Check for updates

Quantifying the Effect of Anticipatory Eye Movement on Successful Ball Hitting Using Fine-Scale Tracking and SHAP-Analysis

Prasetia Utama Putra
prasetia.putra@uni-konstanz.de
Centre for the Advanced Study of Collective Behaviour,
Unviersity of Konstanz
Konstanz, Germany

Fumihiro Kano
fumihiro.kano@uni-konstanz.de,
fkano@ab.mpg.de
Centre for the Advanced Study of Collective Behaviour,
Unviersity of Konstanz,
Max Planck Institute of Animal Behavior
Konstanz, Germany

ABSTRACT

Previous studies have suggested that anticipatory eye movement is important for success in dynamic interception tasks, such as ball sports. However, there are few quantitative studies that explore the connection between specific eye movement and successful ball interception. In this study, we used a state-of-the-art motion-capture and eye-tracking system to monitor the eye and body movements of 91 individuals playing ping-pong against a wall. We analyzed 24 features related to eye movement and action, using nonlinear machine learning models to understand the relationship between these features and successful hits. Our key findings showed that the accuracy and timing of anticipatory saccades to the ball's bounce points were crucial for success. Successful hits were characterized by more precise, just-in-time anticipatory looks and longer duration of ball pursuit compared to unsuccessful hits. This quantitative framework is the first step towards understanding how humans plan their actions based on visual perception in realistic, dynamic tasks.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in interaction design.*

KEYWORDS

anticipatory saccades, SHAP-analysis, gaze event detection, dynamic interception task

ACM Reference Format:
Prasetia Utama Putra and Fumihiro Kano. 2024. Quantifying the Effect of Anticipatory Eye Movement on Successful Ball Hitting Using Fine-Scale Tracking and SHAP-Analysis. In *2024 Symposium on Eye Tracking Research and Applications (ETRA '24), June 04–07, 2024, Glasgow, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3649902.3653007

This work is licensed under a Creative Commons Attribution-Share Alike International 4.0 License.

ETRA '24, June 04–07, 2024, Glasgow, United Kingdom
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0607-3/24/06
https://doi.org/10.1145/3649902.3653007

1 INTRODUCTION

In highly dynamic tasks, human eyes actively anticipate upcoming episodes to plan and execute actions at the right time. This is especially evident in ball sports, where the timing of the hit is critical [Land and Tatler 2009]. In a well-known study by Land [Land and McLeod 2000], a cricket player makes an anticipatory saccade to the ball's bounce point and then pursues its trajectory with their eye movement after the bounce, before hitting the ball. A skilled player is distinguished from a less skilled one by their faster latency in making such an anticipatory saccade. Similarly, Hayhoe [Hayhoe et al. 2005] discovered that when individuals catch a bouncing ball, their eyes anticipate the bounce point and smoothly track the ball. In the experiment, participants initially struggled to follow the unexpectedly altered bounce pattern of a more elastic ball, but their performance went back to normal after a few trials. The same pattern of eye movement—anticipatory saccades to the bounce points followed by the smooth pursuit of the ballis also observed in more dynamic tasks such as squash [Hayhoe et al. 2012] and table tennis [Land and Furneaux 1997]. Such anticipatory saccade is likely driven by the players' internal model or the learning of sophisticated mapping between visually obtained variables and motor responses [Diaz et al. 2013a,b; Hayhoe et al. 2005; Land and Tatler 2009].

Despite the emphasis placed on the significance of anticipation and pursuit in ball sports by many previous studies, there is still a need to study how eye movement affects performance in intercepting tasks. Numerous studies have explored eye movement differences between skilled and less skilled players, but few have examined successful and unsuccessful hits by the same players, except a few [Cesqui et al. 2015]. Cesqui et al. [Cesqui et al. 2015] examined successful and unsuccessful ball interception and found that there was no clear pattern between eye movement patterns and catching success, except for the fact that successful catchers followed the ball for a longer time. However, since this previous study examined a relatively simple task, it remains unclear whether the lack of correlation between a player's ability to hit the ball and their eye movements applies to more complex tasks that are more repetitive and demanding.

Our study addressed this gap using advanced tracking techniques and machine learning analysis. We examined gaze and action features contributing to success or failure in ball interception during a

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-5a1ruq2dq0c5

dynamic task involving hitting a ping-pong ball against a wall. Casual players were monitored using motion-capture and eye-tracking systems. Our key finding revealed that a player's success related to the timing and precision of anticipatory saccades to the ball's bounce point, along with prolonged pursuit before hitting the ball.

2 RELATED WORKS

2.1 Interplay of gaze and action

Previous studies have shown that in tasks involving hitting, catching, and pointing, the eyes play a leading role in guiding the hand's movements [Belardinelli et al. 2016; Land and McLeod 2000; Mrotek and Soechting 2007]. In fast-paced ball sports, visual strategies include anticipation, smooth pursuit, and fixation [Land and Tatler 2009].

Saccadic eye movements are commonly used by individuals to adjust their gaze position to the desired location. These movements can be anticipatory or corrective [Liversedge and Findlay 2000]. Anticipatory saccades occur when the eye moves ahead of a stimulus and its expected location [Hayhoe et al. 2012]. Research suggests a link between the ability to predict the future position of a ball and players' performance in sports like cricket [Land and McLeod 2000; Sarpeshkar et al. 2017], squash [Diaz et al. 2013a; Hayhoe et al. 2012], and table tennis [Land and Furneaux 1997]. Skilled cricket players were found to start their saccades earlier than non-skilled players [Land and McLeod 2000]. However, a recent study contradicted this and suggested that skilled players initiate their predictive saccades later [Sarpeshkar et al. 2017].

On the other hand, corrective saccades are performed to correct any errors in the initial eye movements that correspond to the target [Orban de Xivry and Lefevre 2007]. In a virtual-reality tennis environment, it was observed that players made corrective saccades immediately after the ball bounced to realign their eyes with the ball [Mann et al. 2019]. Whether the eyes need to catch up or "wait" for the target determines the direction of these correction saccades, which can be forward or backward, relative to the direction of the ball.

