Subject identification through Neural SDEs on human gaze trajectories

Mattia Vinciguerra and Ivan Coppola

Department of Computer Science, University of Study of Milan.

Abstract

Several studies have demonstrated that biometric identification based on human gaze trajectories is possible and may be used for authentication. Eye movements provide significant and valuable information about individual behavioral patterns and cognitive processes. The basic idea is that each individual can be identified through their characteristic way of performing fixations, which are small eye micro-movements around a local area of interest that attracts the viewer's focus. Modeling these trajectories represents a complex challenge, as fixations are stochastic, variable in length and duration, and influenced by both perceptual and cognitive factors. **Stochastic Differential Equations (SDEs)** provide a mathematical framework for describing the temporal evolution of systems influenced by both deterministic forces (drift) and stochastic fluctuations (diffusion). Human eye movements exhibit inherent randomness due to neural noise, perceptual variability, and environmental influences. The diffusion term in SDEs explicitly incorporates this randomness, allowing models to reflect the uncertainty and variability observed in real fixation data. While classical SDEs require predefined functional forms for drift and diffusion, **Neural SDEs** extend this framework by parameterizing these functions with neural networks. The results obtained from the experiments presented in this article demonstrate that stochastic deep learning approaches are capable of capturing meaningful temporal dynamics of eye movements, achieving accuracy values above the random-guessing baseline and enabling effective subject identification. The findings contribute to both machine learning and cognitive science, suggesting promising directions for future applications in biometrics, human–computer interaction, and neurological assessment.

GitHub repository

1 Introduction

The study and the analysis of eye movements have been measured since the early 1900 [1, 2] and it has been a wide area of study in various fields. In the last decades human vision attention models have been employed to solve problems and tasks in various fields like image and video processing, automatic captioning, visual question answering, and language understanding, as well as robotics, autonomous driving, and medicine [3, 4]. Since the first pubblication of Itti et al. model [5] in 1998 based on saliency maps and originally published for psychology and neuroscience purposes, this model has been increasing interest in computer vision and pattern recognition community [3]. Biometrics deals with the exploitation of unique, identifiable and quantitatively measurable characteristics of humans in order to authenticate and/or identify them [6].

1.1 Biometrics

Biometric person authentication refers to identifying or verifying persons' identity based on their physical and/or behavioral (learned) characteristics [7]. Biometric identification has become a rapidly expanding research field, offering reliable methods for authentication and security [8, 9], and it has been spreading in humans' life in recent years. For example, it is easy to think that nowadays finger and face recognition are some of the most used unlocking methods for common electronic devices, such as smartphones, laptops, personal devices and so on. Traditionally, biometrics methods have focused on personal physical traits such as fingerprints [10, 11], face [12, 13], iris and retina [14]. Among the possible biometric traits, eye movements have attracted increasing attention due to their inherent link with cognitive processes and their resistance to imitation. In particural, this kind of biometric is based on individual behavioural processes and internal patterns that might be unique to a user [8]. Therefore, these methods result in a less intrusive way to identify users compared to more physical biometrics traits [6].

1.2 Gaze trajectories

Human gaze trajectories refer to the continuous path traced by the point of gaze over discrete time samples as invididuals explore their visual environment. These gaze trajectories derive from eye tracking devices, methods and experiments that can provide more objective insights into brain function and cognitive processing by capturing non-verbal and unconscious responses [15]. This trajectories represent a complex interplay between bottom-up sensory factors and top-down cognitive control mechanisms. These continuous paths are characterized by periods of relative stability (fixations) interrupted by rapid movements (saccades), creating a discontinuous yet predictable pattern that reflects the underlying neural control mechanisms. The contemporary eye movement literature is predominantly structured around two cardinal events: fixations and saccades. These constitute discrete temporal epochs, characterized by distinct onset, offset, and duration parameters, into which the continuous eye-tracker signal is systematically segmented. This temporal parsing may be augmented by additional oculomotor events, including blinks and smooth pursuit movements, depending on the specific analytical framework employed [16]. The conceptualization of fixations exhibits considerable variability across the research literature, reflecting fundamentally different theoretical approaches to understanding this oculomotor phenomenon. One prominent definitional framework characterizes fixations at a functional level [17, 18], emphasizing their role in visual information processing. Within this functional perspective, Gegenfurtner et al. [19] provide additional mechanistic detail by explicating how fixational function is achieved through the orchestrated action of miniature eye movements, including microsaccades, drift, and tremor. An alternative definitional approach characterizes fixations primarily through oculomotor descriptors, specifically defining them as temporal periods during which the eye remains relatively stationary [20, 21]. This definition prioritizes the kinematic properties of eye position rather than the functional outcomes of the fixational state. Similarly, saccades are subject to multiple interpretations. They can be defined functionally [2, 17, 18, 22–24], or characterized in terms of their oculomotor properties [25, 26]. This plurality of definitions introduces certain ambiguities. For instance, fixation may be conceptualized either as a functional process of maintaining a target on the fovea or as a state of ocular stillness relative to the head. These perspectives may lead to apparently contradictory classifications. Consider the case of a participant fixating on an object while moving around a room: under functional definitions, this qualifies as a fixation, yet under the "eye-stillness" definition it does not, since the eyes rotate with respect to the head. The crucial factor here lies in the frame of reference. Stillness of the eye relative to the head only occurs when the fixation target remains stationary with respect to the head. If the head moves relative to the environment while the gaze remains anchored to a world-fixed point, the eyes will rotate relative to the head but still achieve functional stabilization on the fovea.

Saccades are also often described as the intervals between fixations [20, 21]. However, this definition directly depends on how fixation itself is understood. It also leads to two possible scenarios. In the first, any non-saccadic activity (including smooth pursuit and vestibulo-ocular reflex, VOR) is classified as fixation, which aligns with the functional perspective that treats all forms of retinal image stabilization as fixation. In the second, the interval between saccades is considered fixation only if smooth pursuit and VOR are excluded. This latter scenario typically arises in studies where eye trackers provide world-centered gaze coordinates and fixation targets are stationary, thereby eliminating pursuit or VOR episodes. Given their reliance on an "eye-stillness" definition of fixation, the latter interpretation is likely to apply to Holmqvist et al. [20] and Larsson et al. [21], whereas the former interpretation better reflects the perspective of Hessels et al. [16] and Jovancevic-Misic and Hayhoe [27]. Saccades are thus functionally defined as the fast movements that redirect gaze to a new part of the surroundings; each fixation summarises fixational movements that occur within intervals between saccades, in which gaze is held almost stationary, so as to keep onto the fovea (the central part of the retina) the circumscribed region of interest (RoI) within the viewed scene.

1.3 Individual uniqueness

An ideal biometric modality should satisfy several key criteria: it must be unique to each individual, universally applicable, stable over time, easy to measure, cost-effective, and widely acceptable to users. No single biometric meets all these requirements perfectly. For example, fingerprints and retinal patterns are highly distinctive, yet they necessitate specialized sensors and can be perceived as intrusive by users. In contrast, modalities such as voice or facial geometry are easier to capture with inexpensive and unobtrusive devices, such as microphones or cameras, but their discriminative power is generally lower. Prior research has shown that combining multiple complementary biometric traits can significantly improve recognition accuracy compared to relying on a single modality [28, 29].

