of Past States

To estimate AIS in practice, we first have to define the past state \mathbf{X}_{t-1}^- such that it contains all relevant information stored in the past of X about X_t Lizer *et al.* (2012). Formally, we want to define \mathbf{X}_{t-1}^- such that $p(X_t | X_{t-1}, \dots, X_{t-l}) = p(X_t | \mathbf{X}_{t-1}^-)$. In other words, the next value, X_t , is conditionally independent

of all past variables, $X_{t-l}, l > k_{max}$, given \mathbf{X}_{t-1}^- . Non-optimal choices for \mathbf{X}_{t-1}^- may lead to an underestimation of AIS if not all relevant information is covered by \mathbf{X}_{t-1}^- , or they may lead to artificially inflated AIS values if too many variables are included, leading to an under-sampling of the past state.

Here, we find \mathbf{X}_{t-1}^- through a non-uniform embedding procedure Faes *et al.* (2011); Lizier and Rubinov (2012) that selects a subset of variables from all past variables up to a maximum lag, k_{max} (Figure 1C),

$$\mathbf{X}_{t-1}^- = \{X_{t-k}\}, k \in [1, k_{max}]. \quad (6)$$

More formally, our goal is to identify the minimal set of variables, \mathbf{X}_{t-1}^- , that provides maximum information about X_t , $\max_{\mathbf{X}_{t-1}^-} I(\mathbf{X}_{t-1}^-; X_t)$. We optimize \mathbf{X}_{t-1}^- using a greedy forward-selection approach implemented in Wollstadt *et al.* (2019), which iteratively includes variables X_{t-k} if they provide significant, additional information about X_t , conditional on all already selected variables, i.e., $\max_{X_{t-k}} I(X_{t-k}; X_t | \mathbf{X}_{t-1}^-)$ (see Faes *et al.* (2011); Wollstadt *et al.* (2019); Novelli *et al.* (2019); Lizier and Rubinov (2012) for more details). The implementation of the forward-selection approach uses a hierarchical permutation testing scheme to handle estimator bias, while controlling the family-wise error rate during repeated testing Wollstadt *et al.* (2019); Novelli *et al.* (2019). Using statistical testing for the inclusion of variables further provides an automatic stopping criterion for construction of the past state. The implementation of the AIS estimation procedure used is freely available as part of the IDTx1 open-source Python package Wollstadt *et al.* (2019), <https://github.com/pwollstadt/IDTx1>.

2.3.2 Estimating AIS from Discrete Scanpath Data

After optimizing \mathbf{X}_{t-1}^- , we estimate AIS from scanpath data using plug-in estimators Hlaváčková-Schindler *et al.* (2007) that are known to exhibit a bias due to finite sampling (e.g., Paninski (2003); Miller (1955)). Our approach to handling estimator bias is two-fold: First, we apply the bias-correction proposed in Panzeri and Treves (1996); Panzeri *et al.* (2007) and implemented in Ince *et al.* (2009) to final AIS- and entropy-estimates. Second, we use non-parametric permutation testing Novelli *et al.* (2019); Vicente *et al.* (2011) to test MI-estimates for statistical significance.

We here use permutation testing during the optimization as well as to test final AIS estimates. Permutation testing considers the MI-estimate as a test statistic in a test against the null-hypothesis of no relationship between the two variables, where the null-distribution is found through repeated estimation from permuted data Vicente *et al.* (2011). Hierarchical permutation testing is applied during the optimization of the past state, \mathbf{X}_{t-1}^- , by inclusion of past variables if they provide significant information about X_t . Note that, if no past variable provides information in a significant fashion, the inclusion terminates and no AIS is estimated. Hence, the proposed procedure tests for the presence of dependencies between past and current fixation and returns no estimate if no such dependency exists.

2.4 Experiment

As a proof of concept, we estimate AIS from eye tracking data recorded during a visual comparison task, where we varied observer states by adding a time constraint in one condition (see also Wiebel-Herboth *et al.* (2020) for details on the experimental setup).

2.4.1 Participants

We recorded data from 13 participants (one female) with a mean age of 38, ranging from 21 to 53. Data of three participants had to be excluded from the analysis due to failures in the recording process, resulting in a sample of 10 all-male participants. All participants had normal or corrected to normal sight and gave their informed written consent before participating in the experiment.

2.4.2 Task and Experimental Procedure

Participants were asked to identify as fast as possible the difference between a reference and a target image, where both images were identical except for one detail that was changed in the target image. Both images were presented next to each other on a mean gray background (Figure 2A). Participants were asked to indicate the location of the difference by clicking, using a regular computer mouse. The experiment took place in a quiet office environment under normal lighting conditions. Before the start of the experiment, participants were informed about the course of the experiment and received instructions.