While saccades bring the target into the visual field, fixations and smooth pursuit movements are used to maintain a stable image of the target on the fovea [Lisberger 2010; Rashbass 1961]. In squash, a combination of smooth pursuit and catch-up saccades was observed until shortly before individuals intercept the moving ball [Diaz et al. 2013a,b; Hayhoe et al. 2012]. Previous studies have found that features of smooth pursuits, such as pursuit gain, tracking quality, and duration, influence the timing and performance of interception [Cesqui et al. 2015; Delle Monache et al. 2015; Fooken and Spering 2020]. Participants in a Go/NoGo task who had higher error tracking during smooth pursuit also had higher timing errors when responding to the stimulus. In a task involving catching a ball, individuals who pursued the ball for a longer duration had a higher probability of successfully catching it [Cesqui et al. 2015]. Additionally, when aiming a ball for longer distances, it was observed that individuals had more stable tracking of the non-moving target, indicating successful shots. The duration of this stable tracking increased with the difficulty of the shots [Williams et al. 2002].

2.2 SHAP-analysis

Previous studies have made assumptions about the linear relationship between individuals' gaze movement and their performance. For example, skilled cricket players tend to initiate their predictive saccade earlier or later [Land and McLeod 2000]. In binary conditions, a log-linear assumption is used to understand the difference in duration between players with higher and lower success rates in catching a ball [Cesqui et al. 2015].

However, the assumption of a linear or log-linear relationship in real-life tasks, such as a continuous ball interception task where one episode influences the next, has inherent limitations. In such scenarios, multicollinearity of input features is likely to be present. This can lead to high variance in the estimated parameters, thereby reducing the reliability of the results regarding the contributions of input features [Lipovetsky and Conklin 2001].

One potential way to address this limitation is the use of Shapley Additive Explanations (SHAP) analysis. SHAP-analysis is a method used in the field of machine learning to interpret the contribution of input feature on the output of predictive models based on SHAP-value. The unique aspect of SHAP-analysis is its model-agnostic nature, making it applicable to both linear models and more complex non-linear models [Lundberg et al. 2020]. The usability of this approach has been demonstrated through recent quantitative studies across various fields [Lipovetsky and Conklin 2001; Lundberg et al. 2020]. For instance, XGBoost combined with SHAP-value has been applied in predicting New Yorkers' incomes based on socioeconomic factors [Bai et al. 2023] and assessing the likelihood of autism symptoms in children through behavior phenotyping [Perochon et al. 2023]. Furthermore, the integration of SHAP values with LSTM has provided insights into both unrelated and dependent data, such as in studies of guide RNA activity [Wang et al. 2019] and enzyme sequences [Yu et al. 2023]. Due to the potential presence of linearity, non-linearity and multicollinearity in data of real-life interception task, we decided to use SHAP-analysis combined with both linear and non-linear models.

3 METHOD

3.1 Experimental design

A total of 91 participants (36 female, 54 male, and one non-binary) were recruited for the study. The average age of the participants was 25.29 ± 7.09. All participants had no physical disorders, and those with hyperopia or myopia used additional lenses on their Tobii glasses. Five participants were members of a ping-pong club, while the rest had minimal experience or played casually. The study protocol was approved by the Ethics Committee of the University of Konstanz (No.39/2022).

Participants were equipped with a motion capture suit, Tobii glasses 3, and an ECG sensor during the experiments. The motion capture system consists 13 Vicon Vero and 1 Vantage cameras. These equipment components operated at different sampling rates: 100 Hz for the motion capture system, 50 Hz for the Tobii glasses, and 1-2 Hz for the ECG sensor. To synchronize the systems, we sent TTL signals to the glasses recording unit and triggered the recording of ECG signals when the Vicon motion capture system started recording.

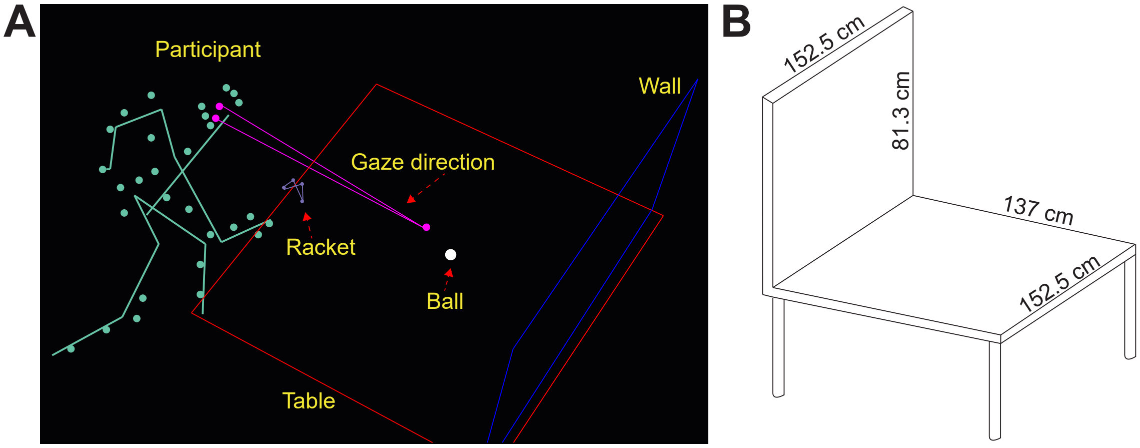


Figure 1: Experimental setup. A: A motion capture system and wearable eye tracker to track the body and eye movement of participants, as well as the position of the table, wall, and racket. B: The dimensions of the table and wall employed during the experiment.

In addition to tracking participants' movements, we also monitored the motion and position of the racket, ball, table, and wall (Figure 1 A and Figure 1 B). To facilitate tracking, markers were attached to the table, wall, and racket, and the ball was coated with reflective powder. The racket had a height of 24.6 cm and a width of 15 cm. The ball had a radius of 4.5 cm. The mass of the racket was 142.2 grams, while the ball weighed 4.04 grams.

Participants took part in both singles and doubles sessions. We provided detailed explanations of experiments and rules before each session with videos and a document. All participants warmed up for 1 to 2 minutes to familiarize themselves with the equipment and the ball. To minimize bias, we randomized the order of the sessions.

Exclusion criteria in this study included the norm score (the ratio of successful outcomes to the total number of episodes) and the percentage of gaze samples. Data from 9 participants were excluded from further analysis since their norm score was less than 0.5 (they failed to hit the ball most of the time) or gaze sampling less than 65%.