The dynamics of gaze trajectories reveal that the selection of subsequent fixation points is neither entirely deterministic nor purely random [30]. In such contexts, visual attention tends to be directed predominantly toward semantically relevant elements of a scene [4, 31], such as human faces.

The use of behavioral biometrics, and in particular the analysis of eye movements for subject identification, is based on deeper reasons than its apparent and mainly practical application [6]. The key idea is that every individual differs in subtle ways from others, differences that can also be described in terms of physical complexity [32–35]. Over time, each person develops a unique identity shaped by personal experiences. This process involves continuously adapting to the environment and building internal models that represent both the outside world and one's role within it. Actions are part of this adaptive process and therefore remain deeply personal and unique [34, 36–38]. Eye movements, and especially gaze patterns, reflect these individual characteristics as well, showing distinctive traits for each person [34, 39].

This paper focuses on classifying users based on their fixation patterns extracted from recorded gaze trajectories in the FIFA Dataset from Moran Cerf's research [4, 31]. The dataset is available at: dataset. The Fixations in Faces (FIFA) dataset was originally created by Cerf and colleagues in their study "Predicting Human Gaze Using Low-Level Saliency Combined with Face Detection" [4, 31]. This dataset contains eye-tracking data from participants who freely viewed natural images. The original purpose was to study how people tend to look at faces and facial features in images. The researchers wanted to create a standard dataset for testing models that combine visual saliency with face detection to predict where people look. In our work, we use this dataset differently. Instead of just predicting where people look, we use it to identify who is looking. We take advantage of the fact that each person has unique fixation patterns, their own way of looking at images. This individual signature in how people move their eyes allows us to classify and identify different observers based on their gaze behavior. The dataset provides a valuable resource for this type of analysis because it contains natural viewing behavior from multiple participants looking at the same set of images. This allows us to study the individual differences in gaze patterns while controlling for the visual content that participants were viewing.

1.3.1 The foraging theory in gaze biometrics

Individual-based approaches in ecology aim to explain how variation among single organisms shapes broader patterns at population, community, and ecosystem levels

[40]. Individuality in behavioural ecology is defined by the combination of phenotypic traits and ecological interactions that make each organism unique. Core aspects include behaviour, habitat use, feeding preferences, and social relations [40].

A paradigmatic case is foraging behaviour, which refers to how animals search for and consume resources [41, 42]. According to optimal foraging theory, individuals continuously balance choices about where to search, what to consume, and when to leave depleted patches [43]. This process creates a characteristic exploration–exploitation cycle, where local resource use alternates with relocation across the environment.

Research has shown consistent inter-individual differences in foraging and resource use, described as animal personality and individual specialization [44]. Personality refers to stable behavioural traits such as exploration, activity, sociability, or boldness, whereas specialization captures persistent differences in diet. Both reflect individuality that remains stable over time.

Recent perspectives suggest that ancestral strategies for food foraging evolved into cognitive forms of information foraging [45–49]. In this view, eye movements represent the primary mean of sampling visual information [39, 50–53]. At the psychological level, foraging has been linked to spatial memory, value-based decision making, and executive control [45]. At the neural level, dopaminergic activity is thought to regulate the balance between focused exploitation and exploratory behaviour [49], in line with active inference models of perception and action [39, 54].

In the context of visual behaviour, this framework describes gaze as a sequence of alternating saccades and fixations. Saccades are rapid eye movements that redirect gaze toward new locations, while fixations are the relatively stable intervals between saccades in which the fovea is aligned with a region of interest in the scene. This explore–exploit pattern, observable in eye movement trajectories, provides a robust analogy with animal foraging [39, 51–53]. Importantly, such dynamics are shaped not only by the visual stimulus and task demands but also by internal states and affective processes, which influence how visual information, including socially relevant cues such as emotions and intentions, is prioritized [55–57].

1.4 Ornstein–Uhlenbeck Process for Fixation Modeling

The Ornstein–Uhlenbeck (OU) process is a mean-reverting Gaussian Markov process that describes stochastic dynamics attracted toward a stable equilibrium. In $\mathbb{R}^d$, an OU process $\{\mathbf{z}(t)\}_{t\geq 0}$ satisfies the linear SDE

$$d\mathbf{z}(t) = -\boldsymbol{\Theta}\big(\mathbf{z}(t) - \boldsymbol{\mu}\big)\,dt + \boldsymbol{\Sigma}^{1/2}\,d\mathbf{w}(t),$$

where $\boldsymbol{\Theta} \in \mathbb{R}^{d\times d}$ is a positive-definite mean-reversion matrix, $\boldsymbol{\mu} \in \mathbb{R}^d$ is the long-run mean (equilibrium), $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is the diffusion covariance, and $\mathbf{w}(t)$ is standard Brownian motion. The OU process is stationary, Gaussian, and ergodic under mild conditions (e.g., $\boldsymbol{\Theta} \succ 0$), with closed-form transition densities [58, 59].

Modeling fixations as an OU process encodes three inductive biases that are well-aligned with human gaze during a fixation: (i) mean reversion around a reference location (the fixation point), (ii) Gaussian, temporally correlated perturbations due to micro-saccades, tremor, drift, and measurement noise, and (iii) continuous-time dynamics that can be evaluated on irregular sampling grids.

1.5 Neural Stochastic Differential Equations (Neural SDEs)

In recent years, several authors have introduced the concept of Neural Stochastic Differential Equations (Neural SDEs) as a natural extension of Stochastic Differential Equations (SDEs), with notable contributions by Tzen and Raginsky (2019) [60], Li et al. (2020) [61], and Hodgkinson et al. (2020) [62]. The dynamics of an SDE consist of a deterministic term and a stochastic term. The main idea is to parameterize the drift and diffusion terms of a differential equation through neural networks. This solution combines two dominant paradigms in modern modeling: the expressive power of deep learning and the mathematical rigor of continuous-time dynamical systems. This approach builds upon the long-standing principle of fitting parameterized differential equations to data, typically optimized through stochastic gradient descent [63]. The integration of neural networks into this framework provides high-capacity, easily trainable function approximators, which allow complex dynamics to be learned directly from data in an **end-to-end** manner. Stochastic Differential Equations themselves have long been employed to describe systems evolving in continuous time under uncertainty. Their applications span a wide range of domains, including particle systems [64–66], financial modeling [67, 68], population dynamics [69, 70], and genetics [71]. By generalizing Ordinary Differential Equations (ODEs) with stochastic components, SDEs provide a flexible mathematical tool for capturing real-world random phenomena [72]. In fact, they are a natural extension of ODEs for modelling systems that evolve in continuous time subject to uncertainty [72].