Trials were recorded under two experimental conditions that were designed to induce two different user states, one relaxed state and one in which participants experienced stress through time pressure. Time pressure was achieved by varying the time available for the participants to complete the search task in each trial: In the first condition, participants had as much time as they needed (time unconstrained condition, TUC). In the second condition, time to finish the task was constraint to 9 s (time constrained condition, TC). The time limit was chosen such that it would lead to a significant performance drop and was determined in pre-tests. In addition, a sequence of nine accelerating tones, presented via headphones indicated the time running up in the TC condition. If participants did not find the difference between the images within the given time range, the next trial was initiated independently of the participant’s response. Performance dropped from 100% in the TUC condition to 65% correct trials on average in the TC condition with an average search time of $m = 4.84$ s, $sem = 0.15$ s (TUC: 100%, average search time: $m = 17.01$ s, $sem = 1.5$ s). This result indicates that the intended manipulation was indeed successful. All participants reported after the experiment that they felt under time pressure in the TC condition. For each condition, 22 trials were recorded. After half of the trials, participants were asked to take a break.

2.4.3 Apparatus and Stimuli

Stimuli were presented on a Dell monitor. Participants saw 44 photographs of varying indoor and outdoor scenes. Images were taken from a publicly available database (Shuffle database, Large Change Images) Sareen *et al.* (2016), http://search.bwh.harvard.edu/new/Shuffle_Images.html. The experimental routine was programmed in Python using Psychopy Peirce (2007, 2009). The participants' gaze behavior was recorded using a Pupil Labs eye tracker, using 120 Hz binocular gaze tracking and 60 fps world camera recordings Kassner *et al.* (2014).

The eye tracker was calibrated at the beginning of the experiment. All calibrations were done using the 9-point calibration routine implemented by Pupil Labs. Gaze points were mapped to the screen via the screen marker solution implemented by pupil labs. For that purpose, the monitor was defined as a surface based on 10 markers attached to the edge of the screen. To validate the calibration, participants were asked to fixate on a fixation dot presented at the center of the screen at the beginning of each trial (Figure 2A). The pupil labs eye tracker offers an accuracy of up to 0.6° and a precision of 0.2° . If online-computed deviations between the recorded gaze position and the fixation dot exceeded 50 px (corresponding to a viewing angle of 1.15°), the eye-tracker was recalibrated.

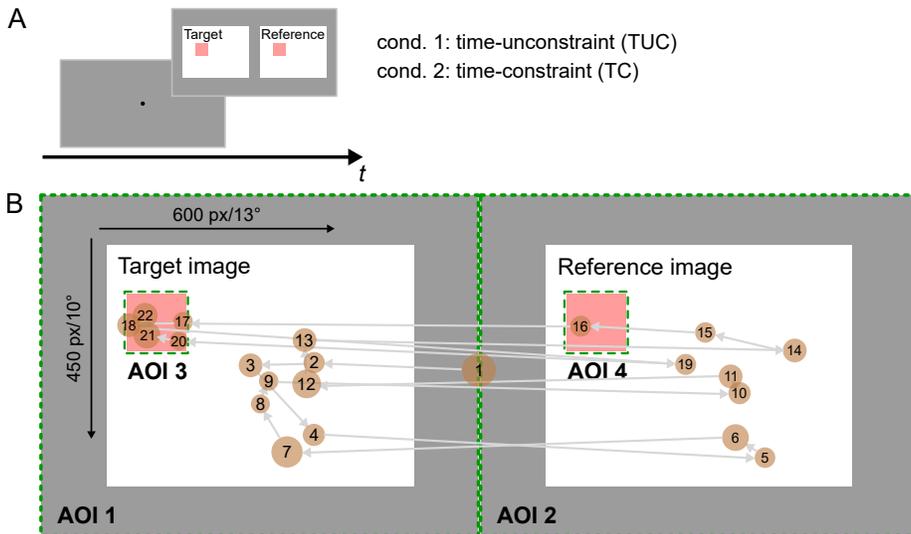


Figure 2: (A) Experimental setup for individual trial consisting of a screen showing the fixation dot and a screen displaying the image pair on a mean grey background. (B) Definition of areas of interest (AOI, green boundaries) on schematic images with target area (red box). The white line denotes an exemplary scanpath, where orange markers indicate ordered fixations and marker size corresponds to fixation time.

Images presented during a trial categorized into easy, medium and difficult with respect to the search task prior to the experiment. Ratings were done by three experimenters independently resulting in 75 % of all ratings to be 100 % consistent, whereas for the remaining 25 % (eleven images) of ratings deviated by one (e.g., easy, easy, and medium). To resolve these inconsistent cases, the median of the ratings was chosen as a label (e.g., easy). The image dataset was split in half, assuring an equal distribution of difficulty among the two. Half of the images were used for the TC condition ($n = 22$) while the other half was used for the TUC condition ($n = 22$). Within each condition, images were shown in a randomized order, with no image shown twice to the same observer.

2.4.4 Preprocessing

Data analysis was done in Python and R R Core Team (2017). Fixations for scanpath representations were computed using the basic Identification by Dispersion-Threshold (IDT) algorithm Salvucci and Goldberg (2000) using a maximum dispersion of 50 px and a minimum duration of 100 ms. Fixations above 1500 ms and data points with a confidence value below 0.9 were excluded from the data analysis. We analyzed data from all trials and did not differentiate between correct and incorrect trials.