3.2 Data pre-processing

Before extracting features, a sanity check was conducted. This involved comparing the number of data points to the recording duration multiplied by the sampling rate. A data cleaning process was also performed to ensure data validity and eliminate noise (Detail: Supplementary Figure 1). In order to validate the synchronization of the data, a manual double check was conducted. The frames obtained from tracking were cross-referenced with the eye tracking recording duration.After that, it was visually confirmed that the time when the ball bounced on the wall aligned with the video recorded using the eyeglasses. To match the sampling frequency of body tracking data, the eye-movement data was up-sampled to 100 Hz using linear interpolation.

For the motion capture system, the data cleaning process involved three main steps. Firstly, a Woltering filter with a mode of MSE and a level of 20 was applied to the motion tracking data, which was then exported to a CSV file. We then removed segments and trajectories that had swapped markers. This was done by considering the distance and velocity of the markers (see Supplementary 3 for detail). Then, we interpolated any missing data. To emphasize the trend, we smoothed the ball tracking data using a moving average with a window of 3 frames.

3.3 Definition of saccades and ball pursuit

In this study, we utilized a modified version of Duchowski's algorithm [Duchowski et al. 2002] to detect saccades in the gaze data. The modifications were made to ensure the algorithm matched our data's specific characteristics, as previous research indicated that no single algorithm worked well in various experimental scenarios [Andersson et al. 2017] (Detail of the eye-movement definition: Supplementary Figure S2).

Saccades were detected by applying Duchnowski's 5-tap filter ([0, 1, 2, 3, 2, 1, 0]) [Duchowski et al. 2002] to the instantaneous visual angle of the two consecutive gaze-in-the-head vectors; saccades were considered during periods when the filtered velocity exceeded $40^{\circ}s^{-1}$, and their onset and offset were defined by acceleration exceeding $15^{\circ}s^{-1}$.

An anticipatory saccade (anticipatory look) was defined as a saccade characterized by a movement pattern that aligns with the ball's trajectory and possesses the largest magnitude during that phase. Saccades occurring after the anticipatory look were classified as correction saccades, whereas all remaining saccades were identified as normal saccades.

In the gaze data, periods that are not related to saccadic movements can be attributed to either fixation or smooth pursuit eye movements. We defined 'ball pursuit' as either fixation or smooth pursuit on the ball. We did not differentiate smooth pursuit eye movement from fixation because we assumed that the looming effect (occurred in phase 3 after the ball bounced on the table) did not require individuals to move their eyes in order to track the ball, as the image of the ball became larger on their retina (Figure 2).

Ball pursuit was defined based on two criteria. The first criterion was that the gaze-in-the-head in visual space relative to the ball should be smaller than 10°. The second criterion required the velocity of the gaze-in-the-head vector to be higher than 30°, with the velocity of the gaze vector over ball vector should range from 0.3 to 1.2. The duration of the ball pursuit needed to be at least 50 ms or longer.

3.4 Definition of eye movement and action features

We used tracking results of gaze, body, racket, table, wall, and ball to extract eye-movement, action, and impact features. For eye movement features, the study focused on anticipatory look and corrective saccade in phase 1 (p1) and phase 2 (p2), as well as ball pursuit variables in phase 3 (p3). The anticipatory look features included the onset, magnitude, and precision of AL, defined as pr_p1_al_on, pr_p1_al_mag, and pr_p1_al_prec. The corrective saccade, pr_p1_cs, was defined as the number of saccades succeeding the anticipatory look. In phase 3, the eye movement features included the onset of ball pursuit (pr_p3_fx_on), its duration (pr_p3_fx_du), and the discrepancy in distance between the ball and gaze from the average distance during ball pursuit (pr_p3_stable).

The action features included the start of the forward swing (ec_start_fs), the distance between the ball and the racket when it started (ec_fs_ball_racket_dist), the ratio of ball and racket speed (ec_fs_ball_racket_ratio) during the forward swing, and the direction of the ball and racket (ec_fs_ball_racket_dir). To determine the start of the forward swing, we measured the acceleration of the distance between the racket and a wall segment.

The moment of impact was defined as the point at which participants made contact with the ball using the racket. From this, we extracted five features: the angle of collision (im_rb_ang_collision), the force exerted by the racket (im_racket_force), the force exerted by the ball (im_racket_ball), the distance of the ball from the center of the racket and the wrist (im_rack_wrist_dist), and the distance of the ball relative to the centroid of the racket (im_rb_dist). Additionally, we measured the deviation of the individual bouncing point position from the centroid of the bouncing point (bouncing_point_to_cent).

3.5 Definition of episodes and phases

An episode (Figure 3) was defined as a sequence of phases that included the impact of the racket with the ball, the bounce of the ball on the wall, the bounce of the ball on the table, and the subsequent impact of the racket with the ball. Any individual action that comprised these sequences of phases was classified as a successful hit, while any action that did not complete the sequence was classified as a failure hit. Within one episode, three phases (Figure 3) could be identified. Phase one (p1) started from the initial impact until the first contact of the ball with the wall. The second phase (p2) encompassed the period between the first bounce and the second bounce of the ball on the table. Lastly, the third phase (p3) extended from the second bounce to the subsequent impact.

3.6 SHAP-analysis with linear, nonlinear, and sequential models

We analyzed the impact of features on hit rates using linear and nonlinear models. Missing values were filled in using the K-Nearest Neighbors imputer [Troyanskaya et al. 2001], except with XGBoost that could handle NaN missing values. All models' hyperparameter values were optimized using grid-search. The imbalanced proportion between unsuccessful (6.4%) and successful (93.3%) classes was addressed by adjusting the weight of each class during training. The Logistic regression model performed best with L1-regularization. At a max-depth of 10, the tree-bagging model achieved its peak accuracy. The XGBoost model performed best with a max-depth of 5. To capture the sequential nature of the input features, a Long Short Term Memory (LSTM) was utilized [Hochreiter and Schmidhuber 1997] and trained with PolyFocal loss [Leng et al. 2022] to handle class imbalance. The inputs of LSTM were a sequence of features from n number of consecutive episodes. We evaluated the performance of our model using multiple metrics, including balanced accuracy (ACC), Matthew Correlation Coefficient (MCC), and Geometric Mean Score (G-MEAN). To ensure a robust evaluation, we carried out five-fold subject-cross-validation, considering the skill levels of the individuals. We performed a five-fold subject-cross-validation twice for LSTM to ensure the stability of its performance.