As mentioned before, the modeling of human gaze trajectories, particularly fixation patterns and saccadic movements, represents a significant challenge in computational neuroscience and human-computer interaction. Fixations, defined as small and relatively stable eye micro-movements occurring around a visual target, provide rich temporal sequences that are distinctive for each individual [6]. Modeling these fixation trajectories is particularly challenging. They are inherently stochastic, variable in length and duration, and strongly influenced by perceptual and cognitive factors [6]. Traditional approaches based on discrete-time models, such as recurrent neural networks, often struggle to capture the irregularity and variability present in real data [73], thus limiting their effectiveness in subject classification tasks. To address these challenges, this study explores the use of **Stochastic Differential Equations (SDEs)** and their Neural extensions for modeling eye fixation trajectories. SDEs provide a continuous-time framework capable of capturing both deterministic dynamics (drift) and stochastic variability (diffusion), making them especially suitable for irregular behavioral data [73]. In particular, variants such as Neural Langevin-type SDEs (LSDE) and Neural Linear Noise SDEs (LNSDE) are investigated [73]. These models allow for expressive and stable latent representations of gaze dynamics, which can be exploited for subject identification.

2 State of the art

Two early studies explored the use of physical features of eye movements for person recognition [74, 75].

The initial exploration of eye movements for biometric purposes was undertaken by the work of Kasprowski and Ober [74]. In their study, classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), decision trees, and Bayesian models were applied to selected cepstral coefficients derived from eye movement signals, yielding promising results. However, their analysis was limited to raw trajectory signals, without explicitly classifying eye movements into physiologically meaningful events. They employed a custom-built head-mounted infrared oculography eye tracker ("OBER2") operating at 250 Hz. Data were collected from $N = 47$ participants, who were asked to follow a jumping stimulus point presented on a standard monitor across twelve successive positions. Each trial lasted less than 10 seconds and produced a fixed-length signal of 2048 samples. To process the recordings, the raw gaze coordinates were normalized and aligned using the stimulus reference signal combined with robust fixation detection. From these normalized signals, several features were extracted, including average velocity direction, distance to the stimulus, inter-eye differences, and spectral representations obtained through Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT). Classification was performed using five standard pattern-matching techniques, as well as their ensemble combination. The ensemble classifier achieved an average recognition error between 7% and 8%.

A second independent study [75] used a commercially available eye tracker (Tobii ET-1750) integrated into a TFT display, sampling at 50 Hz. Data were collected from $N = 12$ subjects while they viewed a simple stimulus: a cross displayed at the center of the screen for one second. In this case, the extracted features included pupil diameter and its temporal derivative, gaze velocity, and the time-varying interocular distance. Feature reduction was performed using DFT, Principal Component Analysis (PCA), and their combination, followed by template matching with a k-nearest neighbors classifier (KNN). The most effective feature was the time derivative of the pupil size, which provided an identification rate between 50% and 60%. The interocular distance, while yielding the best discriminative performance, was excluded as it could be obtained without the use of an eye tracker and therefore was not considered a meaningful biometric feature.

Both studies [74, 75] are inherently task-dependent, as they assume that the same stimulus appears in both the training and testing phases. This setup has the advantage that the reference template (training sample) and the test sample can be accurately aligned. However, this accuracy comes at a cost to convenience and usability. In a security scenario, the user must perform a specific task, which can become learned over time and potentially distracting. If the user internalizes the task—for instance, the order of a moving stimulus—their behavior may change, potentially increasing recognition errors. From a security perspective, repetitive tasks are also easier to imitate, allowing an intruder to mimic the expected input. Various experiments have been considered this stimulus-dependant scenarios: reading [13, 14], static point gazing [75] or jumping point tracking [74–76]. Subsequent studies approached eye movements at a higher conceptual level. In [75], features associated with fixations and saccades were extracted and input into a probabilistic neural network to generate match scores. Similarly, Holland and Komogortsev [77] derived a broader set of high-level features, which were then compared using a Gaussian kernel. Additionally, graph-based models have been proposed, demonstrating competitive performance [78]. Within the framework of hand-crafted features, the STAR method introduced by Friedman et al. [79] is widely regarded as a state-of-the-art approach.

In recent years, deep learning approaches have achieved the best performance in eye movement biometrics, leveraging architectures such as convolutional neural networks

(CNNs) [80–82], residual networks, and densely connected convolutional networks [83]. Jia et al. [80] proposed a recurrent neural network (RNN) with long short-term memory (LSTM) cells that operates directly on raw gaze data. In contrast, the approach by Jäger et al. [82] involves a preprocessing step that transforms raw eye movement signals to isolate micro eye movements based on characteristic velocities, with the resulting scaled signals input to a CNN composed of two separate subnets. Other studies have explored metric learning for eye movement biometrics following an initial feature extraction stage [83, 84].

The current state-of-the-art model, DeepEyedentificationLive (DEL) [82], separates "fast" (e.g., saccadic) and "slow" (e.g., fixational) eye movements, feeding them into two distinct convolutional subnets. Another recent architecture, Eye Know You (EKY) [83], employs exponentially dilated convolutions to achieve strong biometric performance with a relatively compact network (475k learnable parameters).

Overall, these approaches largely rely on classical analyses of gaze trajectories, such as fixation dwell times, distances and direction changes between fixations, and saccade spatial distributions [85]. This reliance persists regardless of the subsequent classification method, whether traditional or deep learning-based. Consequently, much of the complexity inherent in gaze behavior is often neglected, resulting in an underutilization of information that could reveal the subtle uniqueness of individual eye movement patterns [6].

3 Methodology

This chapter present the definition of the implemented pipeline for fixation modeling and subject classification using Neural SDE-based architectures. It also includes a benchmarking study comparing different models in terms of predictive accuracy and loss.

In this work, the focus is on eye movement biometrics that require minimal or no prior knowledge of the underlying task, hereafter referred to as **task-independent** eye movement biometrics. The dataset employed in this study is derived from a free-viewing image task, and the model relies exclusively on raw gaze data. The objective is to investigate whether Neural SDE models can effectively capture the uniqueness of individuals based solely on their gaze dynamics, and to assess whether the combination of deterministic and stochastic diffusion terms can represent distinctive patterns and behaviours specific to each user.

3.1 Dataset

The experiments in this work are conducted on the Fixations In Faces (FiFa) database by Moran Cerf et al. [4]: https://qualinet.github.io/databases/image/fixations_in_faces_fifa_database/. This dataset collects the eye-movement behavior of human observers during free-viewing of images. In particular, it contains the scanpaths of 8 subjects during free-viewing of 200 images, with each image displayed for 2 seconds. Hence, for each subject it collects 200 scanpath recordings, and each one consists of a sequence of 2021 gaze x and y coordinates.

3.2 Preprocessing

Raw gaze data streams from the FiFa database are processed to obtain clean sequences of fixation events. Each scanpath (sequence of gaze samples) is segmented into discrete events, either "fixations" (periods of relatively stable gaze) or "saccades" (rapid eye movements), using the I-VT (Velocity-Threshold Identification) algorithm [86]. Briefly, I-VT operates as follows:

- The gaze velocity between consecutive samples is calculated.
- Samples with velocity above a threshold are labeled as saccades, those below are labeled as part of a fixation.
- Output is a sequence of fixation events (with position and duration) punctuated by saccades.

In this work, only fixation events are retained for modeling and classification.