Scanpaths were defined as sequential fixations of predefined areas of interest (AOI) and thus represent time-series data incorporating temporal as well as spatial information. We defined AOIs as four areas of interest (Figure 2B): (1) the left half of the monitor; (2) the right half of the monitor; (3) the target area in the left image; and (4) the respective target area in the right image. Target areas were defined based on the bounding boxes specifying the location of difference plus an additional frame of 50 px. Our approach aimed at extracting differences in the search process related to a presumably first “general search phase” and a “zooming in and validating phase” at the end of each trial.

3 Results

3.1 Optimization of Past States

We estimated AIS from scanpaths for each trial individually using the IDTxl Python toolbox Wollstadt *et al.* (2019). We first optimized past states, \mathbf{X}_{t-1}^- , while setting k_{max} to 5 previous fixations. This resulted in a wide variety of selected past variables over trials and participants, where, in 74 % of trials, variables with lags greater than one were selected (Figure 3A). Hence, in the majority of trials, fixations prior to the last fixation provided significant information about the next fixation and were relevant for quantifying the predictability of the scanpath. Furthermore, the variability in lags provides evidence for an intra- and inter-individual variance in viewing behavior that should be accounted for by estimation procedures.

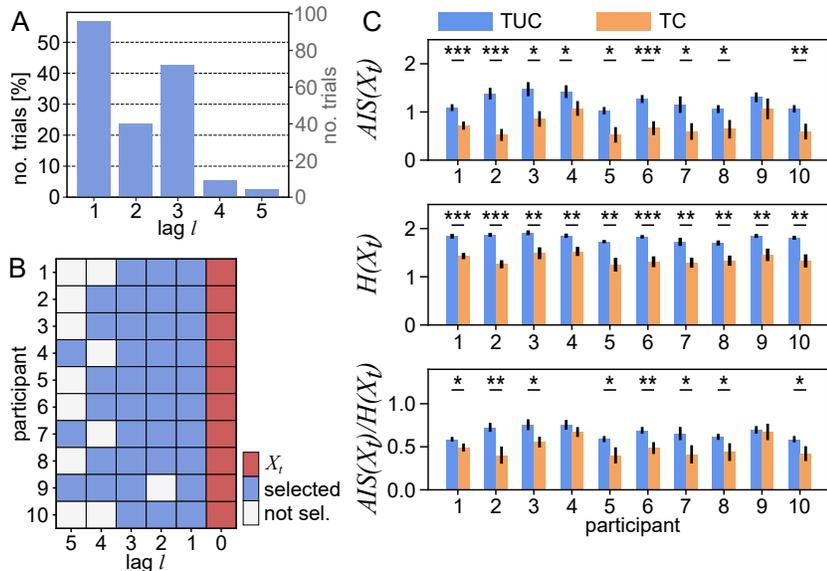


Figure 3: (A) Number of past variables with a given lag l selected through non-uniform embedding over participants and trials. (B) Union past state for each participant used for statistical testing. Selected past variables (blue) with lag l relative to the next fixation, X_t (red). (C) Mean active information storage (AIS, top), entropy (middle), and normalized AIS (bottom), for conditions (*time constraint* (TC) and *time unconstrained* (TUC)) and individual participants ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$, error bars indicate the standard error of the mean).

3.2 Difference in Experimental Conditions

3.2.1 Overall Effect of Condition on Predictability

To test for significant differences in predictability between the two experimental conditions, we fitted a linear mixed effects model with fixed effect *experimental condition* and random effect *participant*, allowing for a varying random slope for the effect of experimental condition on AIS values per participant Fahrmeir *et al.* (2007). For fitting the model, we used the lme4 package Bates *et al.* (2015), written in R R Core Team (2017).

We found a main effect of experimental condition ($\chi^2(1) = 30.054$, $p < 0.001$), while we found no significant effect of the random slope. This indicates an overall effect of experimental condition on predictability when controlling for inter-subject variability. To assess how the experimental condition affected predictability for individual participants, we performed for each participant an independent samples permutation test between AIS in both conditions ($N_{perm} = 5000$, Figure 3C). We found significantly decreased AIS in nine out of ten participants in the TC condition ($p < 0.05$, $AIS_{TC}(X_t) < AIS_{TUC}(X_t)$).

3.2.2 Relationship between AIS and Scanpath Entropy

In a second step, we investigated whether the decrease in AIS reflected a true decline in the predictability of the scanpath or whether it was rather due to a lower scanpath entropy in the TC condition. Since the absolute AIS value is bounded by the entropy of the two variables involved, a reduction in *absolute* AIS may not only be caused by a change in the predictability of a process, but also by a reduction in the processes’ entropy, i.e., a reduction in the information to be predicted.

