THESIS

USING EYE GAZE TO AUTOMATICALLY IDENTIFY FAMILIARITY

Submitted by

Iliana Castillon

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2024

Master's Committee:

Advisor: Nathaniel Blanchard

Sarath Sreedharan
Anne M. Cleary

ABSTRACT

USING EYE GAZE TO AUTOMATICALLY IDENTIFY FAMILIARITY

Understanding internal cognitive states, such as the sensation of familiarity, is crucial not only in the realm of human perception but also in enhancing interactions with artificial intelligence. One such state is the experience of familiarity, a fundamental aspect of human perception that often manifests as an intuitive recognition of faces or places. Automatically identifying cognitive experiences could pave the way for more nuance in human-AI interaction. While other works have shown the feasibility of automatically identifying other internal cognitive states like mind wandering using eye gaze features, the automatic detection of familiarity remains largely unexplored. In this work, we employed a paradigm from cognitive psychology to induce feelings of familiarity. Then, we trained machine learning models to automatically detect familiarity using eye gaze measurements, both in experiments with traditional computer use (e.g., eye tracker attached to monitor) and in virtual reality settings, in a participant independent manner. Familiarity was detected with a Cohen's kappa value, a measurement of accuracy corrected for random guessing, of 0.22 and 0.21, respectively. This work showcases the feasibility of automatically identifying feelings of familiarity and opens the door to exploring automated familiarity detection in other contexts, such as students engaged with a learning task while interacting with an intelligent tutoring system.

ACKNOWLEDGEMENTS

For Nate Blanchard, a great mentor who always made learning fun. Anne Cleary, who's guidance and expertise made this possible. Asa Davis, who reminded me to go outside and be happy. Videep Venkatesha, I couldn't ask for a better collaborator and friend. Lastly for my parents, who provided endless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Familiarity, a sense characterized by a subtle yet compelling feeling of recognition, plays a crucial role in the intricate processes of human memory and information retrieval. Familiarity occurs when individuals encounter situations, events, places, or people and experience a sense of recognition [3]. Anyone who has done a crossword has likely experienced the sense that they know the answer that corresponds with the clue but cannot retrieve it — this is an example of familiarity, and the inability to remember the clue is a recollection failure.

Extensive research in cognitive psychology has focused on the internal state of familiarity, frequently linked with recognition memory. Dual process theories of recognition propose that familiarity is one of two processes involved in recognizing prior experience. The other process, recollection, involves recalling a specific prior instance in which the present stimulus was encountered [4, 5]. While various theories propose different relationships between familiarity and recollection, recently its been proposed that the initial sense of familiarity might prompt the search of memory, eventually leading to recollection [6–8]. Based on this overarching theoretical framework, our current study aimed to capture instances of familiarity as soon as participants sensed it, irrespective of whether this initial familiarity led to successful recall or not.

Previous studies have shown that machine learning algorithms can distinguish between seen and unseen stimuli based on eye movement patterns associated with the onset of recognition memory tests [9]. These memory-related eye movement patterns often occur without participants being consciously aware of them [10]. However, at this time no works have investigated if machine learning models can use eye gaze patterns to automatically identify someone experiencing familiarity. In this thesis, I explore exactly this — are eye gaze patterns linked to familiarity detection? What gaze features as associated with familiarity identification? Are these features consistent across contexts, i.e., desktop computer use and VR?

The ability to detect the subjective sense of familiarity could have practical applications, such as the development of intelligent virtual tutoring systems. Recent studies suggest that experiencing familiarity during recall failure is linked to heightened curiosity and a desire for information-seeking [7]. Therefore, detecting moments of familiarity could assist in creating tutoring systems that respond to students' varying levels of curiosity and inclination to seek information, thereby enhancing their learning experience. For instance, research in the testing effect literature indicates that students learn more effectively when they are required to generate information independently rather than passively receiving it [11].

Furthermore, studies have demonstrated that individuals experiencing heightened curiosity are more motivated to discover answers on their own, rather than being provided with them outright [12]. An intelligent tutoring system could capitalize on this by adaptively encouraging self-directed information-seeking when the learner appears ready and motivated based on detected subjective familiarity. For example, when subjective familiarity suggests that the learner is close to generating an answer independently, the system could offer hints or cues instead of providing the full answer. This approach optimizes the balance between presenting new information and fostering long-term learning through information retrieval, rather than passive reception.

Training models to automatically detect familiarity necessitates a substantial dataset. Therefore, a critical aspect of our study relied on an experimental paradigm from cognitive psychology known for inducing feelings of familiarity. Previous research has demonstrated that exposure to highly similar scenes can evoke a sense of familiarity with new scenes [1, 13, 14]. Our procedure for eliciting familiarity follows that outlined by Cleary et al. (2012) and Okada et al. (2023), inducing familiarity from 3D scenes in Virtual Reality (VR) that share the same spatial configuration as previously viewed scenes [1, 8]. We adopted the same VR procedure and Unity stimuli as Okada et al. (2023), with minor modifications to the Unity program to incorporate eye tracking. This study marks the first attempt to establish an association between specific eye movement patterns and moments of subjectively sensing familiarity.

Chapter 2

Prior Works

While this research introduces a fresh approach by aiming to automatically identify instances of subjective familiarity during scenes reminiscent of past experiences, the concept of automatically detecting subjective recognition and other cognitive states is not unprecedented. Previous studies have demonstrated its potential. For instance, in one study, a model utilizing eye gaze successfully classified whether participants had encountered an image previously, achieving an average accuracy of 68.7% [15]. In Nishimura et al. (2012), models utilized gaze data collected during participants' exposure to an image to determine if they had seen the same image earlier in the experiment, irrespective of whether participants explicitly acknowledged recognition. In contrast, the objective of the present study is to exclusively classify instances where participants are encountering feelings of familiarity. Placing greater emphasis on detecting the internal state of participants. Additionally, the familiarity in this work is evoked from configurally similar, but non-identical scenes.

A significant portion of research on internal state detection has focused on mind wandering, which involves diverting attention away from a specific task. Investigating models that incorporate or depend on features other than gaze has been a key aspect of this research [16–18], although ultimately models built using gaze-based features have been the most effective at detecting mind-wandering [19,20]. Similarly to this, the features we used in this work are all gaze-based. Previous research has attempted to detect mind wandering either during reading [17,21,22], or while watching videos [23, 24]. In addition to global gaze features, certain studies integrated local features derived from the gaze direction concerning the text being read [22] or specific points of interest in the film being watched [23]. The features in our model resemble the global features in these studies, which were context-independent. Moreover, most of these studies employed probe-based detection of mind wandering [17,22,24], given that mind wandering often happens without immediate individual awareness. Nevertheless, self-caught reports of mind wandering have also been integrated [21], and we predominantly utilized this method in our study to precisely identify the moment when participants experienced familiarity.

This work extends the existing body of knowledge by leveraging previous findings and methodologies used in detecting mind wandering to now identify another cognitive state: the sensation of familiarity.

Chapter 3

Two-Dimensional Familiarity Dataset

3.1 Materials and Procedure

3.1.1 Familiarity Task

Expanding upon prior research that utilized a virtual tour paradigm to elicit sensations of familiarity in controlled environments [8, 25], subjects in this study participated in virtual tours of various settings via video presentations displayed on a computer monitor. These presentations featured walkthroughs of virtual landscapes.

In previous work employing the virtual tour task [25], participants encountered unfamiliar settings during the study phase. As they navigated through these study phase scenes via virtual tours, the name of each scene was audibly announced through speakers. For example, if they were viewing an aquarium, a voice would identify it, e.g. "This is a aquarium." Participants were then required to recall each study phase scene along with its corresponding name.

During the test phase, participants were presented with entirely new settings, some of which shared the same spatial layout as scenes previously toured during the study phase. For instance, a museum scene might mirror the spatial arrangement of elements seen in an earlier toured courtyard scene. An example of this is illustrated in Figure 3.3. It's important to note that while scenes in the test phase were otherwise new, they could share a similar spatial configuration with scenes from the study phase. Unlike the study phase, no audio accompanied the viewing of the test phase videos.

To ensure consistent spatial configurations from the study phase to the test phase without directly replicating objects, a grid layout was implemented to create scenes that were spatially mapped but visually distinct, as depicted in Figure 3.1.