3.2.1 Centering fixations for OU modeling.

After event classification with I-VT, fixation coordinates are translated to be zero-centered by subtracting, for each scanpath, the empirical mean across all fixation points. In this way the resulting fixation process fluctuates around the origin, aligning with the Ornstein–Uhlenbeck (OU) prior used in this study. Under an OU assumption, setting $\boldsymbol{\mu} = \mathbf{0}$, the latent dynamics can be written as

$$d\mathbf{z}_t = -\boldsymbol{\Theta}\, \mathbf{z}(t)\, dt + \boldsymbol{\Sigma}^{1/2}\, d\mathbf{w}(t),$$

where $\mathbf{z}(t)$ denotes the centered state, $\boldsymbol{\Theta} \succ 0$ encodes mean-reversion strength toward zero, $\boldsymbol{\Sigma} \succeq 0$ is the diffusion covariance, and $d\mathbf{w}(t)$ is Brownian motion. Zero-centering ensures that the empirical mean corresponds to the OU equilibrium (the origin), simplifying estimation and stabilizing training when coupling OU priors with Neural SDE components.

3.3 Models

To benchmark subject identification from fixations, both baseline neural sequence models and Neural SDE-based architectures are employed:

3.3.1 Baseline Models

Recurrent Neural Network (RNN)

RNN serves as a deterministic sequential baseline that maps a scanpath $\mathbf{x}_{1:T} = \{(x_t, y_t)\}_{t=1}^{T}$ to a subject identity by iteratively updating a hidden state $\mathbf{h}_t$ and producing a sequence summary for classification. Specifically, inputs are normalized and linearly embedded $\tilde{\mathbf{x}}_t = \mathbf{E}\mathbf{x}_t$ before the recurrence, which is defined by $\mathbf{h}_t = \phi(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{xh}\tilde{\mathbf{x}}_t + \mathbf{b}_h)$ with $\mathbf{h}_0 = \mathbf{0}$ and nonlinearity ϕ (e.g., tanh) [87].

Gated Recurrent Unit (GRU)

GRU augments a recurrent cell with *update* and *reset* gates to mitigate vanishing gradients while preserving long-range temporal dependencies in scanpaths. Given input $\mathbf{x}_t$ (gaze coordinates embedding) and previous hidden state $\mathbf{h}_{t-1}$, the gates and state are

$$\mathbf{z}_t = \sigma(\mathbf{W}_{xz}\mathbf{x}_t + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z), \quad \mathbf{r}_t = \sigma(\mathbf{W}_{xr}\mathbf{x}_t + \mathbf{W}_{hr}\mathbf{h}_{t-1} + \mathbf{b}_r),$$

$$\tilde{\mathbf{h}}_t = \tanh\big(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}(\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h\big), \quad \mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t,$$

where σ is the logistic sigmoid and $\odot$ denotes elementwise multiplication. The update gate $\mathbf{z}_t$ controls how much new information overwrites the past, while the reset gate $\mathbf{r}_t$ modulates how much of the previous state is considered when forming the candidate $\tilde{\mathbf{h}}_t$. GRUs typically achieve a favorable accuracy–efficiency trade-off versus LSTMs due to fewer parameters, making them well-suited to long gaze sequences with moderate computational cost [88].

Long Short-Term Memory (LSTM)

LSTM network augments a recurrent cell with an explicit memory state to better preserve long-range dependencies in gaze scanpaths by controlling information flow via input, forget, and output gates. Given input $\mathbf{x}_t$, previous hidden state $\mathbf{h}_{t-1}$, and previous cell state $\mathbf{c}_{t-1}$, the forward update is

$$\mathbf{i}_t = \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i), \quad \mathbf{f}_t = \sigma(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f),$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o), \quad \tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_{xc}\mathbf{x}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c),$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \qquad \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t),$$

where σ is the logistic sigmoid and $\odot$ denotes elementwise multiplication. The forget gate $\mathbf{f}_t$ retains past memory, the input gate $\mathbf{i}_t$ writes new content $\tilde{\mathbf{c}}_t$ into the cell, and the output gate $\mathbf{o}_t$ exposes a filtered view of the memory as the hidden state. Compared to vanilla RNNs, LSTMs more robustly capture long temporal structure in noisy, variable-length gaze sequences at a modest increase in parameters and computing [89].

3.3.2 Neural Differential Equation Models

Neural Ordinary Differential Equation (ODE)

An Ordinary Differential Equation (ODE) is a mathematical equation that describes how a dependent variable changes with respect to a single independent variable, typically time. It is a fundamental tool for modeling continuous-time dynamical systems where the rate of change at any instant depends on the current state and possibly time itself [90]. To model a phenomenon with an ODE means to express the belief that the system's future evolution is completely determined by its current state and the governing differential equation. This embodies several key assumptions:

1. **Deterministic evolution**: Given the current state, the future trajectory is uniquely determined (no randomness);
2. **Markovian property**: The future depends only on the present state, not on the historical path;
3. **Continuous-time dynamics**: The system evolves smoothly without discrete jumps;

4. **Smoothness**: The underlying processes are differentiable with respect to time.

Neural ODEs extend traditional ODEs by replacing the fixed, parametric drift function with a neural network, allowing the dynamics to be learned from data and enabling more flexible, data-adaptive continuous-time models. They represent the hidden state evolution as a continuous-time dynamical system, parameterized by a neural network. Unlike discrete recurrent architectures that update states at fixed intervals, Neural ODEs define the state derivative directly, allowing for adaptive computation and memory-efficient training. For gaze scanpaths, the latent state $\mathbf{z}(t)$ evolves according to:

$$\frac{d\mathbf{z}(t)}{dt} = f(\mathbf{z}(t), t; \theta)$$

with initial condition $\mathbf{z}(0) = g(\mathbf{x}_{1:T})$, where f is a neural network parameterized by θ, and g is an encoder that maps the input scanpath to an initial latent state [90].

Neural Controlled Differential Equation (CDE)

Controlled Differential Equations (CDEs) extend classical ODEs by allowing the dynamics to be driven by an external control path, making it especially well-suited for modeling irregularly sampled time series and sequential data [91]. So the latent state dynamics are entirely determined by the current state and the control path, capturing how inputs directly influence the system's evolution. Since the control $\mathbf{x}(t)$ can be defined continuously (e.g., via splines), CDEs naturally accommodate observations at nonuniform time points.

Neural CDEs extend traditional CDEs exploiting neural networks. For gaze scanpaths, the observed time series $\{\mathbf{x}_t\}_{t=1}^{T}$ is interpolated into a continuous path $\mathbf{x}(t)$ via cubic splines, and the latent state $\mathbf{z}(t)$ evolves according to a controlled differential equation:

$$\frac{d\mathbf{z}(t)}{dt} = f(\mathbf{z}(t), t)\frac{d\mathbf{x}(t)}{dt}$$

with $\mathbf{z}(0) = z_0$, where f is a neural network parametrized by learnable weights, that modulates how changes in $\mathbf{x}(t)$ drive the latent dynamics [91].