We performed two-tailed, independent samples permutation tests for differences in $H(X_t)$ between conditions for each participant ($N_{perm} = 5000$, Figure 3), where we found a significant decline in $H(X_t)$ for the TC condition for all participants ($p < 0.05$, $H_{TC}(X_t) < H_{TUC}(X_t)$). To investigate if the decrease in $H(X_t)$ may fully explain the decrease in $AIS(X_t)$, we further tested for differences in AIS normalized by $H(X_t)$, $AIS(X_t)/H(X_t)$. Here, we found a significant decline in eight out of ten participants ($p < 0.05$, $AIS_{TC}(X_t)/H_{TC}(X_t) < AIS_{TUC}(X_t)/H_{TUC}(X_t)$). This result indicates that the observed decrease in absolute AIS may be at least partially explained by a reduction in fixation entropy, but also by an actual reduction in the regularity or predictability of the scanpath.

Note that, for all statistical comparisons, to avoid spurious effects, we aimed at holding estimation bias constant between the groups to be compared. The estimation bias depends on the number of samples and the size of the variables used Paninski (2003); Panzeri *et al.* (2007). Hence, we fixed the number of samples by discarding samples at the beginning of a trial and created a uniform past state by taking the union of selected past states over all trials and conditions (Figure 3B). Taking the union ensures that the uniform past state contains all relevant variables at the expense of including potentially irrelevant variables in the estimation from some of the trials.

3.3 Decoding of Experimental Condition from Scanpath Predictability

It is commonly assumed that eye movements are partially modulated top-down as a function of task demand (e.g., Tatler *et al.* (2011); Shiferaw *et al.* (2019); Borji and Itti (2014)). Accordingly, in eye tracking analysis, a common objective is to determine if task or observer characteristics can be decoded from eye movements (Coutrot *et al.* (2018); Borji and Itti (2014); Greene *et al.* (2012), see also the work of Yarbus *et al.* for a classic example Yarbus (1967)). To evaluate whether AIS captures relevant information about the observer state induced by our experimental manipulation, we classified the experimental condition (TC versus TUC) from estimated AIS values per image, pooled over participants. Furthermore, we estimated GTE as the conditional entropy of the current fixation given the last, $H(X_t|X_{t-1})$, for each image and compared the classification performance using AIS as input feature to the classification performance using GTE as feature. We hypothesized that, if the long term information captured by

AIS added significant information to the differentiation of the two experimentally induced observer states, then classification accuracy based on AIS should be higher compared to the classification accuracies reached by the GTE.

We followed the same classification approach as Borji and Itti Borji and Itti (2014), who successfully decoded different experimental tasks from fixation statistics recorded during viewing of images from two datasets (see also Greene *et al.* (2012)). The authors used a random undersampling boosting algorithm (RUSBoost) Seiffert *et al.* (2010) with decision trees as base models as well as a k -nearest neighbor classifier (KNN) with $k = 1$. The RUSBoost classifier handles class-imbalances in the number of valid trials entering the analysis. The results are reported as the average classification accuracy over ten repetitions of a five-fold stratified cross validation. We found that the condition could be classified above chance level (0.6725) from AIS values using both the KNN (0.7369 \pm 0.0535 SD) and RUSBoost classifier (0.6924 \pm 0.0693 SD), where the highest accuracy was achieved using the KNN classifier. In comparison, classification from GTE resulted in a lower accuracy of 0.6870 (\pm 0.0602 SD) for the KNN classifier and 0.6747 (\pm 0.0647 SD) for the RUSBoost classifier. We performed a Wilcoxon signed-rank test on the difference in KNN classification performance and found that classification based on AIS as input feature was significantly higher than classification based on GTE ($T = 884.5$, $z = 2.38$, $p < 0.01$). Our results demonstrate that AIS can be used successfully to decode our experimental condition from scanpath predictability, outperforming the classification based on the GTE, indicating that AIS captures more relevant aspects of the task-induced changes in observer state as encoded in eye movement behavior.

4 Discussion

We present AIS Lizier *et al.* (2012) as a novel approach to quantifying the predictability of scanpaths while accounting for long-range temporal dependencies between fixations. We demonstrated how to estimate AIS from scanpath data recorded during a visual comparison task and found that changes in observer states were reflected by changes in estimated AIS, indicating a lower predictability of gaze behavior in more demanding task conditions. Furthermore, we successfully decoded the experimental condition from AIS estimates of individual trials, achieving a higher accuracy compared to decoding the condition from the GTE, which was used in previous work to quantify scanpath predictability and identify differences in user states (e.g., Shiferaw *et al.* (2019)). We thus conclude that more relevant information about the observer state was encoded in the AIS estimate, which accounts for long-term dependencies in viewing behavior.

Current information-theoretic measures of predictability in scanpaths do not incorporate long-range temporal information, which may be important to accurately describe human viewing behavior Hayes and Henderson (2017); Hayes *et al.* (2011); Wiebel-Herboth *et al.* (2020); Hayes and Henderson (2018); Hoppe and Rothkopf (2019). Alternative measures, such as the SR model Hayes *et al.* (2011) or hidden Markov models Simola *et al.* (2008); Coutrot *et al.* (2018), lack

interpretability and their application is not always straightforward Shiferaw *et al.* (2019); Hayes *et al.* (2011). For example, the learning parameter representing the temporal horizon in the SR model has no clear interpretation, such that it is typically set through optimization of an additional criterion. Such an external criterion may not be readily available and may lead to circular analysis designs Kriegeskorte *et al.* (2009).