To evaluate the contribution of each feature on the model's prediction, SHAP values were utilized: TreeShap with interventional feature perturbation (number of background sample was 500) [Lundberg et al. 2018] for XGBoost and GradientShap [Lundberg and Lee 2017] for LSTM. The confidence level was calculated by taking the dot product of baseline SHAP-value from each fold with the value obtained through bagging ("comparison" SHAP-value). In each fold, a model was first trained and the baseline SHAP-value was calculated using the test data. Then, a subset of the training data was selected, and the model was retrained from scratch. The "comparison" SHAP-value was then computed using the test data. A confidence level greater than 0 indicates that the SHAP-value results are robust.

4 RESULTS

4.1 Linear and non-linear models

Comparison between linear and non-linear models revealed that XGBoost outperformed both Bagging and Logistic regression in accurately determining whether a hit is considered optimal (success) or not (failure) (Tabel1). XGBoost achieved a 2% higher accuracy rate compared to the other algorithms, as well as a 0.09 higher MCC score than Logistic regression. These results indicate that the relationship between the features and outcomes is better modeled as a non-linear problem rather than a linear one.

To analyze the contribution of each feature to the outcome, we used SHAP values (confidence level equals to .85 ± .12; Detail: Figure 4). The average of the absolute SHAP values (Figure 4) showed that eye-movement-related features had a smaller effect compared to impact and action features. The feature with the highest SHAP-value was the steadiness of the pursuit of the ball (p3_stable). On the other hand, the duration and onset of ball pursuit had a lower

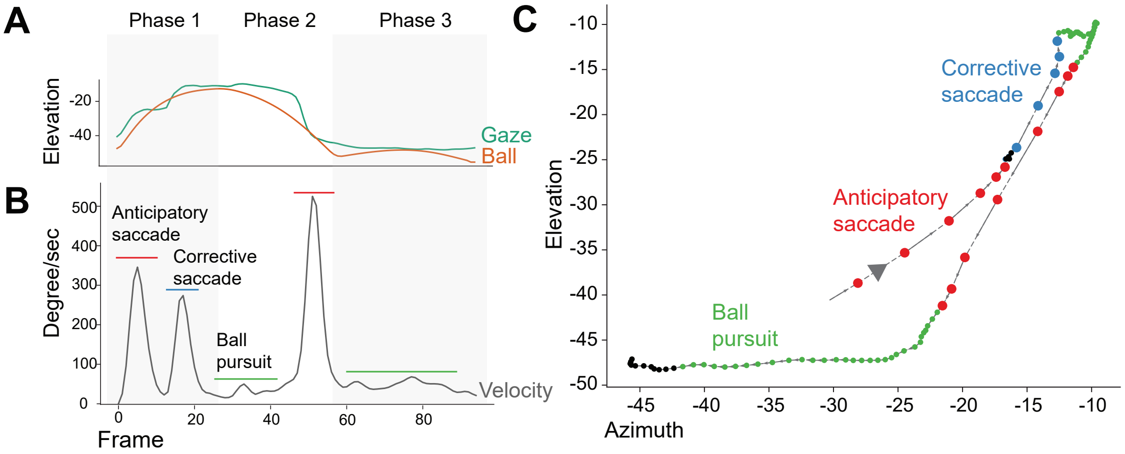


Figure 2: Gaze and ball vector in the visual field (all figures were generated using the same data). A: The elevation of gaze and ball vector in the visual field. B: The velocity of gaze vector in phase 1, 2, and 3. C: Each segment of gaze vector is labeled with a type of eye movement episode.

Table 1: The balanced accuracy (ACC), Matthew Correlation Coefficient (MCC), and Geometric Mean Score (G-MEAN) of linear (Logistic regression), nonlinear (Bagging and XGBoost), and sequential (LSTM) models. Important inputs excluded p1_al, p2_al, and p3_fx since those features showed no contribution to the final outcome of the model.

Model	Inputs	ACC	MCC	G-MEAN
Logistic (L1-regularization)	All-features	78.20	0.31	0.78
Bagging (DecisionTree-25 depth)	-	77.79	0.36	0.77
Xgboost (5-depth)	-	80.60	0.40	0.80
LSTM (3-sequence-length)	-	79.88	0.57	0.78
Xgboost (5-depth)	Important	80.50	0.40	0.80
LSTM (3-sequence-length)	-	80.78	0.56	0.80

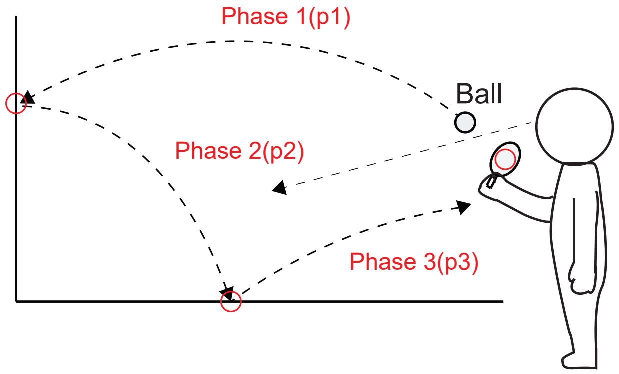


Figure 3: Three phases defined an episode.