In this paradigm, participants underwent two study-test blocks. Each study phase comprised 16 study scenes, while each test phase consisted of 32 test scenes. Among these test scenes, 16 scenes

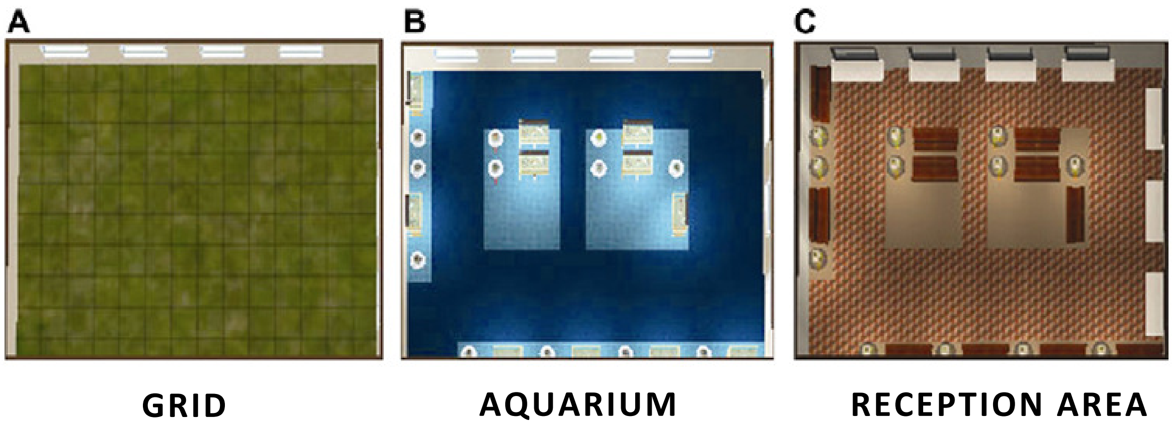

Figure 3.1: (A) Grid used to create same spatial configuration; (B) Sample "Aquarium" study scene; (C) Sample "Reception Area" test scene corresponding to "Aquarium" study scene [1].

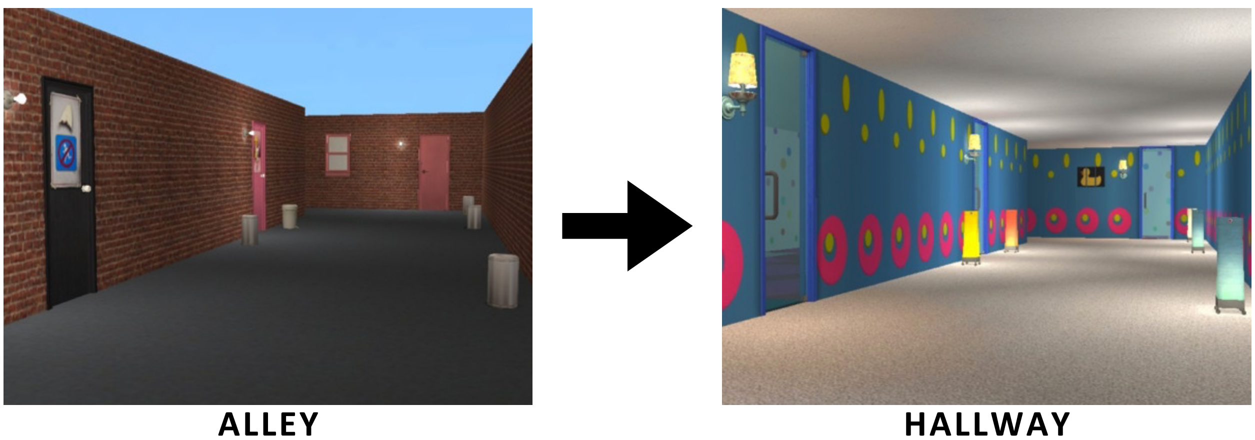

Figure 3.2: Sample "Alley" study Scene and configurally similar "Hallway" test scene [1].

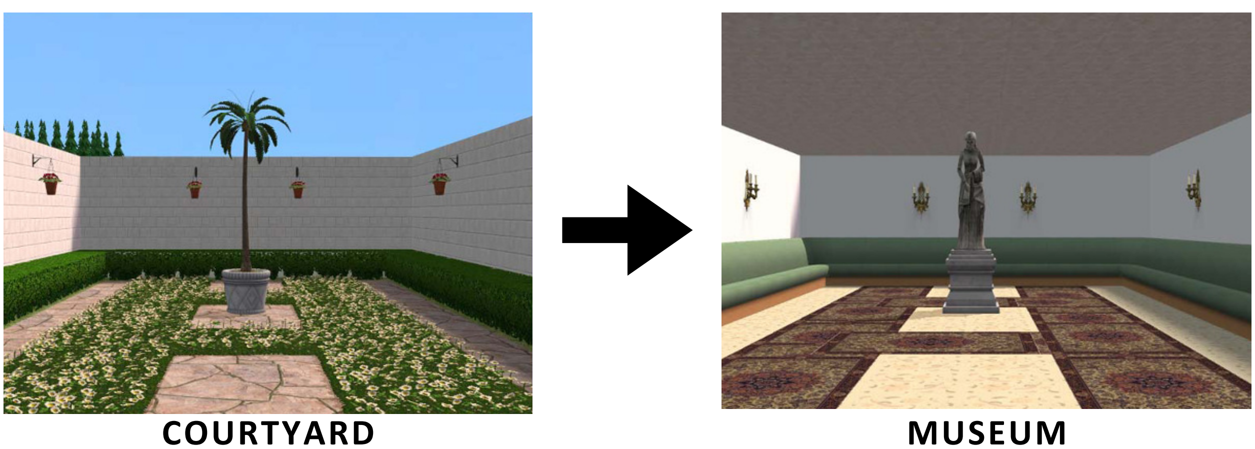

Figure 3.3: Sample "Courtyard" study scene and configurally similar "Museum" test scene [1].

corresponded to studied scenes, and the remaining 16 corresponded to unstudied scenes in terms of their spatial layout. To ensure counterbalancing across participants, four experiment versions were utilized to distribute stimuli evenly across different conditions.

Previous research utilizing this paradigm has demonstrated that participants are more likely to experience a sense of familiarity with a novel test scene that shares a spatial layout with a study scene, even when they are unable to identify the source of that familiarity [1, 7, 8, 25].

The present study aimed to employ the same paradigm, while also measuring eye gaze features as participants completed the virtual tour task. The specific tours utilized in this study were those employed by Okada et al. (2023) in their Experiments 2a and 2b [8].

3.1.2 Participants

The participants in this study comprised 61 undergraduate students enrolled in a psychology class at Colorado State University. They volunteered to take part in the study in exchange for course credit. Sample-size was based on Experiments 2a and 2b of [8].

3.1.3 Procedure

Participants were brought into a test room where they sat at a desk with a computer connected to an eye tracker and webcam. A photo of experiment hardware can be found in Figure 3.4. At the beginning of the experiment, the eye tracker was calibrated for each participant.

Once the task commenced, participants were directed to watch a series of study videos, during which they were pulled on rails through various scenes as previously discussed in the familiarity task. When the test phase commenced, participants were instructed to press the 'up arrow key' on the keyboard whenever they experienced a sense of familiarity. This key was labeled with a bright yellow sticker to facilitate easy identification. Participants were instructed to keep their finger on the key, ready to press it, to avoid the need to look down at the keyboard.

The task duration for participants was approximately an hour. Each participant completed two study-test blocks, involving the viewing of a total of 96 videos.