In contrast, AIS paired with novel estimation techniques, namely non-uniform embedding using a recently proposed estimation algorithm Faes *et al.* (2011); Wollstadt *et al.* (2019); Novelli *et al.* (2019), allows optimizing the temporal horizon accounted for in a purely data-driven fashion. Furthermore, the optimized past state allows for a clear interpretation in units of past samples, i.e., fixations, offering additional explanatory value. Lastly, the past state is optimized individually per participant, accounting for inter-individual variation and including cases that are best modeled by a Markov chain of order one as a special case. In the latter case, i.e., if the optimized past state contains only the past fixation with lag one, AIS and GTE are complementary such that a change in GTE corresponds to an equivalent change in AIS and vice versa. When applying AIS estimation to scanpath data, we found significant temporal relationships in scanpaths beyond first-order transitions and high inter-individual variability. Both findings underline the importance of accounting for long-range temporal dependencies as well as inter-individual differences when modeling scanpath data, in particular when quantifying the regularity or predictability of gaze behavior.

As a proof of concept, we applied the AIS estimation to eye tracking data recorded from a visual comparison task, in which two different observer states were induced. In the TC condition, participants experienced a higher task demand compared to the TUC condition. Here, we found a significant decline in predictability measured by AIS for higher task demand. This result is in line with the majority of studies utilizing GTE, which find an increase in GTE and thus lower predictability under increased task difficulty (see Shiferaw *et al.* (2019) for a review). Furthermore, we were able to decode observer states from AIS and that decoding accuracy was higher for AIS than for GTE. We conclude that AIS is able to detect changes in predictability due to changes in task demand, while avoiding methodological ad hoc choices and being more versatile with respect to temporal correlations present in gaze behavior. Furthermore, we argue that AIS provides a more immediate measure of “predictability” as its calculation incorporates the *maximum* amount of information that can be gained from the past of a process about its next state Wibral *et al.* (2014); Crutchfield and Feldman (2003). Lastly, note that we here normalized our estimate of predictability by the entropy of the next fixation, also termed *stationary gaze entropy* Shiferaw *et al.* (2019). We emphasize that such a normalization is necessary to exclude that changes in GTE or AIS are purely due to a change in fixation entropy between experimental conditions.

Being able to quantify changes in user states using approaches such as the one presented here, is central, for example, in many human machine cooperation scenarios Krüger *et al.* (2017). Being able to detect user states and changes therein

allows adapting machine behavior, e.g., to improve the interaction. Imagine for example a teaching assistance system, which, to provide optimal support for a student, must be able to assess whether a change in task demand, e.g., increasing the level of difficulty, is appropriate or overextending for the student. Only then the system can adjust to the right level of information supply or offer additional support for solving the task (see, e.g., Celiktutan and Demiris (2018)). For such an assessment of the human state, gaze behavior has been suggested as a rich data source, whose analysis can provide unobtrusive insights into a user’s cognitive or emotional state (e.g., Shiferaw *et al.* (2019); Zagermann *et al.* (2018); Majaranta and Bulling (2014)). For the analysis of gaze behavior, in particular information-theoretic measures have been suggested as promising markers of human states. We here extend existing work in this field by including previously neglected temporal correlations in the analysis of scanpath predictability, which may improve classification of user states, as demonstrated here by decoding the experimental condition from AIS. We therefore suggest AIS as a novel information-theoretic approach to the analysis of gaze behavior in user state modeling.

5 Conclusions

AIS is a promising measure for analyzing the predictability of scanpath data. Future work should extend its application to eye tracking data, for example, by exploiting the possibility to interpret AIS in a local (sample-wise) fashion Lizier *et al.* (2012). Information-theoretic quantities such as entropy or mutual information allow for an interpretation for individual realizations of the random variables involved, which allows quantifying the local entropy or local predictability of a single fixation in time. Such a localized description of fixation sequences allows for a more fine-grained quantification of gaze behavior, up to the quantification of the predictability of a single fixation. Applying AIS in its localized version thus opens the possibility of using information-theoretic measures in real-time applications such as online monitoring or assistance (e.g., Ebeid and Gwizdka (2018)). Furthermore, the application should be extended to other tasks, in particular free viewing, to gain further insights on how predictability changes as a function of the task at hand and given more natural viewing conditions. In addition, further comparison to existing scanpath analysis approaches in dedicated experiments are needed and are subject to future work. This will be a next important step to evaluate its potential for real-world gaze-based applications. In addition, note that our study is limited by its relatively small sample size and an all-male sample. Future research should therefore extend the application of AIS to larger and more diverse groups. Lastly, AIS estimation may be applied to non-discretized, raw (x, y) -fixation coordinates by foregoing the definition of AOIs and applying estimators for continuous data Kraskov *et al.* (2004); Khan *et al.* (2007). Further studies may explore these possibilities to further evaluate and extend the application of AIS to scanpath data.