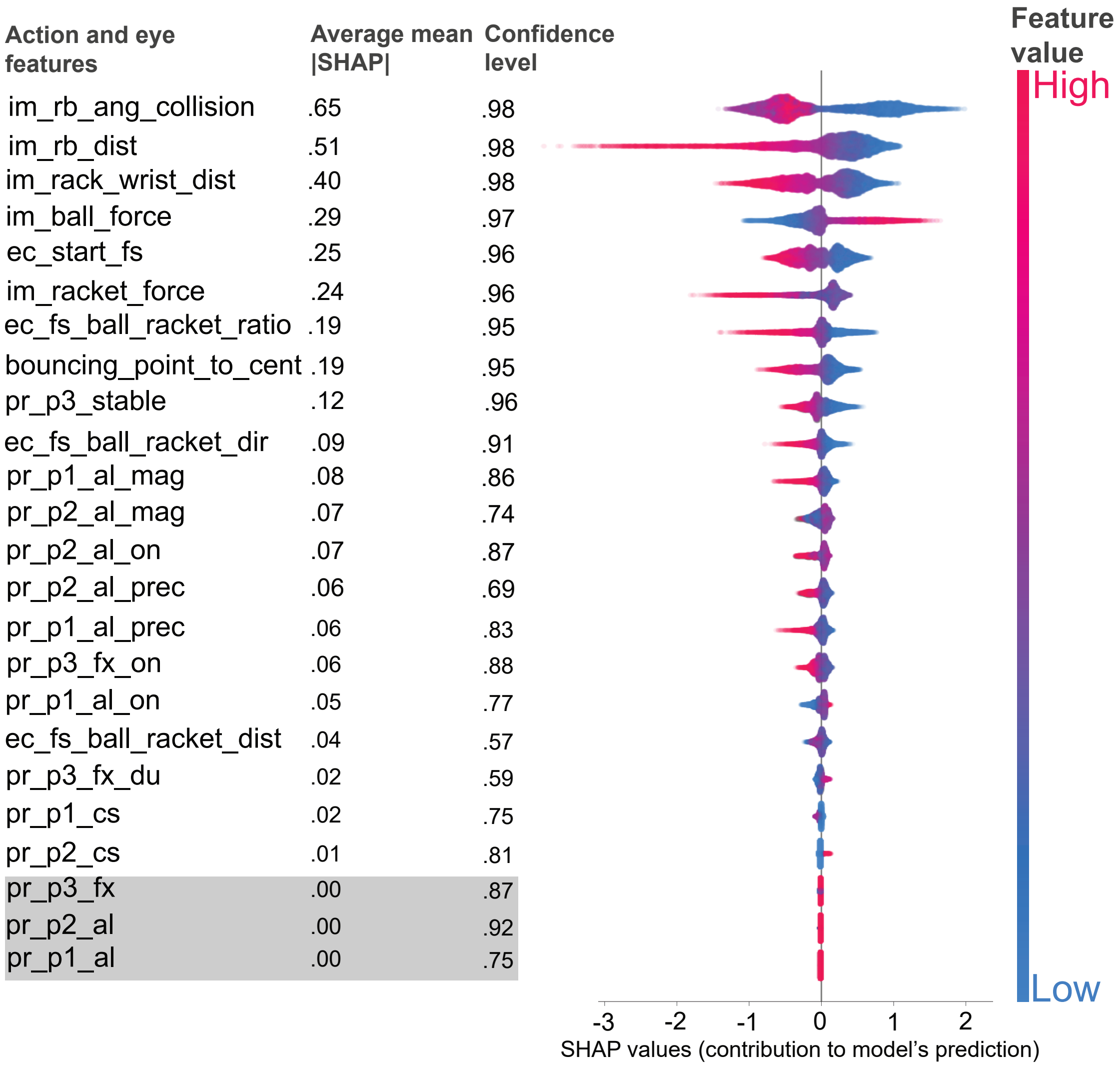
contribution to the final outcome. Additionally, whether the subject performed ball pursuit or not had almost no effect on the outcome.

In all phases, eye-movement features contributed to the success of a hit in all phases (Figure 5). However, the specific contributions of these features varied across each phase. In phase 1, an earlier onset of anticipatory look had a negative effect on the final outcome. In phase 2, an earlier onset showed a positive correlation. Additionally, the smaller magnitude in phase 1 had a positive linear relationship with the SHAP-value, while in phase 2, it had a non-linear correlation. However, the precision of anticipatory look in both phases showed a positive relationship with the SHAP-value. Moreover, the model's accuracy in predicting successful and unsuccessful hits was not significantly influenced by whether individuals performed anticipatory look in phase 1 and 2, and ball pursuit in phase 3.

4.2 Sequential model

The use of LSTM improved the performance of classifying the episodes. Although the balance accuracy and G-MEAN score did not change significantly, the MCC score increased by 0.16. The model's performance might have improved because of the balanced class proportion achieved during the striding process. However, upon examination, it was discovered that the class proportion remained imbalanced, with a success rate of 94.5% and an unsuccessful rate of 5.4%.

When using LSTM for data modeling, it's important to consider that the significance of input features at a specific time step can be influenced by preceding input data. This can lead to varying correlation trends between the SHAP-value of LSTM and the features. Our results showed a less clear correlation of the precision of



Action and eye features	Average mean |SHAP|	Confidence level
im_rb_ang_collision	.65	.98
im_rb_dist	.51	.98
im_rack_wrist_dist	.40	.98
im_ball_force	.29	.97
ec_start_fs	.25	.96
im_racket_force	.24	.96
ec_fs_ball_racket_ratio	.19	.95
bouncing_point_to_cent	.19	.95
pr_p3_stable	.12	.96
ec_fs_ball_racket_dir	.09	.91
pr_p1_al_mag	.08	.86
pr_p2_al_mag	.07	.74
pr_p2_al_on	.07	.87
pr_p2_al_prec	.06	.69
pr_p1_al_prec	.06	.83
pr_p3_fx_on	.06	.88
pr_p1_al_on	.05	.77
ec_fs_ball_racket_dist	.04	.57
pr_p3_fx_du	.02	.59
pr_p1_cs	.02	.75
pr_p2_cs	.01	.81
pr_p3_fx	.00	.87
pr_p2_al	.00	.92
pr_p1_al	.00	.75

Figure 4: The SHAP values of each feature from XGBoost. im, ec, and pc stand for impact, action, and eye-movement, respectively. The phases 1, 2, and 3 are represented by p1, p2, and p3. Features that have contributions close to zero are highlighted in grey. Negative SHAP values contribute to predicting unsuccessful episodes, while positive values correlate with successful episodes.

anticipatory look in phase 1 and the onset of ball pursuit in phase 3 at episode $t-2$ and $t-1$, as well as the magnitude of anticipatory look in phase 2 at episode t. However, the correlation of other eye-movement features, such as the precision of anticipatory look in phase 2 and the stability of ball pursuit in phase 3, remained consistent over time and exhibited a similar trend to the correlation between XGboost's SHAP-value (Figure 4) and eye-movement features.

Additionally, it is worth noting that compared to XGboost's SHAP-value (Figure 4), the confidence level of LSTM's SHAP-value (Figure 6) was lower. This result is likely caused by the stochastic process involved in training the LSTM and the change of data variance during bagging in computing the confidence level [Gawlikowski et al. 2021].