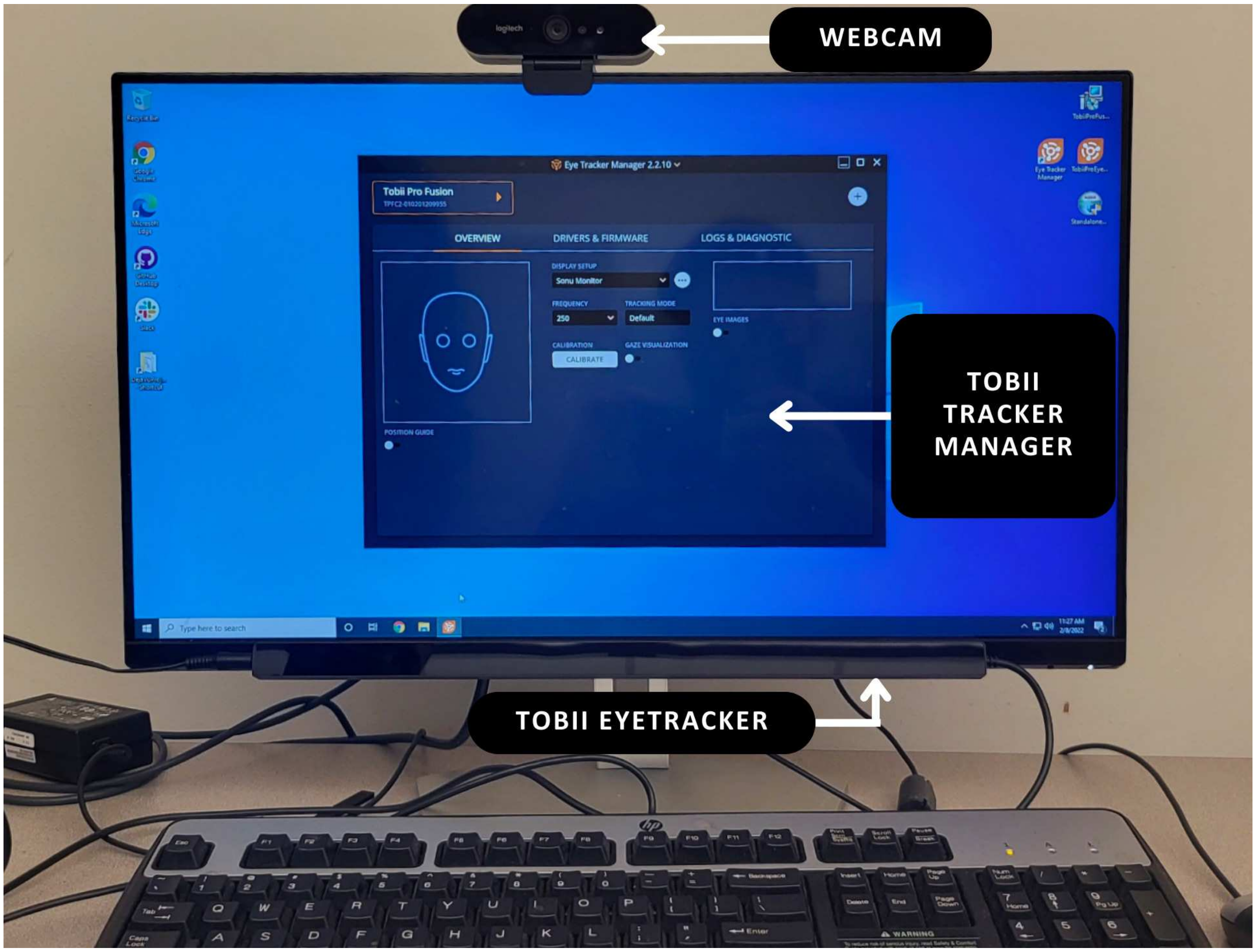

Figure 3.4: Experiment hardware including webcam and eyetracker [2].

3.1.4 Eye Tracking and Feature Generation

In this study, we employed the Tobii Pro Fusion eye tracker in conjunction with PyTrack, which is an end-to-end open-source solution designed for the analysis and visualization of eye tracking data [26]. The Tobii Pro Fusion eye tracker captures 250 images per second (250 Hz) and is equipped with two built-in pupil tracking modules. PyTrack facilitated the extraction of parameters of interest, including blinks, saccade count, average pupil size, etc. A list of all the extracted features can be found in Figure 3.5.

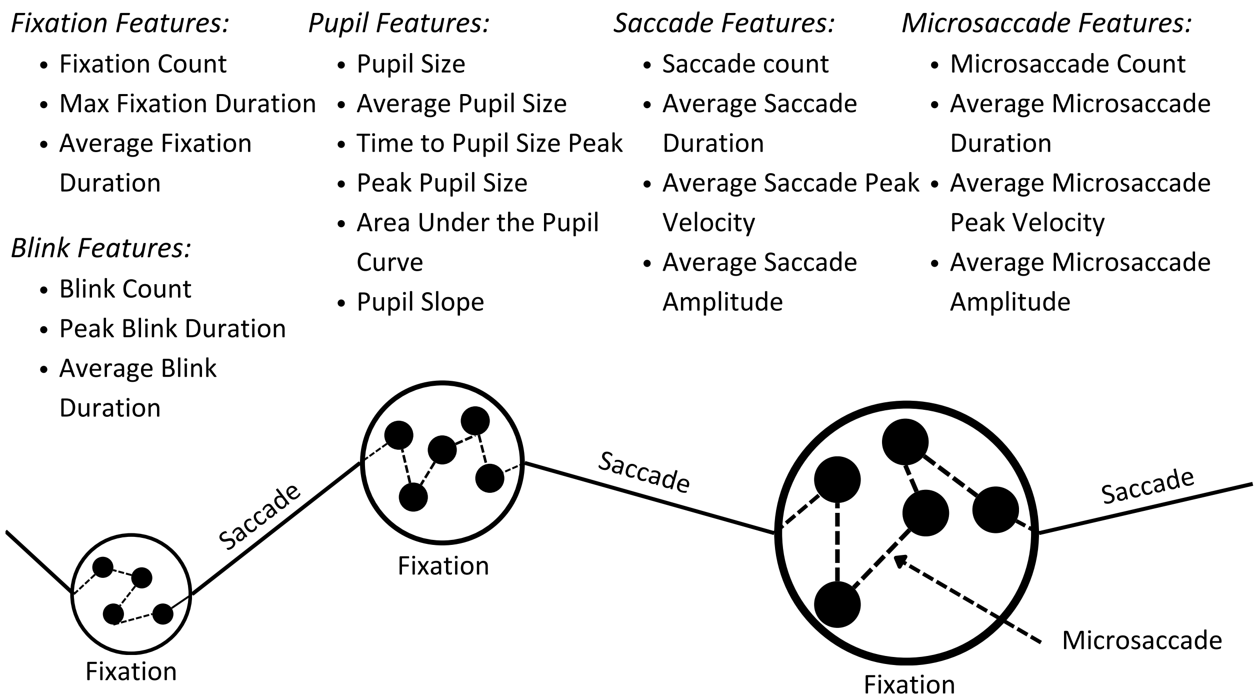

Figure 3.5: Eye gaze features.

3.2 Data Preprocessing

Participants reported a total of 698 instances of familiarity via self-report. However, we recognized the potential for bias introduced by the act of self-reporting, as participants may exhibit patterns such as looking down to identify the up key, which could influence our findings. To mitigate this possibility, we implemented a 2-second buffer of data removal from the time before the button was pressed. This buffer was determined by analyzing the videos for any indication of participants looking down to self-report. Our analysis indicated that a minimum buffer of 1.5 seconds was necessary, but we extended it to 2 seconds as a precautionary measure. Subsequently, we examined the 1-second window of time preceding the two-second buffer and utilized this window to extract eye gaze features, as depicted in Figure 3.6.

Unfortunately, in order to extract features, we required at least three seconds of data prior to the key press (comprising the two-second buffer and the one-second window). Among the 698 instances of self-reported familiarity, only 263 instances had the requisite three seconds of data

Table 3.1: Descriptive statistics of eye gaze features (2 second buffer and 1 second window).

	Familiarity				Non-Familiarity			
Feature	**Mean**	**SD**	**Min**	**Max**	**Mean**	SD	**Min**	**Max**
Fixation Count	3.34	2.53	0.00	5.00	2.77	2.31	0.00	4.00
Fixation Duration (ms)	42.70	47.04	0.00	244.00	46.49	50.78	0.00	245.00
Saccade Count	2.27	1.65	0.00	3.00	1.94	1.45	0.00	3.00
Saccade Duration (ms)	68.85	78.01	0.00	246.00	77.28	86.40	0.00	246.00
Microsaccade Count	0.35	0.64	0.00	1.00	0.42	0.72	0.00	1.00
Microsaccade Duration (ms)	2.54	4.23	0.00	14.00	2.76	4.15	0.00	12.50
Blink Count	0.85	0.46	0.00	2.00	0.74	0.52	0.00	2.00
Blink Duration (ms)	129.57	98.35	0.00	246.00	104.38	97.92	0.00	246.00

before the button press. This discrepancy likely occurred because the onset of familiarity occurred more rapidly than initially anticipated.