Abbreviations

The following abbreviations are used in this manuscript:

AIS	Active Information Storage
GTE	Gaze Transition Entropy
IDTxl	Information Dynamics Toolkit xl
IDT	Identification by Dispersion-Threshold
MI	Mutual Information
SR Model	Successor Representation Model

References

- Henderson, J.M. Human gaze control during real-world scene perception. *Trends Cogn. Sci.* **2003**, *7*, 498–504.
- Hayhoe, M.; Ballard, D. Eye movements in natural behavior. *Trends Cogn. Sci.* **2005**, *9*, 188–194.
- Tatler, B.W.; Hayhoe, M.M.; Land, M.F.; Ballard, D.H. Eye guidance in natural vision: reinterpreting salience. *J. Vis.* **2011**, *11*, 5.
- Foulsham, T. Eye movements and their functions in everyday tasks. *Eye* **2015**, *29*, 196–199.
- Shiferaw, B.; Downey, L.; Crewther, D. A review of gaze entropy as a measure of visual scanning efficiency. *Neurosci. Biobehav. Rev.* **2019**, *96*, 353–366.
- Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; Van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*; Oxford University Press: Oxford, UK, 2011.
- Hayes, T.R.; Henderson, J.M. Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *J. Vis.* **2017**, *17*, 23, 1–17.
- Raptis, G.E.; Katsini, C.; Avouris, N.; Belk, M.; Fidas, C.; Samaras, G. Using eye gaze data & visual activities to infer human cognitive styles: Method & feasibility studies. In Proceedings of the UMAP 2017 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 9–12 July 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 164–173.
- Baranes, A.; Oudeyer, P.Y.; Gottlieb, J. Eye movements reveal epistemic curiosity in human observers. *Vis. Res.* **2015**, *117*, 81–90.
- Allsop, J.; Gray, R.; Bühlhoff, H.H.; Chuang, L. Effects of Anxiety and cognitive load on instrument scanning behavior in a flight simulation. In Proceedings of the 2nd Workshop on Eye Tracking and Visualization (ETVIS) 2016, Baltimore, MD, USA, 23–28 October 2016; pp. 55–59.

- Ebeid, I.A.; Gwizdka, J. Real-Time Gaze Transition Entropy. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Warsaw, Poland, 14–17 June 2018; p. 94.
- Hao, Q.; Sbert, M.; Ma, L. Gaze information channel in cognitive comprehension of poster reading. *Entropy* **2019**, *21*, 1–24.
- Di Stasi, L.L.; Diaz-Piedra, C.; Rieiro, H.; Carrión, J.M.S.; Berrido, M.M.; Olivares, G.; Catena, A. Gaze entropy reflects surgical task load. *Surg. Endosc.* **2016**, *30*, 5034–5043.
- Shiferaw, B.A.; Downey, L.A.; Westlake, J.; Stevens, B.; Rajaratnam, S.M.W.; Berlowitz, D.J.; Swann, P.; E., H.M. Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *Sci. Rep.* **2018**, *8*, 2220.
- Schieber, F.; Gilland, J. Visual entropy metric reveals differences in drivers' eye gaze complexity across variations in age and subsidiary task load. *Proc. Hum. Factors Ergon. Soc.* **2008**, *3*, 1883–1887.
- Krejtz, K.; Szmidt, T.; Duchowski, A.T.; Krejtz, I. Entropy-based statistical analysis of eye movement transitions. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA), Safety Harbor, CA, USA, 26–29 March 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 159–166.
- Krejtz, K.; Duchowski, A.; Szmidt, T.; Krejtz, I.; Perilli, F.G.; Pires, A.; Vilaro, A.; Villalobos, N. Gaze transition entropy. *ACM Trans. Appl. Percept.* **2015**, *13*, 4.
- Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
- Chanijani, S.S.M.; Klein, P.; Bukhari, S.S.; Kuhn, J.; Dengel, A. Entropy based transition analysis of eye movement on physics representational competence. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp): Adjunct, Heidelberg, Germany, 12–16 September 2016; ACM: New York, NY, USA, 2016; pp. 1027–1034.
- Diaz-Piedra, C.; Rieiro, H.; Cherino, A.; Fuentes, L.J.; Catena, A.; Di Stasi, L.L. The effects of flight complexity on gaze entropy: An experimental study with fighter pilots. *Appl. Ergon.* **2019**, *77*, 92–99.
- Allsop, J.; Gray, R. Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *J. Appl. Res. Mem. Cogn.* **2014**, *3*, 63–71.
- van Dijk, H.; van de Merwe, K.; Zon, R. A coherent impression of the pilots' situation awareness: studying relevant human factors tools. *Int. J. Aviat. Psychol.* **2011**, *21*, 343–356.