5 DISCUSSION AND CONCLUSION

This study examined how eye movement (gaze) and action (hand and racket) features are related to the success rate of hitting a ball in a dynamic and repetitive task. We analyzed 24 features related to hand, racket, and eye movements affecting ball trajectories. Importantly, while the outcome of the task can be affected by both the performance of individuals within the task and individual differences in skill levels, we addressed this issue through subject cross-validation.

We conducted comparisons between linear (Logistic Regression) and nonlinear (Bagging and XGBoost) models, as well as between sequential (LSTM) and nonsequential (XGBoost) models. Results revealed that the nonlinear models outperformed the linear models, and the sequential model outperformed the nonsequential model,

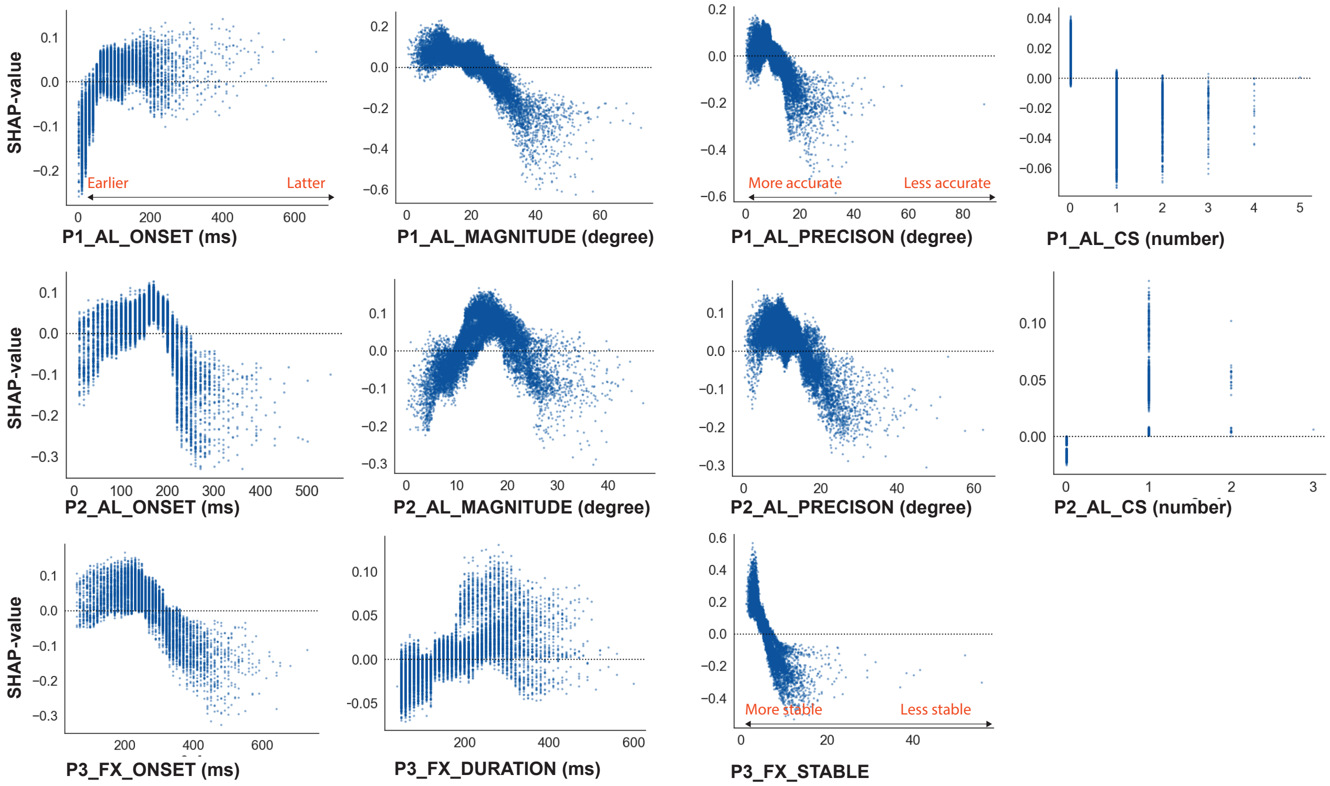


Figure 5: The SHAP values associated with eye-movement features. AL, CS, and FX refers to anticipatory saccade (anticipatory look), corrective saccades, and ball pursuit. P1, P2, and P3 represent phase 1, phase 2, and phase 3. Negative SHAP values indicate contribution to failure episodes, while positive values show contribution to success episodes. Dotted line indicates zero.

in terms of accuracy in predicting the success rate (Tabel1). These outcomes highlight the task's nonlinear and sequential characteristics.

To understand how each feature contributes to predicting the success rate, we analyzed SHAP values in both sequential (LSTM) and non-sequential (XGBoost) models. Not surprisingly, the action (hand and racket) features showed strong contributions to both models' predictions (Figure 4 and 5), while the eye movement (gaze) features had weaker overall contributions. Additionally, the contributions of eye movement features varied between sequential and non-sequential models (Figure 6). However, we observed consistent correlations between each eye movement feature and SHAP values across both sequential and non-sequential models, with confidence levels indicating moderate to high robustness of the results (Figure 4 and 6). Nonetheless, by examining the magnitudes and confidence levels of the SHAP values, we could identify that certain eye movement features were less related to the task performances (e.g., the occurrence of anticipatory look during phase 1 and phase 2, the magnitude of anticipatory look during phase 2, the timing and occurrence of ball pursuit in phase 3). The task-related eye movement features are: 1) timing and precision of anticipatory saccades and the number of corrective saccades during phase 1; 2) timing and precision of anticipatory saccades during phase 2, as well as the number of corrective saccades; 3) duration, and stability of ball pursuit in phase 3.

More specifically, in phase 1, the anticipatory saccades started later after the racket hit the ball in successful episodes. Second, the anticipatory saccades were more accurate, and fewer corrective saccades were made in successful episodes compared to failure episodes.

In phase 2, the optimal timings for anticipatory saccades in successful episodes occurred 100-200 ms after the ball bounced off the wall, compared to those in failed episodes. Additionally, these anticipatory saccades were more accurate. In contrast to phase 1, there were more corrective saccades in successful episodes compared to failure episodes in phase 2. Finally, in phase 3, successful episodes were characterized by greater stability in tracking the ball trajectory and a longer duration of ball pursuit, compared to failure episodes.