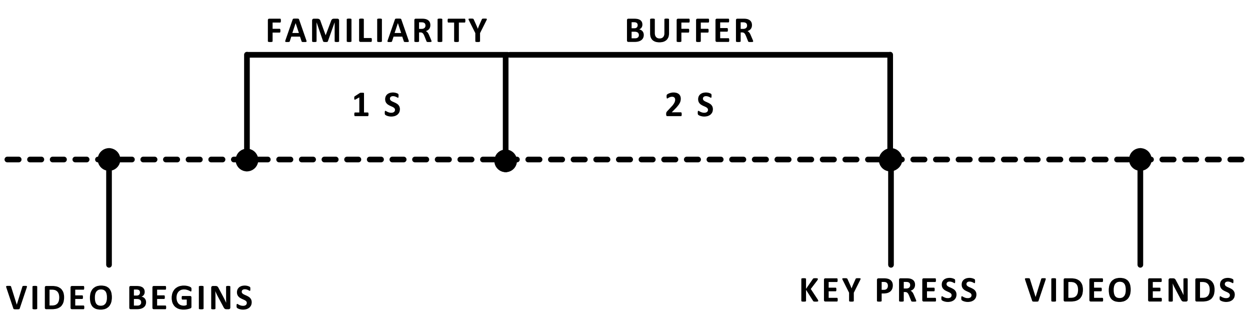

Figure 3.6: Familiarity instances timeline for 2D condition.

To generate negative instances, we randomly selected test scenes where participants did not report experiencing familiarity. To ensure a balanced dataset, we sampled an equal number of negative instances per participant as we had valid positive instances. This process yielded a dataset consisting of 263 valid instances of familiarity and 263 negative instances. The duration of each of these instances was one second.

Table 3.2: SVC Model Results

Model	Buffer	Window	Cohen's Kappa	F1 Score	Accuracy
SVC	2 s	1 s	0.22 (0.42)	0.56 (0.20)	0.61 (0.21)

3.3 Training and Evaluation

Hyperopt [27] was used for distributed hyperparameter optimization in our model search. Among the classification algorithms we evaluated were AdaBoost (AB), Naive Bayes (NB), Logistic Regression (LR), Support Vector Classifier (SVC), Random Forrest (RF), and K-Nearest Neighbors (KNN).

To train and test each machine learning model, we employed Leave-One-Participant-Out Cross-Validation. For this evaluation protocol, n is equal to the number of participants in the dataset. We train each model on n-1 participants and test on the participant that was left out of training. This process is repeated until every participant's samples have been used as the test set. Evaluation metrics are averaged across all n folds.

Our final model selection was based on identifying the model with the highest average Cohen's Kappa across all n folds [28]. Cohen's Kappa ranges between –1 and 1, with 0 representing chance performance (equivalent to random guessing), 1 indicating perfect performance, and –1 corresponding to perfectly imperfect performance (where the classifier always predicts the opposite label).

3.4 Results

Using the Support Vector Classifier *(C = 95, gamma = 0.001, kernel = rbf)* algorithm, we trained a machine learning model that identified familiarity with a Cohen's Kappa of 0.22 (SD = 0.42) and an F1 score of 0.56 (SD = 0.23). This performance is in line with other predictors of internal cognitive states, such as mind wandering [22].

The substantial standard deviation values observed among metrics, as illustrated in Table 3.2, indicates significant variation in detection performance among participants. This varied performance indicates individual differences in eye gaze patterns that emerge during the experience of familiarity. Additionally, the number of instances reported by participants appears to impact the model's performance. Participants who reported only one or two instances of familiarity exhibited extreme ends of the Cohen's Kappa distribution. However, the majority of participants' kappa scores ranged from 0 to 0.65, as depicted in Figure 3.7.

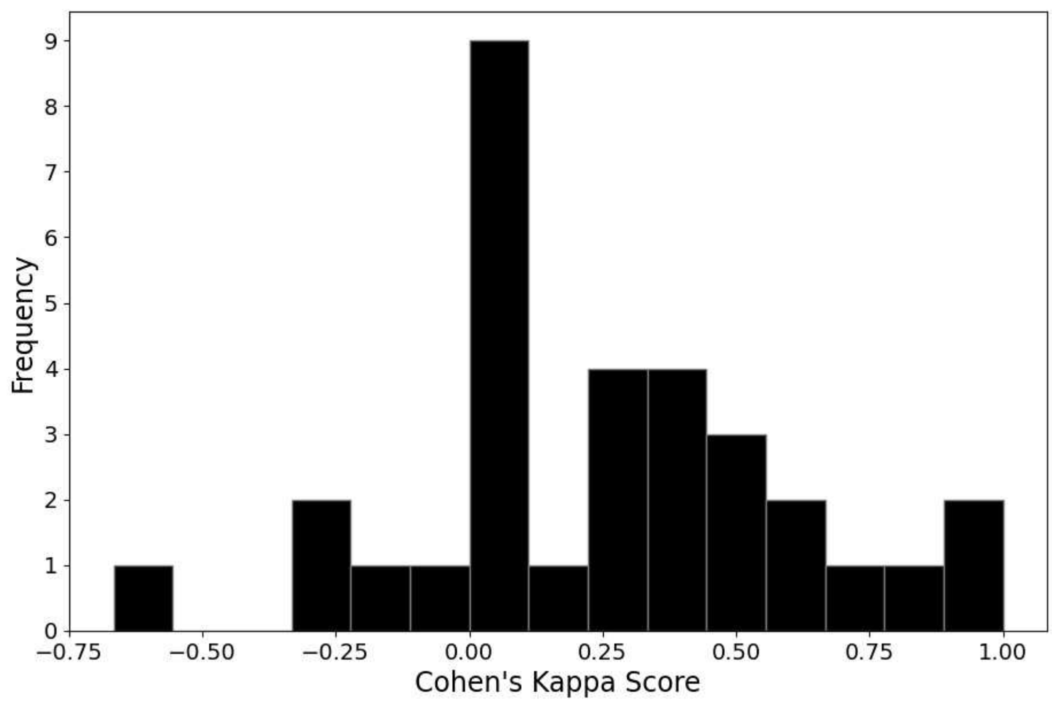

Figure 3.7: Distribution of Cohen's Kappa scores for 2D condition.

Additionally we experimented with training the same SVC model using only one feature at a time. The 3 best performing features were average blink duration, blink count, and fixation count. These model results can be seen in Table 3.3. Along with this, we investigated which features demonstrated statistically significant variances between instances of positive and negative familiarity. To accomplish this, we performed hypothesis tests and pinpointed found that there was a significant difference between fixation count, blink count, average blink duration, saccade count, and microsaccade velocity between positive and negative familiarity instances. Additional descriptive statistics for features can be found in Table 3.1. All of the features that had the best

Table 3.3: Best Performing Features for 2D Condition

Feature	Kappa	F1
avg_blink_duration	0.17 (0.26)	0.55 (0.26)
blink_cnt	0.16 (0.39)	0.52 (0.29)
fix_cnt	0.12 (0.34)	0.52 (0.26)

performance when trained on individually, were also the ones with a difference that was statistically significant.

Chapter 4

Virtual Reality Familiarity Dataset

4.1 Materials and Procedure

4.1.1 Familiarity Task

The structure of the VR experiment closely resembled the two-dimensional study described in section 3.1.1, encompassing both "study" and "test" phases. However, in this version of the task, static versions of the scenes used in Okada et al.'s (2023) Experiment 3 were created to facilitate the measurement of eye movement patterns [8]. Participants were not guided through the scenes but were instead positioned in the center of a scene and given the ability to turn their heads to explore from that position. Additionally, participants were not probed after every scene ended. Instead, they were instructed to press a button on the handheld VR remote controller to indicate they were experiencing a sense of familiarity. On average, each scene lasted 46 seconds.

4.1.2 Participants

The participants in this study consisted of 26 undergraduate students enrolled in a psychology class at Colorado State University. They volunteered to participate in the study in exchange for course credit. Sample size was based on Experiment 3 of [8].

4.1.3 Procedure

Participants were escorted into a designated test room where they were directed to sit in a chair positioned at the center of the room. They were instructed to remain seated throughout the duration of the experiment to prevent motion sickness. Once seated, each participant was fitted with the HTC Vive Pro Eye headset and provided guidance on adjusting the earphones to achieve a comfortable sound level. Once fitted and wearing the VR headset, participants were given VR hand controllers and briefed on their general usage.

Subsequently, participants underwent a calibration procedure for the eye tracking component of the study. This calibration process involved adjusting the interpupillary distance for each participant, followed by a series of directed eye movements while the eye tracker tracked their gaze. This procedure allowed for automatic calibration of the eye tracking to the participants' movements. Participants held the VR hand controllers in their dominate hand with their thumb over the button throughout the duration of the experiment.