Neural Stochastic Differential Equation (Neural SDE)

Stochastic Differential Equations (SDEs) extend ODEs by incorporating stochasticity through a diffusion term, enabling the modeling of uncertainty and noise inherent in real-world dynamical systems. Neural SDEs extend traditional SDEs by parameterizing both the drift and diffusion coefficients with neural networks, allowing for flexible, data-driven modeling of stochastic continuous-time dynamics. For gaze scanpaths, Neural SDEs provide a natural framework to capture the variability and unpredictability in human eye movement patterns while maintaining continuous-time dynamics. The latent state $\mathbf{z}(t)$ evolves according to:

$$d\mathbf{z}(t) = f(\mathbf{z}(t), t; \theta_f)\, dt + g(\mathbf{z}(t), t; \theta_g)\, d\mathbf{w}(t)$$

with initial condition $\mathbf{z}(0) = h(\mathbf{x}_{1:T}; \theta_h)$, where f and g are neural networks parameterizing the drift and diffusion coefficients respectively, $d\mathbf{w}(t)$ is a multi-dimensional Brownian motion, and $\theta = \{\theta_f, \theta_g, \theta_h\}$ are learnable parameters.

The drift term $f(\mathbf{z}(t), t; \theta_f)$ captures the deterministic evolution of the latent dynamics, similar to Neural ODEs, while the diffusion term $g(\mathbf{z}(t), t; \theta_g)\, d\mathbf{w}(t)$ models the stochastic fluctuations. The diffusion coefficient g can be either a scalar function (diagonal noise) or a matrix-valued function (correlated noise), depending on the complexity required for the specific application (this work considers a diagonal noise). The theoretical foundation and training methodology for Neural SDEs are detailed in [61].

Neural Langevin-type Stochastic Differential Equation (Neural LSDE)

Neural LSDE constitutes one of the three stable Neural SDE classes proposed by Oh et al. [73], specifically designed to address stability issues in modeling irregular time series data.

The Neural LSDE is defined by the following stochastic differential equation:

$$d\mathbf{z}(t) = \gamma(\mathbf{z}(t); \theta_\gamma)\, dt + \sigma(t; \theta_\sigma)\, d\mathbf{w}(t)$$

with $\mathbf{z}(0) = h(\mathbf{x}; \theta_h)$, where the initial condition $h(\cdot; \theta_h)$ is an affine function with parameter θ_h that maps the input scanpath to the initial latent state, $\mathbf{z}(t)$ is the latent state, the drift function $\gamma(\cdot; \theta_\gamma)$ is a neural network with parameter θ_γ, the diffusion function $\sigma(\cdot; \theta_\sigma)$ is a neural network with parameter θ_σ and $d\mathbf{w}(t)$ is the Brownian motion. Note that the drift function of Neural LSDEs is not explicitly dependent on time.

Neural Linear Noise Stochastic Differential Equation (Neural LNSDE)

Neural LNSDE constitutes the second class of stable Neural SDE architectures proposed by Oh et al. [73], characterized by linear multiplicative noise that scales proportionally with the latent state.

The Neural LNSDE is defined by the following stochastic differential equation:

$$d\mathbf{z}(t) = \gamma(t, \mathbf{z}(t); \theta_\gamma)\, dt + \sigma(t; \theta_\sigma)\mathbf{z}(t)\, dW(t)$$

with $\mathbf{z}(0) = h(\mathbf{x}; \theta_h)$, where the initial condition $h(\cdot; \theta_h)$ is an affine function with parameter θ_h that maps the input scanpath to the initial latent state, $\mathbf{z}(t)$ is the latent state, the drift function $\gamma(\cdot; \theta_\gamma)$ is a neural network with parameter θ_γ, the diffusion function $\sigma(\cdot; \theta_\sigma)$ is a neural network with parameter θ_σ and $d\mathbf{w}(t)$ is the Brownian motion.

The distinguishing feature of the Neural LNSDE is the multiplicative noise structure, where the diffusion term $\sigma(t; \theta_\sigma)\mathbf{z}(t)$ scales linearly with the latent state. This creates state-dependent stochasticity that can capture phenomena where uncertainty grows proportionally to the magnitude of the underlying process.

Neural Geometric Stochastic Differential Equation (Neural GSDE)

Neural GSDE represents the third stable SDE variant in Oh et al. [73].

The Neural GSDE is defined by:

$$\frac{d\mathbf{z}(t)}{\mathbf{z}(t)} = \gamma(t, \mathbf{z}(t); \theta_\gamma)\, dt + \sigma(t; \theta_\sigma) dW(t)$$

with $\mathbf{z}(0) = h(\mathbf{x}; \theta_h)$, where the initial condition $h(\cdot; \theta_h)$ is an affine function with parameter θ_h that maps the input scanpath to the initial latent state, $\mathbf{z}(t)$ is the latent state, the drift function $\gamma(\cdot; \theta_\gamma)$ is a neural network with parameter θ_γ, the diffusion function $\sigma(\cdot; \theta_\sigma)$ is a neural network with parameter θ_σ and $d\mathbf{w}(t)$ is the Brownian motion.

3.4 Incorporating a Controlled Path to Neural SDEs.

To further improve empirical performance and effectively capture sequential observations in gaze scanpaths, we incorporate a controlled path into the drift functions of the proposed Neural SDEs following the approach of Oh et al. [73]. We define an augmented latent state $\overline{\mathbf{z}}(t)$ that incorporates a controlled path in a nonlinear way as follows:

$$\overline{\mathbf{z}}(t) = \zeta(t, \mathbf{z}(t), X(t); \theta_\zeta),$$

where $X(t)$ is the controlled path (a cubic spline interpolation of the input scanpath data) and ζ is a neural network parameterized by θ_ζ. We then replace $\mathbf{z}(t)$ in the drift functions of the proposed Neural SDEs with $\overline{\mathbf{z}}(t)$ as defined above.

This controlled path extension allows the drift dynamics to depend not only on the current latent state but also on the history and temporal structure of the gaze sequence, enabling better capture of sequential patterns in fixation data. The controlled path $X(t)$ provides a continuous representation of the discrete scanpath observations, bridging the gap between the irregular sampling of eye movement data and the continuous-time SDE formulation.

As demonstrated in Oh et al. [73], when combined with $\bar{\mathbf{z}}(t)$, the proposed Neural SDEs including Neural LSDE, Neural LNSDE, and Neural GSDE maintain their unique strong solution properties. The effectiveness of incorporating controlled paths is validated through their comprehensive ablation studies, showing improved performance in capturing sequential dependencies in time series data.

However, this controlled path formulation introduces an important limitation: the model becomes inherently discriminative rather than generative, as it requires the complete input scanpath $X(t)$ to define the controlled path, thereby preventing the generation of novel gaze sequences from the learned dynamics alone.

3.5 Training and Metrics

In order to perform subject classification, a linear classifier is added at the end of each model, which input is the last state $\mathbf{h}_T$ of the model evolution. This provide the class probabilities $\hat{\mathbf{p}} = \text{softmax}(\mathbf{W}_{ho}\mathbf{h}_T + \mathbf{b}_o)$. All models are trained to minimize the cross-entropy loss:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log \hat{y}_i \tag{1}$$

where y is the true class label and $\hat{y}$ is the predicted probability. A grid search is performed over key hyperparameters (batch size and hidden dimension) for each architecture.