Our quantitative results, based on the contributions of eye movement features to success and failure, can be interpreted in line with previous findings. Specifically, previous studies have suggested that anticipatory looks after the ball hits reflect the participants' expectations about their strokes [Land and Furneaux 1997]. Thus, one possible interpretation of our observation during phase 1 is that participants' improved stroke expectations enhanced participants'

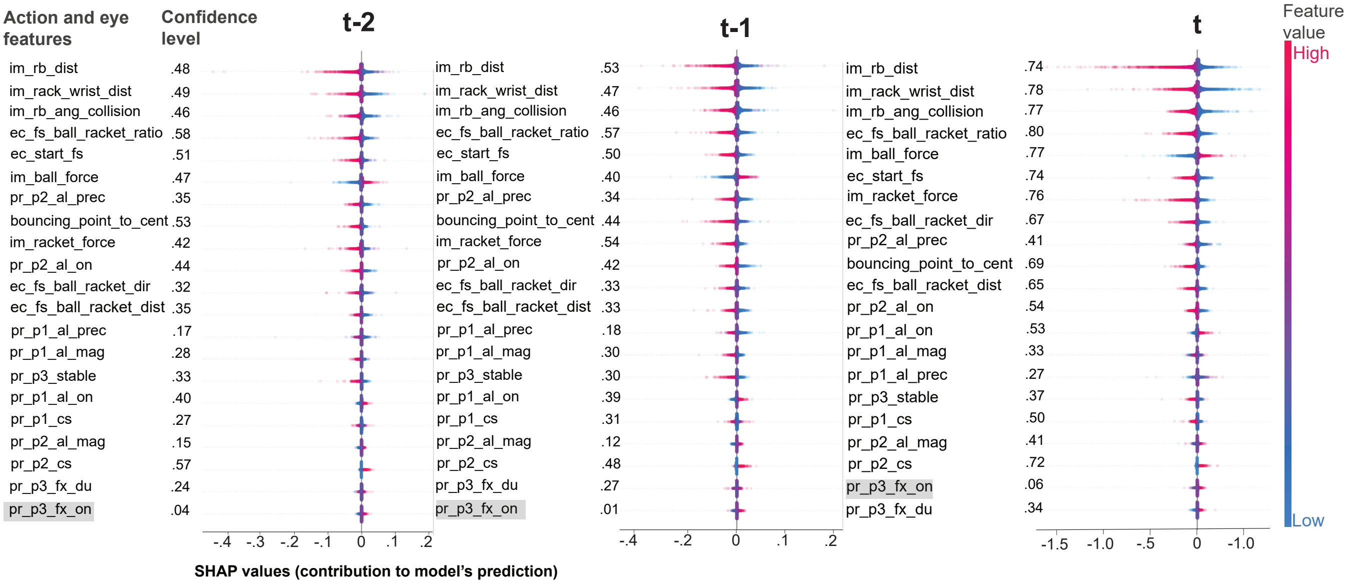


Figure 6: Features contribution using LSTM model from each time to the final outcome. The current episode is denoted as t, while the preceding episodes are denoted as $t-1$ and $t-2$. Across time, the onset of ball pursuit in phase 3 had low confidence, highlighted in grey.

precision and efficiency in their anticipatory gaze. As participants become more certain about controlling their strokes, they can fixate their gaze on an intended location and wait to perform an anticipatory saccade. This may have led to a lower magnitude and later initiation of anticipatory gaze, as well as fewer corrective saccades.

Our observations during phase 2 are consistent with previous studies demonstrating that participants could make accurate anticipatory looks at the bouncing point at the right timings in ball intersection task [Diaz et al. 2013a,b; Hayhoe et al. 2012; Land and McLeod 2000]. During phase 2, the participants' anticipatory gazes were likely influenced by the visual characteristics of the ball, as it bounced off the wall, rather than by their expectations about their strokes from phase 1. Consequently, while we observed that a higher number of corrective saccades correlated either negatively or positively with participant performance in phases 1 and 2 respectively, the negative correlation in phase 1 might suggest a failure in forming stroke expectations, whereas the positive correlation in phase 2 could be attributed to anticipating the ball's trajectory.

Our observations in phase 3, characterized by longer and more stable ball pursuit, may be linked to successful anticipation in phase 2. Additionally, these findings align with previous research in aiming sports such as basketball shooting and billiards, where a phenomenon known as the 'quiet eye'—a prolonged fixation on a target—has been shown to improve success rates [Vickers 1996; Williams et al. 2002].

One of the main limitations of this study is the difficulty in establishing a causal relationship between eye-movement and the final outcome of an action. While a combination of SHAP-analysis and prediction models like XGBoost can help us understand how each feature contributes to the outcome, they cannot determine how manipulating input features will directly cause a change in the outcome [Moraffah et al. 2020]. To investigate this, controlled experiments are necessary, which consider confounding variables, and use causal inference models to model the relationship between inputs and outcomes.

In conclusion, this study utilized a motion capture system and wearable eye-tracking technology to evaluate individuals' eye-movement and interception skills in dynamic tasks. We introduced a quantitative analysis framework accompanied by a comprehensive dataset. We used SHAP-analysis to show how eye movement and action features significantly influence the outcomes of dynamic tasks. Our results revealed that participants' performances are most accurately described by nonlinear and sequential models. The validity of our analysis is supported by its consistency with previous studies, particularly regarding the characteristics of anticipatory looks at the ball's bouncing points. Overall, our research lays the groundwork for a holistic approach to understanding how humans strategize and plan their actions based on visual eye-movement in realistic, dynamic settings.

ACKNOWLEDGMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2117 (ID: 422037984). We would also like to thank our research assistants, Riya and Mihane, and Kano's lab member for their help in collecting data. Last, we extend our gratitude to all the members of the Centre for the Advanced Study of Collective Behaviour, particularly Mathias Günther and Jan Peter, for their assistance in setting up our motion capture lab.

REFERENCES

Richard Andersson, Linnea Larsson, Kenneth Holmqvist, Martin Stridh, and Marcus Nyström. 2017. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior research methods* 49 (2017), 616–637.

Ruiqiao Bai, Jacqueline CK Lam, and Victor OK Li. 2023. What dictates income in New York City? SHAP analysis of income estimation based on Socio-economic and Spatial Information Gaussian Processes (SSIG). *Humanities and Social Sciences Communications* 10, 1 (2023), 1–14.