Once the experimental procedure commenced, participants were sequentially immersed in each scene for a fixed duration. Within each scene, participants had the freedom to explore their surroundings by turning their head to look around. For the study portion of the experiment, participants were instructed to: *Do your best to try to remember that scene along with what its name is. While viewing each of these scenes, a voice will play through the VR headphones telling the name of the scene. For example, while viewing a golf course, the voice would say "This is a golf course. Golf course." Try to also remember the name so that you can convey this later on if asked about earlier-viewed scenes.* After the study phase ended, participants were asked if they needed a break from the VR immersion.

For the test portion, before the scenes began to play, participants saw the instruction: *"If the scene starts to feel familiar to you, push the button under your THUMB to indicate that it feels familiar. Try to do this AS SOON as you start to feel a sense of familiarity with the scene. Specifically, if the scene reminds you of a specific scene that you viewed earlier. Let the experimenter know what that scene is that this scene is reminding you of. Sometimes, a scene may remind you of a similar-looking scene from earlier. Whenever this happens (even if you did not push the button) please tell the experimenter the name of the earlier-viewed scene. Even if the test scene did not remind you of a specific earlier-viewed scene."* When participants pressed the button to indicate familiarity, the experimenter was made aware through a message logged to the Unity terminal. Participants were then asked if they could identify possible reasons for the familiarity, and were continuously reminded that sometimes they may be able to identify a reason for any perceived familiarity with a scene and other times they may not. Most of the answers that participants gave regarding the source of their familiarity corresponded to earlier viewed scenes. However, some participants indicated some scenes reminded them of other locations, such as "a friend's basement." These answers were logged on paper by the experimenters and recorded on a microphone. Similarly to the two-dimensional familiarity task, participants completed two blocks of the study and test phases.

4.1.4 Eye Tracking and Feature Generation

The HTC Vive Pro Eye is a virtual reality headset with built-in infrared-based eye tracking technology developed by HTC Corporation. Prior research suggests that the HTC Vive Pro Eye validly measures eye movement metrics of interest to researchers [29]. The headset was used to collect eye tracking data within the Unity game engine from participants while they were immersed in the familiarity task environments. Eye tracking data was collected using the SRanipal software development kit (SDK) version 1.3.6.8 for Unity provided by the HTC Corporation [30]. Prior studies have demonstrated that timestamped eye tracking data from this device, collected with Unity and the SRanipal SDK, can accurately assess saccadic eye movements [31].

The SRanipal SDK facilitated the recording of the following eye measurements: pupil position, pupil diameter, eye openness, gaze origin, and gaze direction. Data was collected into a buffer at approximately 120 Hz in a dedicated thread using the SRanipal callback registration function, and this buffer was subsequently written to a file in the form of comma-separated values (CSVs) at the conclusion of each scene.

In addition to eye tracking data points, we recorded the current Unix timestamp from the computer running the program using the DateTime struct. While the SRanipal SDK does offer a timestamp data point, previous research has highlighted inaccuracies and errors in previous versions of the SDK [31]. Although potential bugs related to the timestamp may have been addressed in the latest version of the SRanipal SDK, we chose to utilize system time rather than of confirming the resolution of these issues. Furthermore, using the Unity ActionBasedController class, we moni-

Table 4.1: Descriptive statistics of eye gaze features (100 millisecond buffer and 3 second window.)

	Familiarity				Non-Familiarity			
Feature	**Mean**	**SD**	**Min**	**Max**	**Mean**	SD	**Min**	**Max**
Fixation Count	8.53	3.96	1.00	11.00	8.78	4.33	0.00	12.00
Fixation Duration (ms)	32.90	15.60	14.00	276.00	34.73	20.73	0.00	344.00
Saccade Count	4.39	2.10	0.00	5.00	4.50	2.29	0.00	6.00
Saccade Duration (ms)	47.60	30.87	0.00	227.00	42.09	33.11	0.00	356.00
Microsaccade Count	0.31	0.60	0.00	3.00	0.33	0.62	0.00	3.00
Microsaccade Duration (ms)	2.34	4.27	0.00	21.00	2.59	4.39	0.00	15.83
Blink Count	01.02	0.90	0.00	4.00	1.09	0.86	0.00	4.00
Blink Duration (ms)	38.35	56.85	0.00	355.00	48.99	73.02	0.00	592.00

tored whether participants were pressing the button on the HTC Vive Controller as instructed to indicate a sense of familiarity.

This data collection approach resulted in the generation of two CSV files per participant, one for each block. Each file contained nearly one hundred thousand timestamped rows of data. Each row contained the eye measurements described above, the status of the familiarity indication button (pressed or unpressed), and the current VR scene the participant was in at that point in time. We acquired a row of data every 8.33 milliseconds on average throughout the experiment.

Similarly to the methods described in section 3.1.4, PyTrack was used to extract parameters of interest [26]. Table 4.1 shows descriptive statistics for a portion of eye gaze features discussed in Section 4.4.1. All of eye gaze features generated from this experiment can be found listed in Figure 3.5.

4.2 Data Preprocessing

Participants reported 538 instances of familiarity. On average, the button indicating familiarity was pressed approximately 15.29 seconds into a scene (SD = 9.11). To extract features based on reported familiarity, we retrieved data from the moments preceding participants' button presses. We discarded a short buffer prior to the button press as has been commonly used in other model detection attempts [16, 21, 32, 33]. This approach ensures that the model prediction is based on the eye data patterns leading up to the report rather than physiological patterns from the act of reporting. However, the discarded buffer was shorter than in previous experiments, as participants were immersed in the VR environment while wearing the headset, reducing the need for actions such as looking away from the screen to confirm button presses. Further discussion regarding the buffer windows can be found in section 4.4.1.

To generate negative training instances, we randomly sampled test videos where participants did not report familiarity. To ensure a balanced dataset, we generated an equal number of negative instances per participant as we had reported positive familiarity instances. This approach resulted in a dataset containing 1,076 entries, with 538 positive familiarity instances and 538 negative familiarity instances. All of these instances ranged from 1 to 3 seconds in duration, depending on the window size being experimented with, as further discussed in section 4.4.1.

4.3 Training and Evaluation

Similarly to the two-dimensional study, we used Hyperopt for distributed hyperparameter optimization in our model search [27]. Among the classification algorithms we evaluated were AdaBoost (AB), Naive Bayes (NB), Logistic Regression (LR), Support Vector Classifier (SVC), Random Forrest (RF), and K-Nearest Neighbors (KNN). All of these models were trained using Leave-One-Participant-Out Cross-Validation. We guided our model search using the highest average Cohen's Kappa across all folds. More details on this validation protocol and the evaluation metric can be found in section 3.3.

4.4 Results

4.4.1 Buffer and Window Size Experiments

No Buffer

To encompass all potential instances of familiarity and to ascertain the extent to which eye-pattern indicators of subjective familiarity extend before the button-press, we conducted experiments sampling various window sizes. These windows range from 1 to 3 seconds pre-button-press. Additionally, to address concerns about the model potentially learning eye movement patterns associated with the VR hand-controller button-press itself, we conducted model searches both with and without using a buffer.

When we sampled 1 second of data immediately before the button press without using a buffer, the Random Forest models yielded an unexpectedly high Cohen's Kappa score of 0.73, as shown in Table 4.2. However, this result deviates significantly from prior research attempting to identify internal cognitive states using eye tracking [16, 19, 22–24, 34].

We hypothesised the observed high performance may be attributed to physiological patterns exhibited by participants in the moments immediately before the button press, stemming from the action of physically pressing the button. This could potentially confound the interpretation of the eye tracking data and lead to inflated model performance. Consequently, we concluded that a buffer period was necessary between the familiarity window and the button press to mitigate this confounding effect.

However, determining the exact duration of this buffer period poses a challenge. While cognitive processes are known to occur in fractions of a second [35], and the button-press itself is estimated to occur within 100 ms [36], the precise millisecond-level timing between experiencing subjective familiarity and pressing a handheld VR controller button remains uncertain. In order to err on the side of caution, a buffer period was integrated. This approach aimed at ensuring that any potential physiological patterns associated with the button-press were adequately separated from the eye movement data indicative of subjective familiarity, thus reducing the risk of confounding effects and ensuring the reliability of the analysis.