Performance is evaluated using the following metrics:

- **Number of epochs to convergence**: the training epoch achieving the lowest validation loss.
- **Cross entropy loss**: on validation and test sets.
- **Fixation-level classification accuracy**: proportion of individual fixations correctly classified.
- **Scanpath-level classification accuracy**: majority voting over fixation-level predictions in a given scanpath.

4 Experiments and Results

This chapter presents the empirical evaluation of the proposed baseline and Neural SDE–based architectures for subject identification from eye-fixation scanpaths. After training each model to minimize cross-entropy loss, we assess their performance in terms of validation loss, fixation-level accuracy, and scanpath-level accuracy. Table 1 summarizes these metrics for all models (best values are in **bold**, second best are in *italic*).

Table 1: Comparison of validation loss, fixation-level accuracy, and scanpath-level accuracy across models.

Model	Epoch	Val Loss	Fixation Acc.	Scanpath Acc.
RNN	19	1.970	0.255	0.312
GRU	33	1.500	0.502	0.705
LSTM	35	1.580	0.489	0.675
CDE	92	1.900	0.323	0.433
ODE	33	1.990	0.212	0.267
SDE	45	1.990	0.223	0.271
LSDE	165	1.270	**0.583**	**0.787**
LNSDE	169	1.217	*0.571*	*0.775*
GSDE	17	1.960	0.265	0.342

The results reveal several key insights. First, among the classical sequence models, the GRU achieves the lowest validation loss (1.500) and substantially outperforms the vanilla RNN and LSTM in both fixation-level (0.532) and scanpath-level (0.767) accuracy. This confirms the GRU's effectiveness in capturing temporal dependencies in noisy gaze sequences while remaining parameter-efficient.

Second, Neural CDEs and ODEs, which offer continuous-time modeling, exhibit competitive performance relative to RNNs but do not surpass the GRU baseline. The Neural ODE yields the poorest fixation (0.212) and scanpath (0.267) accuracies, indicating that purely deterministic continuous dynamics struggle with the variability inherent in eye-fixation data.

Third, integrating stochasticity via generic Neural SDEs marginally improves over the ODE, but still lags behind GRU. This highlights that noise alone is insufficient without stability guarantees.

Crucially, the stable Neural SDE variants (LSDE and LNSDE) demonstrate the strongest performance. The LSDE achieves the highest fixation accuracy (0.583) and scanpath accuracy (0.787), closely followed by the LNSDE (fixation 0.571, scanpath 0.775). These results underscore the benefits of incorporating stable, state-dependent diffusion dynamics, which better capture the stochastic structure of fixation patterns. The GSDE, while guaranteeing geometric constraints, underperforms relative to the linear-noise and Langevin variants, suggesting that positivity constraints may not align optimally with zero-centered fixation data.

Overall, the benchmarking study confirms that stable Neural SDEs, particularly the LSDE and LNSDE, offer significant gains in discriminative accuracy for subject classification from eye-fixations, validating the proposed pipeline's ability to leverage stochastic and stability-enhanced continuous-time dynamics.

5 Conclusions

In this work, we presented a comprehensive pipeline for modeling eye-fixation scanpaths and performing subject identification via a range of neural sequence models and continuous-time stochastic differential equation (SDE) architectures. Beginning with data acquisition from the Fixations In Faces (FiFa) database, we applied I-VT event classification and centering preprocessing to align fixation coordinates with Ornstein–Uhlenbeck priors. We then benchmarked traditional recurrent neural networks (RNN, GRU, LSTM) and neural differential equation models (Neural CDE, Neural ODE, Neural SDE) alongside three novel stable SDE variants (Langevin-type, Linear Noise, and Geometric SDE) for discriminative classification.

Our experiments demonstrate that gated recurrent architectures (GRU, LSTM) outperform vanilla RNNs and continuous-time baselines in both fixation-level and scanpath-level accuracy, confirming their efficacy in capturing temporal patterns in noisy gaze data. However, the incorporation of stochasticity alone (Neural SDE) and deterministic continuous dynamics (Neural ODE) does not suffice to surpass these baselines. Crucially, the stable Neural SDE variants—particularly the Langevin-type SDE (LSDE) and Linear Noise SDE (LNSDE)—achieve the best overall performance, with substantially lower validation losses and the highest classification accuracies (fixation-level $\approx$ 0.58, scanpath-level $\approx$ 0.79). This highlights the importance of stability-ensuring formulations that guarantee well-posed dynamics and mean-reverting or state-dependent noise tailored to fixation processes.

We further extended these models with a controlled path mechanism, enabling the drift dynamics to leverage the continuous scanpath as a control signal. While this augmentation improves empirical results by better capturing sequential dependencies, it also renders the model inherently discriminative and unable to generate novel sequences without observed trajectories.

Overall, this study confirms that stable Neural SDEs provide a powerful framework for eye-tracking–based subject identification, effectively balancing expressivity, stochastic modeling, and stability. Future work may explore generative extensions and a benchmark classification based on saccades, that can greatly improve scanpath classification.

References

[1] Dodge, R., Cline, T.S.: The angle velocity of eye movements. Psychological Review **8**(2), 145 (1901)

[2] Dodge, R.: Five types of eye movement in the horizontal meridian plane of the field of regard. American journal of physiology-legacy content **8**(4), 307–329 (1903)

[3] Cartella, G., Cornia, M., Cuculo, V., D'Amelio, A., Zanca, D., Boccignone, G., Cucchiara, R.: Trends, Applications, and Challenges in Human Attention Modelling (2024). https://arxiv.org/abs/2402.18673

[4] Cerf, M., Harel, J., Einhäuser, W., Koch, C.: Predicting human gaze using low-level saliency combined with face detection. Advances in neural information processing systems **20** (2007)

[5] Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(11), 1254–1259 (1998) https://doi.org/10.1109/34.730558

[6] D'Amelio, A., Patania, S., Bursic, S., Cuculo, V., Boccignone, G.: Using gaze for behavioural biometrics. Sensors **23**(3), 1262 (2023)

[7] Bolle, R.M., Connell, J.H., Pankanti, S., Ratha, N.K., Senior, A.W.: Guide to Biometrics. Springer, ??? (2013)

[8] Al Zaidawi, S.M.K., Prinzler, M.H., Lührs, J., Maneth, S.: An extensive study of user identification via eye movements across multiple datasets. Signal Processing: Image Communication **108**, 116804 (2022)

[9] Schröder, C., Al Zaidawi, S.M.K., Prinzler, M.H., Maneth, S., Zachmann, G.:

Robustness of eye movement biometrics against varying stimuli and varying trajectory length. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–7 (2020)

[10] Hrechak, A.K., McHugh, J.A.: Automated fingerprint recognition using structural matching. Pattern Recognition **23**(8), 893–904 (1990) https://doi.org/10.1016/0031-3203(90)90134-7

[11] Galbally, J., Haraksim, R., Beslay, L.: A study of age and ageing in fingerprint biometrics. IEEE Transactions on Information Forensics and Security **14**(5), 1351–1365 (2019) https://doi.org/10.1109/TIFS.2018.2878160