- Gotardi, G.; Schor, P.; Van Der Kamp, J.; Navarro, M.; Orth, D.; Savelsbergh, G.; Polastri, P.F.; Oudejans, R.; Rodrigues, S.T. The influence of anxiety on visual entropy of experienced drivers. In Proceedings of the 3rd Workshop on Eye Tracking and Visualization (ETVIS), Warsaw, Poland, 15 June 2018; ACM: New York, NY, USA, 2018; pp. 1–4.
- Simola, J.; Salojärvi, J.; Kojo, I. Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cogn. Syst. Res.* **2008**, *9*, 237–251.
- Hayes, T.R.; Petrov, A.A.; Sederberg, P.B. A novel method for analyzing sequential eye movements reveals strategic influence on Raven’s Advanced Progressive Matrices. *J. Vis.* **2011**, *11*, 10.
- Kübler, T.C.; Rothe, C.; Schiefer, U.; Rosenstiel, W.; Kasneci, E. SubMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behav. Res. Methods* **2017**, *49*, 1048–1064.
- Coutrot, A.; Hsiao, J.H.; Chan, A.B. Scanpath modeling and classification with hidden Markov models. *Behav. Res. Methods* **2018**, *50*, 362–379.
- Wiebel-Herboth, C.B.; Krüger, M.; Hasenjäger, M. Interactions between inter- and intra-individual effects on gaze behavior. In Proceedings of the Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP), Genoa, Italy, 12–18 July 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 35–40.
- Sutton, R.S. Learning to predict by the methods of temporal differences. *Mach. Learn.* **1988**, *3*, 9–44.
- Dayan, P. Improving generalization for temporal difference learning: The successor representation. *Neural Comput.* **1993**, *5*, 613–624.
- Hayes, T.R.; Henderson, J.M. Scan patterns during scene viewing predict individual differences in clinical traits in a normative sample. *PLoS ONE* **2018**, *13*, e0196654.
- Hoppe, D.; Rothkopf, C.A. Multi-step planning of eye movements in visual search. *Sci. Rep.* **2019**, *9*, 1–12.
- Julian, B.; Mondal, D. Exact goodness-of-fit tests for Markov chains. *Biometrics* **2013**, *69*, 488–496.
- Król, M.; Król, M.E. A novel eye movement data transformation technique that preserves temporal information: A demonstration in a face processing task. *Sensors* **2019**, *19*, 2377.
- Lizier, J.T.; Prokopenko, M.; Zomaya, A.Y. Local measures of information storage in complex distributed computation. *Inf. Sci.* **2012**, *208*, 39–54.

- Faes, L.; Nollo, G.; Porta, A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Phys. Rev. E* **2011**, *5*, 051112.
- Wollstadt, P.; Lizier, J.T.; Vicente, R.; Finn, C.; Martínez-Zarzuela, M.; Mediano, P.A.M.; Novelli, L.; Wibral, M. IDTxl: The Information Dynamics Toolkit xl: a Python package for the efficient analysis of multivariate information dynamics in networks. *J. Open Source Softw.* **2019**, *4*, 1081.
- Novelli, L.; Wollstadt, P.; Mediano, P.A.M.; Wibral, M.; Lizier, J.T. Large-scale directed network inference with multivariate transfer entropy and hierarchical statistical testing. *Netw. Neurosci.* **2019**, *3*, 827–847.
- Gómez, C.; Lizier, J.T.; Schaum, M.; Wollstadt, P.; Grützner, C.; Uhlhaas, P.; Freitag, C.M.; Schlitt, S.; Bölte, S.; Hornero, R.; others. Reduced predictable information in brain signals in autism spectrum disorder. *Front. Neuroinformatics* **2014**, *8*, 9.
- Brodski-Guerniero, A.; Paasch, G.F.; Wollstadt, P.; Özdemir, I.; Lizier, J.T.; Wibral, M. Information-theoretic evidence for predictive coding in the face-processing system. *J. Neurosci.* **2017**, *37*, 8273–8283.
- Wollstadt, P.; Sellers, K.K.; Rudelt, L.; Priesemann, V.; Hutt, A.; Fröhlich, F.; Wibral, M. Breakdown of local information processing may underlie isoflurane anesthesia effects. *PLoS Comput. Biol.* **2017**, *13*, e1005511.
- Faes, L.; Porta, A.; Rossato, G.; Adami, A.; Tonon, D.; Corica, A.; Nollo, G. Investigating the mechanisms of cardiovascular and cerebrovascular regulation in orthostatic syncope through an information decomposition strategy. *Auton. Neurosci.* **2013**, *178*, 76–82.
- Wang, X.R.; Miller, J.M.; Lizier, J.T.; Prokopenko, M.; Rossi, L.F. Quantifying and Tracing Information Cascades in Swarms. *PLoS ONE* **2012**, *7*, e40084.
- Lizier, J.T.; Prokopenko, M.; Zomaya, A.Y. Coherent information structure in complex computation. *Theory Biosci.* **2012**, *131*, 193–203.
- Shiferaw, B.A.; Crewther, D.P.; Downey, L.A. Gaze entropy measures detect alcohol-induced driver impairment. *Drug Alcohol Depend.* **2019**, *204*, 107519.
- MacKay, D.J.C. *Information Theory, Inference, and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2005.
- Wibral, M.; Lizier, J.T.; Vögler, S.; Priesemann, V.; Galuske, R. Local active information storage as a tool to understand distributed neural information processing. *Front. Neuroinformatics* **2014**, *8*, 1–11.
- Crutchfield, J.P.; Feldman, D.P. Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos* **2003**, *13*, 25–54.