Buffer

Experimentation with various buffer and window sizes was carried out. It was observed that upon implementing a 100 millisecond buffer, model performance dropped significantly. Prior studies have established that the act of pressing a button occurs in only 100 ms [36, 37], thus providing a basis for this window size encapsulating the physiological patterns that may have emerged due

Table 4.2: No Buffer Model Search Results. Metrics were evaluated in a leave-one-participant-out paradigm (Standard Deviation in parenthesis)

Window	Model	Cohen's Kappa	F1 Score
1 sec	RF	0.73 (0.15)	0.85 (0.07)
2 sec	RF	0.18 (0.18)	0.58 (0.09)
3 sec	KNN	0.19 (0.18)	0.59 (0.09)

to the physical button press. Additionally, model performance with this buffer size aligned the most with prior work investigating the relationship between internal cognitive states and eye gaze patterns [19, 22–24, 34], in addition to the previous two-dimensional experiment.

Upon establishing the use of a buffer, experimentation with various window sizes was conducted. As other safeguard to exclude physiological patterns that may have emerged from the act of pressing the button, we proceeded to use a 3 second window along with the 100 millisecond buffer. This way, in-case we did happen to catch some of these patterns, they would only be a small portion of the positive instances. Additionally, the 3 second window most closely matched the window size in related work investigating the eye patterns exhibited in relation to mind wandering, another cognitive states [22].

As we increased the window size from 1 second to 3 seconds, we found that our best model results were usually in in the one 1 second window, regardless of the buffer size. There are a number of possible reasons for this finding, including: 1) eye gaze patterns related to the onset of familiarity may occur rapidly with respect to the subjective sensation, allowing participants to quickly push the button the moment they sense familiarity (within a tight time-window). 2) As we expanded the familiarity window, we may have begun sampling moments of eye gaze patterns where participants were not experiencing familiarity. 3) The closer the window is to the button-press, the more likely it is that eye gaze patterns reflect the participant's engagement in pressing the VR hand-controller button. The exact model evaluation metrics can be seen in Table 4.3.

Although cognitive processes tend to occur in fractions of a second [35] and research suggests that a button-press occurs within 100 ms [36,37], we cannot be certain at what point in the temporal

Table 4.3: Buffer-Window Model Search Results. Participants are evaluated in a leave-one-participant-out paradigm and the average (Standard Deviation in parenthesis)

Buffer	Window	Model	Cohen's Kappa	F1 Score
100 ms	1 sec	KNN	0.23 (0.15)	0.61 (0.07)
	2 sec	AB	0.19 (0.17)	0.59 (0.09)
	3 sec	RF	0.21 (0.16)	0.60 (0.08)
250 ms	1 sec	AB	0.17 (0.13)	0.59 (0.07)
	2 sec	LR	0.17 (0.22)	0.59 (0.11)
	3 sec	LR	0.14 (0.20)	0.57 (0.10)
500 ms	1 sec	KNN	0.14 (0.17)	0.57 (0.09)
	2 sec	KNN	0.18 (0.14)	0.59 (0.09)
	3 sec	AB	0.13 (0.15)	0.55 (0.08)
1000 ms	1 sec	RF	0.16 (0.15)	0.58 (0.08)
	2 sec	RF	0.17 (0.20)	0.59 (0.10)
	3 sec	LR	0.17 (0.22)	0.17 (0.11)

stream the eye gaze patterns reflect the sensation of familiarity versus the act of pushing the VR hand-controller button. Therefore, we continued our analysis using the 100 millisecond buffer along with 3 seconds of eye gaze data as a conservative approach, an illustration of this buffer-window size can be seen in Figure 4.1.

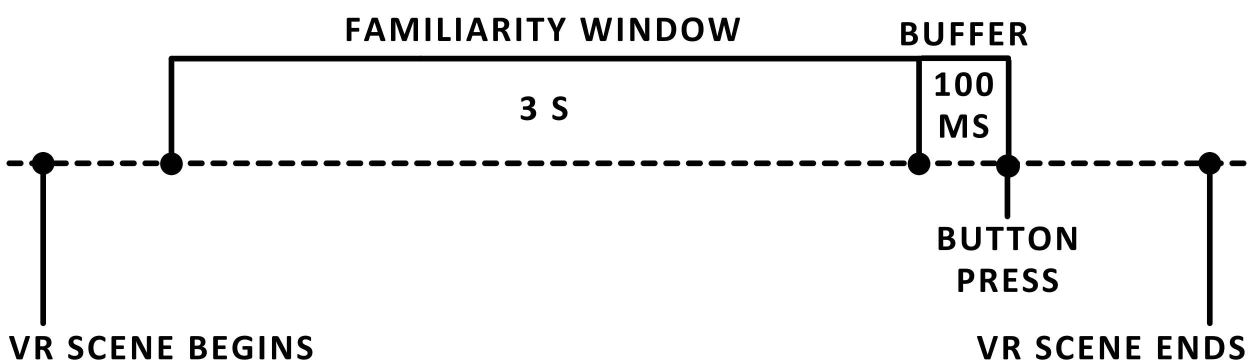

Figure 4.1: Familiarity instances timeline for VR condition.

Table 4.4: RF Model Results

Model	Buffer	Window	Cohen's Kappa	F1 Score	Accuracy
RF	100 ms	3 s	0.21 (0.16)	0.60 (0.08)	0.60 (0.07)

Table 4.5: VR Best Performing Features

Feature	Kappa	F1
pupil_AUC	0.13 (0.14)	0.56 (0.10)
avg_saccade_amplitude	0.09 (0.13)	0.53 (0.13)
avg_fix_cnt	0.08 (0.18)	0.52 (0.13)

Using the KNN *(n-neighbors = 95)* classification algorithm, our best model resulted in a kappa value of 0.21 (SD = 0.16). Additional evaluation metrics can be seen in Table 4.4. While this Cohen's Kappa value is slightly lower than the best model's value for the two-dimensional experiment, the standard deviation value is significantly lower. This indicates that overall, the model was not predicting with high confidence for certain participants and with extremely low confidence for others. Figure 4.2 shows the distribution of kappa values, which appears relatively normal.

Additionally, we were interested in exploring which features exhibited statistically significant differences between positive and negative familiarity instances. To achieve this, we conducted hypothesis tests and identified several features that showed statistically significant differences between positive and negative familiarity instances. These significantly different features included average blink duration, average saccade duration, and average pupil size. Figure 4.3 illustrates the differences in pupil size for positive and negative instances.

When training the KNN model using each feature individually, the features with the best performance were pupil area under the curve, average saccade amplitude, and average fixation duration. Unlike the two-dimensional experiment, all of the best performing features did not demonstrate a difference that was statistically significant.

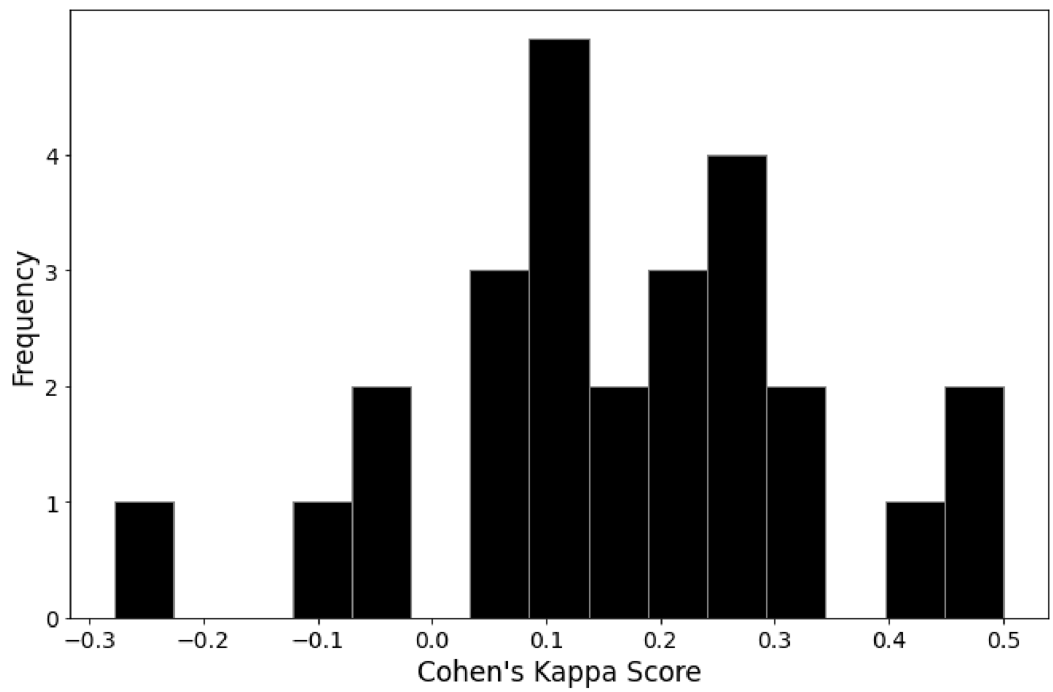

Figure 4.2: Distribution of Cohen's Kappa scores for VR condition.