[12] Wang, M., Deng, W.: Deep face recognition: A survey. Neurocomputing **429**, 215–244 (2021)

[13] Cuculo, V., D'amelio, A., Grossi, G., Lanzarotti, R., Lin, J.: Robust single-sample face recognition by sparsity-driven sub-dictionary learning using deep features. Sensors **19**(1), 146 (2019)

[14] Rattani, A., Derakhshani, R.: Ocular biometrics in the visible spectrum: A survey. Image and Vision Computing **59**, 1–16 (2017)

[15] Das, R., Gayathri, V., Divyajyoti, Jose, S., Bartos, I., Klimenko, S., Mishra, C.K.: Inferring additional physics through unmodeled signal reconstructions. Physical Review D **112**(2), 023011 (2025)

[16] Hessels, R.S., Niehorster, D.C., Nyström, M., Andersson, R., Hooge, I.T.: Is the eye-movement field confused about fixations and saccades? a survey among 124 researchers. Royal Society open science **5**(8), 180502 (2018)

[17] Hessels, R.S., Niehorster, D.C., Kemner, C., Hooge, I.T.: Noise-robust fixation

detection in eye movement data: Identification by two-means clustering (i2mc). Behavior research methods **49**(5), 1802–1823 (2017)

[18] Rolfs, M.: Microsaccades: small steps on a long way. Vision research **49**(20), 2415–2441 (2009)

[19] Gegenfurtner, A., Lehtinen, E., Säljö, R.: Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. Educational psychology review **23**(4), 523–552 (2011)

[20] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Weijer, J.: Eye Tracking: A Comprehensive Guide to Methods and Measures. oup Oxford, ??? (2011)

[21] Larsson, L., Nyström, M., Andersson, R., Stridh, M.: Detection of fixations and smooth pursuit movements in high-speed eye-tracking data. Biomedical Signal Processing and Control **18**, 145–152 (2015)

[22] Leigh, R.J., Zee, D.S.: The Neurology of Eye Movements. Oxford university press, ??? (2015)

[23] Sparks, D.L.: The brainstem control of saccadic eye movements. Nature Reviews Neuroscience **3**(12), 952–964 (2002)

[24] Duchowski, A.T., Duchowski, A.T.: Eye Tracking Methodology: Theory and Practice. Springer, ??? (2017)

[25] Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological bulletin **124**(3), 372 (1998)

[26] Shic, F., Chawarska, K., Scassellati, B.: The amorphous fixation measure revisited: With applications to autism. In: 30th Annual Meeting of the Cognitive

Science Society, pp. 1–6 (2008). Presented at the 30th Annual Meeting of the Cognitive Science Society ...

[27] Jovancevic-Misic, J., Hayhoe, M.: Adaptive gaze control in natural environments. Journal of Neuroscience **29**(19), 6234–6238 (2009)

[28] Brunelli, R., Falavigna, D.: Person identification using multiple cues. IEEE transactions on pattern analysis and machine intelligence **17**(10), 955–966 (2002)

[29] Ross, A., Jain, A.: Information fusion in biometrics. Pattern recognition letters **24**(13), 2115–2125 (2003)

[30] Canosa, R.L.: Real-world vision: Selective perception and task. ACM Transactions on Applied Perception (TAP) **6**(2), 1–34 (2009)

[31] Cerf, M., Frady, E.P., Koch, C.: Faces and text attract gaze independent of the task: Experimental data and computer model. Journal of vision **9**(12), 10–10 (2009)

[32] Friston, K.J.: The labile brain. ii. transients, complexity and selection. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **355**(1394), 237–252 (2000)

[33] Ainley, V., Apps, M.A., Fotopoulou, A., Tsakiris, M.: 'bodily precision': a predictive coding account of individual differences in interoceptive accuracy. Philosophical Transactions of the Royal Society B: Biological Sciences **371**(1708), 20160003 (2016)

[34] Rudrauf, D., Bennequin, D., Granic, I., Landini, G., Friston, K., Williford, K.: A mathematical model of embodied consciousness. Journal of theoretical biology **428**, 106–131 (2017)

[35] Solms, M., Friston, K.: How and why consciousness arises: some considerations from physics and physiology. Journal of Consciousness Studies **25**(5-6), 202–238 (2018)

[36] Atzil, S., Gao, W., Fradkin, I., Barrett, L.: Growing a social brain. Nature Human Behaviour, 2 (9), 624-636 (2018)

[37] De Vignemont, F., Alsmith, A.J.: The Subject's Matter: Self-consciousness and the Body. MIT Press, ??? (2017)

[38] Apps, M.A., Tsakiris, M.: The free-energy self: a predictive coding account of self-recognition. Neuroscience & Biobehavioral Reviews **41**, 85–97 (2014)

[39] Mirza, M.B., Adams, R.A., Mathys, C.D., Friston, K.J.: Scene construction, visual foraging, and active inference. Frontiers in computational neuroscience **10**, 56 (2016)

[40] Trappes, R.: Individual differences, uniqueness, and individuality in behavioural ecology. Studies in History and Philosophy of Science **96**, 18–26 (2022)

[41] Stephens, D.W., Krebs, J.R.: Foraging Theory vol. 6. Princeton university press, ??? (1986)

[42] Viswanathan, G.M., Da Luz, M.G., Raposo, E.P., Stanley, H.E.: The Physics of Foraging: an Introduction to Random Searches and Biological Encounters. Cambridge University Press, ??? (2011)

[43] Bartumeus, F., Catalan, J.: Optimal search behavior and classic foraging theory. Journal of Physics A: Mathematical and Theoretical **42**(43), 434002 (2009)

[44] Toscano, B.J., Gownaris, N.J., Heerhartz, S.M., Monaco, C.J.: Personality, foraging behavior and specialization: integrating behavioral and food web ecology

at the individual level. Oecologia **182**(1), 55–69 (2016)

[45] Todd, P.M., Hills, T.T.: Foraging in mind. Current Directions in Psychological Science **29**(3), 309–315 (2020)

[46] Budaev, S., Jørgensen, C., Mangel, M., Eliassen, S., Giske, J.: Decision-making from the animal perspective: bridging ecology and subjective cognition. Frontiers in Ecology and Evolution **7**, 164 (2019)

[47] Rosati, A.G.: Foraging cognition: reviving the ecological intelligence hypothesis. Trends in cognitive sciences **21**(9), 691–702 (2017)

[48] Priolli, P.: Information foraging theory: Adaptive interaction with information. Oxford University Press, Oxford-New York (2007)

[49] Hills, T.T.: Animal foraging and the evolution of goal-directed cognition. Cognitive science **30**(1), 3–41 (2006)

[50] Liversedge, S.P., Findlay, J.M.: Saccadic eye movements and cognition. Trends in cognitive sciences **4**(1), 6–14 (2000)

[51] Wolfe, J.M.: When is it time to move to the next raspberry bush? foraging rules in human visual search. Journal of vision **13**(3), 10–10 (2013)