Anna Belardinelli, Madeleine Y Stepper, and Martin V Butz. 2016. It's in the eyes: Planning precise manual actions before execution. *Journal of vision* 16, 1 (2016), 18–18.

Benedetta Cesqui, Maura Mezzetti, Francesco Lacquaniti, and Andrea d'Avella. 2015. Gaze behavior in one-handed catching and its relation with interceptive performance: What the eyes can't tell. *PLoS One* 10, 3 (2015), e0119445.

Sergio Delle Monache, Francesco Lacquaniti, and Gianfranco Bosco. 2015. Eye movements and manual interception of ballistic trajectories: effects of law of motion perturbations and occlusions. *Experimental Brain Research* 233 (2015), 359–374.

Gabriel Diaz, Joseph Cooper, and Mary Hayhoe. 2013a. Memory and prediction in natural gaze control. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 1628 (2013), 20130064.

Gabriel Diaz, Joseph Cooper, Constantin Rothkopf, and Mary Hayhoe. 2013b. Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of vision* 13, 1 (2013), 20–20.

Andrew Duchowski, Eric Medlin, Nathan Cournia, Hunter Murphy, Anand Gramopadhye, Santosh Nair, Jeenal Vorah, and Brian Melloy. 2002. 3-D eye movement analysis. *Behavior Research Methods, Instruments, & Computers* 34 (2002), 573–591.

Jolande Fooken and Miriam Spering. 2020. Eye movements as a readout of sensorimotor decision processes. *Journal of Neurophysiology* 123, 4 (2020), 1439–1447.

Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. 2021. A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342* (2021).

Mary Hayhoe, Neil Mennie, Brian Sullivan, and Keith Gorgos. 2005. The Role of Internal Models and Prediction in Catching Balls. *From Reactive to Anticipatory Cognitive Embodied Systems*.

M. M. Hayhoe, T. McKinney, K. Chajka, and J. B. Pelz. 2012. Predictive eye movements in natural vision. *Exp Brain Res* 217, 1 (2012), 125–36. https://doi.org/10.1007/s00221-011-2979-2

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

Michael Land and Benjamin Tatler. 2009. *Looking and Acting: Vision and eye movements in natural behaviour.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570943.001.0001

M. F. Land and S. Furneaux. 1997. The knowledge base of the oculomotor system. *Philos Trans R Soc Lond B Biol Sci* 352, 1358 (1997), 1231–9. https://doi.org/10.1098/rstb.1997.0105

Michael F Land and Peter McLeod. 2000. From eye movements to actions: how batsmen hit the ball. *Nature neuroscience* 3, 12 (2000), 1340–1345.

Zhaoqi Leng, Mingxing Tan, Chenxi Liu, Ekin Dogus Cubuk, Xiaojie Shi, Shuyang Cheng, and Dragomir Anguelov. 2022. Polyloss: A polynomial expansion perspective of classification loss functions. *arXiv preprint arXiv:2204.12511* (2022).

Stan Lipovetsky and Michael Conklin. 2001. Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry* 17, 4 (2001), 319–330.

Stephen G Lisberger. 2010. Visual guidance of smooth-pursuit eye movements: sensation, action, and what happens in between. *Neuron* 66, 4 (2010), 477–491.

Simon P Liversedge and John M Findlay. 2000. Saccadic eye movements and cognition. *Trends in cognitive sciences* 4, 1 (2000), 6–14.

Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence* 2, 1 (2020), 56–67.

Scott M Lundberg, Gabriel G Erion, and Su-In Lee. 2018. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888* (2018).

Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

David L Mann, Hiroki Nakamoto, Nadine Logt, Lieke Sikkink, and Eli Brenner. 2019. Predictive eye movements when hitting a bouncing ball. *Journal of vision* 19, 14 (2019), 28–28.

Raha Moraffah, Mansooreh Karami, Ruocheng Guo, Adrienne Raglin, and Huan Liu. 2020. Causal interpretability for machine learning-problems, methods and evaluation. *ACM SIGKDD Explorations Newsletter* 22, 1 (2020), 18–33.

Leigh A Mrotek and John F Soechting. 2007. Target interception: hand–eye coordination and strategies. *Journal of Neuroscience* 27, 27 (2007), 7297–7309.

Jean-Jacques Orban de Xivry and Philippe Lefevre. 2007. Saccades and pursuit: two outcomes of a single sensorimotor process. *The Journal of physiology* 584, 1 (2007), 11–23.

Sam Perochon, J Matias Di Martino, Kimberly LH Carpenter, Scott Compton, Naomi Davis, Brian Eichner, Steven Espinosa, Lauren Franz, Pradeep Raj Krishnappa Babu, Guillermo Sapiro, et al. 2023. Early detection of autism using digital behavioral phenotyping. *Nature Medicine* 29, 10 (2023), 2489–2497.

Cyril Rashbass. 1961. The relationship between saccadic and smooth tracking eye movements. *The Journal of physiology* 159, 2 (1961), 326.

Vishnu Sarpeshkar, Bruce Abernethy, and David L Mann. 2017. Visual strategies underpinning the development of visual–motor expertise when hitting a ball. *Journal of Experimental Psychology: Human Perception and Performance* 43, 10 (2017), 1744.

Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics* 17, 6 (2001), 520–525.

J. N. Vickers. 1996. Visual control when aiming at a far target. *J Exp Psychol Hum Percept Perform* 22, 2 (1996), 342–54. https://doi.org/10.1037//0096-1523.22.2.342

Daqi Wang, Chengdong Zhang, Bei Wang, Bin Li, Qiang Wang, Dong Liu, Hongyan Wang, Yan Zhou, Leming Shi, Feng Lan, et al. 2019. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nature communications* 10, 1 (2019), 4284.

A Mark Williams, Robert N Singer, and Shane G Frehlich. 2002. Quiet eye duration, expertise, and task complexity in near and far aiming tasks. *Journal of motor behavior* 34, 2 (2002), 197–207.

Han Yu, Huaxiang Deng, Jiahui He, Jay D Keasling, and Xiaozhou Luo. 2023. UniKP: a unified framework for the prediction of enzyme kinetic parameters. *Nature Communications* 14, 1 (2023), 8211.