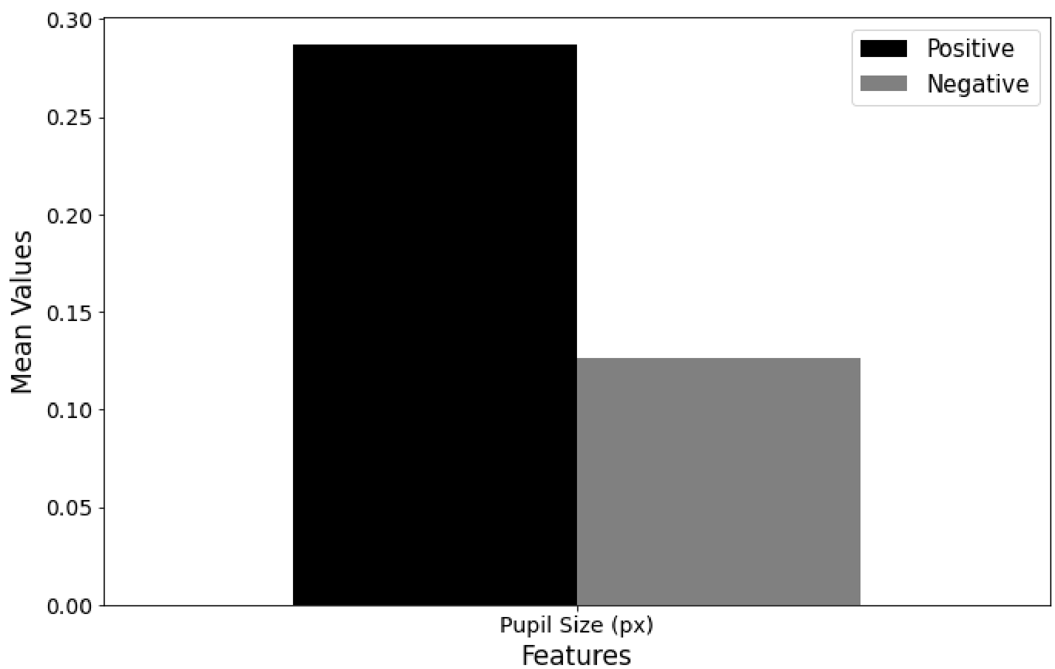

Figure 4.3: Significantly different pupil size for VR condition.

Chapter 5

Discussion

This work explores the feasibility of automatically identifying when a person is experiencing a sense of familiarity, an internal cognitive state associated with feelings of curiosity and information-seeking behaviors [7], using physiological eye gaze features. The findings from both our two-dimensional and more immersive three-dimensional virtual reality (VR) experiments suggest that the internal subjective state of familiarity does indeed manifest through eye movements. The capability to detect the sensation of familiarity through eye gaze patterns is analogous to identifying other internal cognitive states, such as mind wandering [16, 17, 23, 38].

5.1 Key Findings and Implications

The successful identification of familiarity using eye gaze patterns, as evidenced by the Cohen's Kappa values obtained in our models, suggests a tangible link between physiological responses and subjective cognitive states. This finding underscores the potential of utilizing eye tracking technology across varied environments, ranging from traditional 2D setups to more immersive 3D virtual reality (VR) scenarios. Of particular significance is the implication for developing intelligent systems, particularly in educational technology. The ability to detect when a person is experiencing familiarity opens avenues for designing adaptive learning environments. For example, intelligent tutoring systems could tailor content delivery based on learners' sense of familiarity-driven curiosity, potentially leading to more engaging and effective learning experiences [7].

One particularly interesting finding is the increased pupil size for familiarity instances in both the two-dimensional and VR experiment. While only the VR experiment showed a statistically significant difference, prior research has examined changes in pupil size shortly after the presentation of a stimulus, exploring its connection to later experiences of tip-of-the-tongue states. They observed significant differences in pupil size, with larger pupil diameter after stimulus presentation correlated with the sensation of almost recalling a word [39].

These findings not only contribute to the field of Human-Computer Interaction and Cognitive Science but also raise various opportunities for future work, including the extension of this research to real-world contexts.

5.2 Comparative Analysis of 2D and VR Environments

The differences in standard deviations for evaluation metrics between the 2D and more immersive 3D VR setups could be attributed to the distinct nature of interactions in these environments. The immersive nature of VR might lead to better eye tracking performance since users' eyes are always within tracking range, regardless of their orientation or head movements. This continuous tracking capability inherent to VR setups could contribute to reduced variability in eye gaze data compared to traditional 2D setups, where users' eye movements may be constrained by the screen boundaries.

However, when considering the application of this work to intelligent tutoring systems, the reliance on VR environments may pose challenges. Traditional online classroom settings typically do not involve immersive VR experiences. Therefore, implementing eye tracking technology in such settings may require alternative approaches to ensure accurate and reliable data collection. This highlights the need to tailor eye tracking methodologies and algorithms to suit the specific context of use.

5.3 Limitations and Future Work

Despite the promising findings, this study has several limitations, many of which underscore the need for further research to deepen understanding of the subjective sense of familiarity in relation to eye gaze. One notable limitation is the absence of an examination of pupil diameter at stimulus onset as it pertains to eventual downstream reporting of subjective familiarity. It is plausible that pupil size may precede the subjective sense of familiarity. For instance, Ryals et al. (2021) investigated pupil size within a short time window extending forward from stimulus onset in relation to eventual reporting of a tip-of-the-tongue state. They found significant differences in pupil size, with larger pupil diameter following stimulus onset being associated with the sensation of a word being on the tip of the tongue [39]. While our findings also matched these results with a significantly larger pupil size for positive familiarity instances, we cannot make a direct comparison since this work assessed pupil size within a short time window extending backward from the response button press. Thus, this is another avenue to be explored regarding the physiological responses linked to the subjective sense of familiarity.

Also, while consistent with other works automatically identifying internal states [16, 17], the high standard deviation in the 2D experiment model performance suggests significant variability in individual eye gaze patterns — this variability complicates efforts to integrate automated detection into AI [40]. Moreover, the constraint of data collection, where only instances with a three-second window prior to reporting familiarity were utilized, might have restricted the breadth of our analysis. Future research should endeavor to standardize the button press procedure across familiar and unfamiliar response options. For instance, researchers could require participants to press a designated button (e.g., right hand controller button for familiar scenes, left hand controller button for unfamiliar scenes) immediately upon identifying the familiarity or unfamiliarity of a scene. Alternatively, future studies could explore the possibility of eliminating the button press altogether, opting for a probe-based methodology instead. While this approach would preclude the assessment of familiarity experience in real-time, it would enable researchers to evaluate whether machine learning algorithms can discern differences between scenes that elicit a sense of familiarity and those that do not. By adopting such approaches, the scope of investigation could be broadened and potentially uncover new insights into the relationship between eye gaze patterns and subjective cognitive states.

Although the sample sizes in this study were determined based on prior behavioral research utilizing both 2D and more immersive 3D VR methodologies [8], there was no established precedent for computing the necessary sample size specifically for eye gaze data in these paradigms. As a result, the sample size for eye gaze data, particularly in the more immersive 3D VR experiment, may have been relatively small, potentially impacting the generalizability of the findings. To address this limitation, future research could explore methods for determining adequate sample sizes specifically for eye gaze data collection within virtual reality environments. Additionally, further investigation could involve the integration of other physiological measures, such as heart rate or skin conductance [16], to enhance the detection and understanding of cognitive states.

Finally, the largest question that looms from this work is the feasibility of discerning distinct types of internal states. While our study demonstrates the potential for identifying general internal states such as familiarity using eye gaze patterns, a more ambitious goal is to accurately classify specific internal states, such as curiosity, tip-of-the-tongue states, or mind wandering, in real-time. Ideally, an intelligent tutoring system would not only detect that a user is experiencing an internal state but also accurately identify the specific state occurring at any given moment. Future research should therefore aim to investigate whether machine learning algorithms can be trained to differentiate between various types of internal subjective states based on eye gaze patterns and other physiological signals. This endeavor would require robust experimental designs along with large and diverse datasets encompassing a wide range of internal states.