[52] Ehinger, K.A., Wolfe, J.M.: When is it time to move to the next map? optimal foraging in guided visual search. Attention, Perception, & Psychophysics **78**(7), 2135–2151 (2016)

[53] Cain, M.S., Vul, E., Clark, K., Mitroff, S.R.: A bayesian optimal foraging model of human visual search. Psychological science **23**(9), 1047–1054 (2012)

[54] Friston, K.J., Shiner, T., FitzGerald, T., Galea, J.M., Adams, R., Brown, H.,

Dolan, R.J., Moran, R., Stephan, K.E., Bestmann, S.: Dopamine, affordance and active inference. PLoS computational biology **8**(1), 1002327 (2012)

[55] Shepherd, S.V., Platt, M.L.: Spontaneous social orienting and gaze following in ringtailed lemurs (lemur catta). Animal cognition **11**(1), 13–20 (2008)

[56] Guy, N., Azulay, H., Kardosh, R., Weiss, Y., Hassin, R.R., Israel, S., Pertzov, Y.: A novel perceptual trait: Gaze predilection for faces during visual exploration. Scientific reports **9**(1), 10714 (2019)

[57] Boccignone, G., Cuculo, V., D'Amelio, A., Grossi, G., Lanzarotti, R.: Give ear to my face: Modelling multimodal attention to social interactions. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp. 0–0 (2018)

[58] Fein-Ashley, J.: Methods for parameter estimation of the ornstein–uhlenbeck process. arXiv (2024) arXiv:2404.11526 [q-fin.CP]

[59] Eckhoff, P.: On diffusion processes with variable drift rates as models of evidence accumulation in a visual discrimination task. PLoS Computational Biology **4**(1), 1000183 (2008) https://doi.org/10.1371/journal.pcbi.0040062

[60] Tzen, B., Raginsky, M.: Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit. arXiv preprint arXiv:1905.09883 (2019)

[61] Li, X., Wong, T.-K.L., Chen, R.T., Duvenaud, D.: Scalable gradients for stochastic differential equations. In: International Conference on Artificial Intelligence and Statistics, pp. 3870–3882 (2020). PMLR

[62] Hodgkinson, L., Heide, C., Roosta, F., Mahoney, M.W.: Stochastic normalizing flows. arXiv preprint arXiv:2002.09547 (2020)

[63] Giles, M., Glasserman, P.: Smoking adjoints: Fast monte carlo greeks. Risk **19**(1), 88–92 (2006)

[64] Coffey, W., Kalmykov, Y.P.: The Langevin Equation: with Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering vol. 27. World Scientific, ??? (2012)

[65] Pavliotis, G.A.: Stochastic processes and applications. Texts in applied mathematics **60** (2014)

[66] Lelievre, T., Stoltz, G.: Partial differential equations and stochastic methods in molecular dynamics. Acta Numerica **25**, 681–880 (2016)

[67] Black, F., Scholes, M.: The pricing of options and corporate liabilities. Journal of political economy **81**(3), 637–654 (1973)

[68] Brigo, D., Mercurio, F.: Interest Rate models—Theory and Practice: With Smile, Inflation and Credit. Springer, ??? (2006)

[69] Arató, M.: A famous nonlinear stochastic equation (lotka-volterra model with diffusion). Mathematical and Computer Modelling **38**(7-9), 709–726 (2003)

[70] Soboleva, T., Pleasants, A.: Population growth as a nonlinear stochastic process. Mathematical and Computer Modelling **38**(11-13), 1437–1442 (2003)

[71] Huillet, T.: On wright–fisher diffusion and its relatives. Journal of Statistical Mechanics: Theory and Experiment **2007**(11), 11006 (2007)

[72] Kidger, P.: On neural differential equations. arXiv preprint arXiv:2202.02435 (2022)

[73] Oh, Y., Lim, D.-Y., Kim, S.: Stable neural stochastic differential equations in

analyzing irregular time series data. arXiv preprint arXiv:2402.14989 (2024)

[74] Kasprowski, P.: Human identification using eye movements. Institute of Computer Science, 1–111 (2004)

[75] Bednarik, R., Kinnunen, T., Mihaila, A., Fränti, P.: Eye-movements as a biometric. In: Scandinavian Conference on Image Analysis, pp. 780–789 (2005). Springer

[76] Kasprowski, P., Ober, J.: Enhancing eye-movement-based biometric identification method by using voting classifiers. In: Biometric Technology for Human Identification II, vol. 5779, pp. 314–323 (2005). SPIE

[77] Holland, C., Komogortsev, O.V.: Biometric identification via eye movement scanpaths in reading. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–8 (2011). IEEE

[78] Rigas, I., Economou, G., Fotopoulos, S.: Biometric identification based on the eye movements and graph matching techniques. Pattern Recognition Letters **33**(6), 786–792 (2012)

[79] Friedman, L., Nixon, M.S., Komogortsev, O.V.: Method to assess the temporal persistence of potential biometric features: Application to oculomotor, gait, face and brain structure databases. PloS one **12**(6), 0178501 (2017)

[80] Jia, S., Seccia, A., Antonenko, P., Lamb, R., Keil, A., Schneps, M., Pomplun, M., *et al.*: Biometric recognition through eye movements using a recurrent neural network. In: 2018 IEEE International Conference on Big Knowledge (ICBK), pp. 57–64 (2018). IEEE

[81] Jäger, L.A., Makowski, S., Prasse, P., Liehr, S., Seidler, M., Scheffer, T.: Deep

eyedentification: Biometric identification using micro-movements of the eye. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 299–314 (2019). Springer

[82] Makowski, S., Prasse, P., Reich, D.R., Krakowczyk, D., Jäger, L.A., Scheffer, T.: Deepeyedentificationlive: Oculomotoric biometric identification and presentation-attack detection using deep neural networks. IEEE Transactions on Biometrics, Behavior, and Identity Science **3**(4), 506–518 (2021)

[83] Lohr, D., Griffith, H., Komogortsev, O.V.: Eye know you: Metric learning for end-to-end biometric authentication using eye movements from a longitudinal dataset. IEEE Transactions on Biometrics, Behavior, and Identity Science **4**(2), 276–288 (2022)

[84] Abdelwahab, A., Landwehr, N.: Deep distributional sequence embeddings based on a wasserstein loss. Neural Processing Letters **54**(5), 3749–3769 (2022)

[85] Yin, J., Sun, J., Li, J., Liu, K.: An effective gaze-based authentication method with the spatiotemporal feature of eye movement. Sensors **22**(8), 3002 (2022)

[86] Salvucci, D.D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, pp. 71–78. ACM, ??? (2000)

[87] Elman, J.L.: Finding structure in time. Cognitive Science **14**(2), 179–211 (1990) https://doi.org/10.1207/s15516709cog1402_1

[88] Cho, K., Merriënboer, B., Gülçehre, Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)

[89] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997)

[90] Chicone, C.: Ordinary Differential Equations with Applications, 2nd edn. Springer, ??? (2006)

[91] Kidger, P., Morrill, J., Foster, J., Lyons, T.: Neural controlled differential equations for irregular time series. Advances in neural information processing systems **33**, 6696–6707 (2020)