- Lizier, J.T.; Rubinov, M. Multivariate Construction of Effective Computational Networks from Observational data. Preprint no.: 25/2012, Max Planck Institute for Mathematics in the Sciences, 2012. Available online: <https://www.mis.mpg.de/publications/preprints/2012/prepr2012-25.html> (accessed on 15 December 2012).
- Hlaváčková-Schindler, K.; Paluš, M.; Vejmelka, M.; Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* **2007**, *441*, 1–46.
- Paninski, L. Estimation of entropy and mutual information. *Neural Comput.* **2003**, *15*, 1191–1253.
- Miller, G. Note on the bias of information estimates. In *Information Theory in Psychology II-B*; Quastler, H., Ed.; Free Press: Glencoe, IL, USA, 1955; pp. 95–100.
- Panzeri, S.; Treves, A. Analytical estimates of limited sampling biases in different information measures. *Netw. Comput. Neural Syst.* **1996**, *7*, 87–107.
- Panzeri, S.; Senatore, R.; Montemurro, M.A.; Petersen, R.S. Correcting for the sampling bias problem in spike train information measures. *J. Neurophysiol.* **2007**, *98*, 1064–1072.
- Ince, R.A.A.; Petersen, R.S.; Swan, D.C.; Panzeri, S. Python for information theoretic analysis of neural data. *Front. Neuroinformatics* **2009**, *3*, 4.
- Vicente, R.; Wibral, M.; Lindner, M.; Pipa, G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **2011**, *30*, 45–67.
- Sareen, P.; Ehinger, K.A.; Wolfe, J.M. CB Database: A change blindness database for objects in natural indoor scenes. *Behav. Res. Methods* **2016**, *48*, 1343–1348.
- Peirce, J.W. PsychoPy – Psychophysics software in Python. *J. Neurosci. Methods* **2007**, *162*, 8–13.
- Peirce, J.W. Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinformatics* **2009**, *10*, 1–8.
- Kassner, M.; Patera, W.; Bulling, A. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp): Adjunct Publication; ACM: New York, NY, USA, 13–17 September 2014; pp. 1151–1160.
- R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.

- Salvucci, D.D.; Goldberg, J.H. Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 Symposium on Eye Tracking Research and Applications (ETRA), Palm Beach Gardens, FL, USA, 6–8 November 2000; ACM: New York, NY, USA, 2000; pp. 71–78.
- Fahrmeir, L.; Kneib, T.; Lang, S.; Marx, B. *Regression*; Springer: Berlin/Heidelberg, Germany, 2007.
- Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48.
- Borji, A.; Itti, L. Defending Yarbus: Eye movements reveal observers’ task. *J. Vis.* **2014**, *14*, 1–22.
- Greene, M.R.; Liu, T.; Wolfe, J.M. Reconsidering Yarbus: A failure to predict observers’ task from eye movement patterns. *Vis. Res.* **2012**, *62*, 1–8.
- Yarbus, A.L. *Eye Movements and Vision*; Plenum: New York, NY, USA, 1967.
- Seiffert, C.; Khoshgoftaar, T.M.; Van Hülse, J.; Napolitano, A. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man, Cybern. Part A Syst. Humans* **2010**, *40*, 185–197.
- Kriegeskorte, N.; Simmons, W.K.; Bellgowan, P.S.F.; Baker, C.I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **2009**, *12*, 535–540.
- Krüger, M.; Wiebel, C.B.; Wersing, H. From tools towards cooperative assistants. In Proceedings of the 5th International Conference on Human Agent Interaction (HAI), Bielefeld, Germany, 17–20 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 287–294.
- Celiktutan, O.; Demiris, Y. Inferring human knowledgeability from eye gaze in mobile learning environments. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
- Zagermann, J.; Pfeil, U.; Reiterer, H. Studying eye movements as a basis for measuring cognitive load. In Proceedings of the Extended Abstracts of the 2018 Conference on Human Factors in Computing Systems (CHI), Montreal, QC, Canada, 21–26 April 2018; pp. 1–6.
- Majoranta, P.; Bulling, A. Eye tracking and eye-based human-computer interaction. In *Advances in Physiological Computing*; Springer: London, United Kingdom, 2014; pp. 39–65.
- Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 16.
- Khan, S.; Bandyopadhyay, S.; Ganguly, A.R.; Saigal, S.; Erickson III, D.J.; Protopopescu, V.; Ostrouchov, G. Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Phys. Rev. E* **2007**, *76*, 026209.