Chapter 6

Conclusion

In conclusion, this study underscores the promise of employing eye tracking technology to discern an individual's subjective sense of familiarity, a significant cognitive state. Despite the challenges highlighted, the results provide a solid foundation for advancing research and practical applications in the fields of affective computing and cognitive science.

By showcasing the feasibility of detecting familiarity through eye gaze patterns, this work opens up new avenues for exploring the relationship between physiological responses and subjective cognitive states. Moreover, it highlights the potential of integrating eye tracking technology into various domains, including education, virtual reality experiences.

Moving forward, addressing the limitations identified and further refining methodologies will be crucial for maximizing the utility and reliability of eye tracking technology in detecting cognitive states. Additionally, future research endeavors should aim to expand the scope of investigation to encompass a broader range of cognitive states and explore approaches for enhancing detection accuracy and applicability.

Ultimately, the insights gained from this study contribute to the ongoing dialogue surrounding affective computing and cognitive science, providing a basis for the development of more adaptive technologies.

Bibliography

[1] Anne M. Cleary, Alan S. Brown, Benjamin D. Sawyer, Jason S. Nomi, Adaeze C. Ajoku, and Anthony J. Ryals. Familiarity from the configuration of objects in 3-dimensional space and its relation to déjà vu: A virtual reality investigation. *Consciousness and Cognition*, 21(2), 2012.

[2] Logan K. Seabolt. Eye've seen this before: Building a gaze data analysis tool for déjà vu detection, 2022.

[3] American Psychological Association. *APA Dictionary of Psychology*. American Psychological Association, 2023.

[4] George Mandler. Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science*, 3(5), 2008.

[5] Anne M. Cleary. Recognition memory, familiarity, and déjà vu experiences. *Current Directions in Psychological Science*, 17(5), 2008.

[6] Anne M Cleary, Zachary C Irving, and Caitlin Mills. What flips attention? *Cognitive Science*, 47(4):e13274, 2023.

[7] Katherine L McNeely-White and Anne M Cleary. Piquing curiosity: Déjà vu-like states are associated with feelings of curiosity and information-seeking behaviors. *Journal of Intelligence*, 11(6):112, 2023.

[8] Noah S. Okada, Katherine L McNeely-White, Anne M Cleary, Brooke N Carlaw, Daniel L Drane, Thomas D Parsons, Timothy McMahan, Joseph Neisser, and Nigel P Pedersen. A virtual reality paradigm with dynamic scene stimuli for use in memory research. *Behavior Research Methods*, pages 1–24, 2023.

[9] George Nishimura and Aldo Faisal. Déjà vu: Classification of memory using eye movements. 2015.

[10] Anthony J Ryals, Jane X Wang, Kelly L Polnaszek, and Joel L Voss. Hippocampal contribution to implicit configuration memory expressed via eye movements during scene exploration. *Hippocampus*, 25(9):1028–1041, 2015.

[11] Christopher A Rowland. The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological bulletin*, 140(6):1432, 2014.

[12] Janet Metcalfe, Treva Kennedy-Pyers, and Matti Vuorre. Curiosity and the desire for agency: wait, wait… don't tell me! *Cognitive Research: Principles and Implications*, 6:1–8, 2021.

[13] Anne M. Cleary, Anthony J. Ryals, and Jason S. Nomi. Can déjà vu result from similarity to a prior experience? support for the similarity hypothesis of déjà vu. *Psychonomic Bulletin & Review*, 16, 2009.

[14] Alan S. Brown and Elizabeth J. Marsh. evoking false beliefs about autobiographical experience. *Psychonomic Bulletin & Review*, 15, 2008.

[15] Aldo Faisal George Nishimura. Déjà vu: Classification of memory using eye movements. 2015.

[16] Nathaniel Blanchard, Robert Bixler, Tera Joyce, and Sidney D'Mello. Automated physiological-based detection of mind wandering during learning. In *Intelligent Tutoring Systems: 12th International Conference, ITS 2014, Honolulu, HI, USA, June 5-9, 2014. Proceedings 12*, pages 55–60. Springer International Publishing, 2014.

[17] Robert Bixler, Nathaniel Blanchard, Luke Garrison, and Sidney D'Mello. Automatic detection of mind wandering during reading using gaze and physiology. In *ICMI '15: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 299–306. Association for Computing Machinery, 2015.

[18] Kalina Christoff, Alan M. Gordon, Jonathan Smallwood, Rachelle Smith, and Jonathan W. Schooler. Experience sampling during fmri reveals default network and executive system

contributions to mind wandering. *Proceedings of the National Academy of Sciences of the Uinited States of America*, 2009.

[19] Stephen Hutt, Kristina Krasich, Caitlin Mills, Nigel Bosch, Shelby White, James R. Brockmole, and Sidney K. D'Mello. Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction*, 29:821–867, 2019.

[20] Vishal Kuvar, Julia W. Y. Kam, Stephen Hutt, and Caitlin Mills. Detecting when the mind wanders off task in real-time: An overview and systematic review. *ICMI '23: Proceedings of the 25th International Conference on Multimodal Interaction*, pages 163–173, 2023.

[21] Myrthe Faber, Robert Bixler, and Sidney K. D'Mello. An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*, 50:134,150, 2017.

[22] Robert Bixler and Sidney D'Mello. Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, 26:33–68, 2015.

[23] Caitlin Mills, Robert Earl Bixler, Xinyi Wang, and Sidney K. D'Mello. Automatic gaze-based detection of mind wandering during film viewing. In *Educational Data Mining*, 2016.

[24] Stephen Hutt, Jessica Hardey, Robert Earl Bixler, Angela E. B. Stewart, Evan F. Risko, and Sidney K. D'Mello. Gaze-based detection of mind wandering during lecture viewing. In *Educational Data Mining*, 2017.

[25] Anne M. Cleary and Alexander B. Claxton. Déjà vu: An illusion of prediction. *Psychological Science*, 29(4):635–644, 2018. PMID: 29494276.

[26] Upamanyu Ghose, Arvind A. Srinivasan, W. Paul Boyce3, Hong Xu, and Eng Siong Chng1. Pytrack: An end-to-end analysis toolkit for eye tracking. *Behavior Research Methods*, 52(2588–2603), 2020.

[27] James Bergstra, Daniel Yamins, and David Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 115–123, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

[28] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[29] Immo Schuetz and Katja Fiehler. Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. *Journal of Eye Movement Research*, 15(3), 2022.

[30] Vive eye and facial tracking sdk 1.3.6.8.

[31] Yu Imaoka, Andri Flury, and Eling D. de Bruin. Assessing saccadic eye movements with head-mounted display virtual reality technology. *Frontiers in Psychiatry*, 11(572938), 2020.

[32] Vishal Kuvar, Nathaniel Blanchard, Alexander Colby, Laura Allen, and Caitlin Mills. Automatically detecting task-unrelated thoughts during conversations using keystroke analysis. *User Modeling and User-Adapted Interaction*, pages 617–641, 2023.

[33] Angela Stewart, Nigel Bosch, Huili Chen, Patrick J. Donnelly, and Sidney K. D'Mello. Where's your mind at? video-based mind wandering detection during film viewing. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, UMAP '16, page 295–296, New York, NY, USA, 2016. Association for Computing Machinery.

[34] Sidney D'Mello, Jonathan Cobian, and Matthew Hunter. Automatic gaze-based detection of mind wandering during reading. In *Educational Data Mining 2013*, 2013.

[35] F.C. Donders. On the speed of mental processes. *Acta Psychologica*, 30:412–431, 1969.

[36] Antti Oulasvirta, Sunjun Kim, and Byungjoo Lee. Neuromechanics of a button press. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery.

[37] Sunjun Kim, Jeongmin Son, Geehyuk Lee, Hwan Kim, and Woohun Lee. Tapboard: making a touch screen keyboard more touchable. pages 2833–2834, 04 2013.

[38] Stephen Hutt, Caitlin Mills, Shelby White, Patrick J Donnelly, and Sidney K D'Mello. The eyes have it: Gaze-based detection of mind wandering during learning with an intelligent tutoring system. *International Educational Data Mining Society*, 2016.

[39] Anthony J Ryals, Megan E Kelly, and Anne M Cleary. Increased pupil dilation during tip-of-the-tongue states. *Consciousness and Cognition*, 92:103152, 2021.

[40] Iliana Castillon, Videep Venkatesha, Nikhil Krishnaswamy, and Nathaniel Blanchard. Multimodal features for group dynamic-aware agents. In *Interdisciplinary Approaches to Getting AI Experts and Education Stakeholders Talking Workshop at AIEd.*, 2022.