國立臺灣大學理學院心理學系

碩士論文

Department of Psychology

College of Science

National Taiwan University

Master's Thesis

缺乏無意識處理連續閃爍抑制下影像記憶性的證據

Lack of Evidence for Unconscious Processing of Image Memorability Under Continuous Flash Suppression

徐永一

Yung-Yi Hsu

指導教授：謝伯讓 博士

Advisor: Po-Jang Hsieh Ph.D.

中華民國114年1月

January 2025

doi:10.6342/NTU202500760

摘要

先前的研究指出，刺激的記憶性（memorability）——即單次暴露後被記住的可能性——在個體之間具有一致性，且難以通過主動的認知努力加以控制。由於記憶性具有廣泛存在且無法自主控制的特性，本研究探討了記憶性是否可能是一種無意識的過程。在實驗一中，我們使用了突破連續閃爍抑制（breaking continuous flash suppression，簡稱 bCFS）範式，測量高記憶性與低記憶性人臉圖像的可見性閾值，以推斷它們在無意識處理上的差異。結果顯示，與統計上顯著的種族熟悉效果相比，這兩組之間沒有差異。在實驗二中，我們成功在大規模樣本重現了先前關於記憶性的研究結果，但在較小的子集樣本中未發現統計上的可靠性，這表明記憶性分數可能僅在較大規模的研究中具有可靠性。在實驗三中，我們檢驗了記憶性分數與 bCFS 反應時間之間的相關性，分析了大量圖像，但未發現顯著的關聯性。綜上所述，我們的研究未能支持記憶性差異在連續閃爍抑制下的無意識處理中出現的假設。這表明，記憶性在認知處理中的變化可能主要發生於意識層面。未來研究需進一步探討成功記憶編碼的關鍵因素，並深入闡明記憶性在感知處理中的作用。

關鍵詞：記憶性、無意識資訊處理、連續閃爍抑制

doi:10.6342/NTU202500760

Lack of Evidence for Unconscious Processing of Image Memorability Under Continuous Flash Suppression

Yung-Yi Hsu

Abstract

Prior studies indicate that the memorability of a stimulus, defined as the likelihood of it being remembered after a single exposure, remains consistent across individuals and is largely uncontrollable through top-down cognitive effort. Because of its pervasive and inherently involuntary nature, we investigated the potential for memorability to be an unconscious process in this study. In Experiment One, we used the breaking continuous flash suppression (bCFS) paradigm to measure the visibility thresholds of high- and low-memorability face images, aiming to infer differences in their unconscious processing. The results revealed no difference between the two groups compared with the statistically significant race-familiarity effect. In Experiment Two, we replicated previous findings on memorability at a large scale but found no statistical reliability in a smaller subset, suggesting that memorability scores may only be reliable at larger scales. In Experiment Three, we examined the correlation between memorability scores and bCFS reaction times across a large set of images but found no significant relationship. In summary, our findings provide no evidence to support the hypothesis that memorability differences emerge during unconscious processing under CFS. This suggests that variations in memorability during cognitive processing likely arise at the conscious level. Future research is needed to identify the key factors

doi:10.6342/NTU202500760

contributing to successful memory encoding and to further elucidate the role of memorability in perceptual processing.

Keywords: Memorability, Unconscious Information Processing, Continuous Flash Suppression

doi:10.6342/NTU202500760

Table of Contents

摘要.....i

Abstract.....ii

Table of Contents.....iv

List of Tables.....vi

List of Figures.....vii

1. Introduction.....1

2. Materials and Methods.....6

2.1 Participants.....6

2.2 Stimuli.....6

2.3 Procedure.....10

2.4 Apparatus.....15

2.5 Data Analysis.....16

3. Results.....22

3.1 Experiment One.....22

3.2 Experiment Two.....22

3.3 Experiment Three.....25

doi:10.6342/NTU202500760

4. Discussion 26

4.1 Experiment One 26

4.2 Experiment Two 29

4.3 Experiment Three 31

4.4 General Discussion 34

4.5 Conclusion 35

References 62

doi:10.6342/NTU202500760

List of Tables

Table 1 Analysis 1-1: Repeated Measures ANOVA Table for Experiment One (Excluding Participants Based on General Performance)............................36

Table 2 Analysis 1-2: Repeated Measures ANOVA Table for Experiment One (Excluding Participants Based on Performance Within Each Condition).....37

Table 3 Analysis 2-1: Repeated Measures ANOVA Table for d’ in Experiment Two (Group-Level Analysis) 38

Table 4 Analysis 2-1: Repeated Measures ANOVA Table for Hit Rates in Experiment Two (Group-Level Analysis)..39

Table 5 Analysis 2-1: Repeated Measures ANOVA Table for False Alarm Rates in Experiment Two (Group-Level Analysis) 40

Table 6 Analysis 2-2: Repeated Measures ANOVA Table for d’ in Experiment Two (Image-Level Analysis) 41

Table 7 Analysis 2-2: Repeated Measures ANOVA Table for Hit Rates in Experiment Two (Image-Level Analysis)..42

Table 8 Analysis 2-2: Repeated Measures ANOVA Table for False Alarm Rates in Experiment Two (Image-Level Analysis) 43

doi:10.6342/NTU202500760

List of Figures

Figure 1 Experiment One Procedure..44

Figure 2 Experiment Two Procedure...45

Figure 3 Experiment Three Procedure (bCFS Experiment) 46

Figure 4 Experiment Three Procedure (Memory Experiment)...............................47

Figure 5 Analysis 1-1: Results (Excluding participants according to average performance)..48

Figure 6 Analysis 1-2: Results (Excluding participants according to performance within condition)..49

Figure 7 Analysis 2-1: Results of d' Analysis (Group Level) 50

Figure 8 Analysis 2-1: Results of Hit Rates Analysis (Group Level).....................51

Figure 9 Analysis 2-1: Results of False Alarm Rates Analysis (Group Level)..........52

Figure 10 Analysis 2-2: Results of d' Analysis (Image Level) 53

Figure 11 Analysis 2-2: Results of Hit Rates and False Alarm Rates Analysis (Image Level) 54

Figure 12 Analysis 3-1: Results of Correlation Between Memorability Scores and bCFS Reaction Times (Within Subject) 55

doi:10.6342/NTU202500760

Figure 13 Analysis 3-2: Results of Correlation Between Memorability Scores and bCFS Reaction Times (Without Correction) 56

Figure 14 Analysis 3-2: Results of Correlation Between Memorability Scores and bCFS Reaction Times (with Correction) 57

Figure 15 Analysis 3-3: Results of Comparison bCFS Reaction Times Between Memorized and Forgotten Images 58

Figure 16 Bar Plots Showing the Hit Rates for Different Intervals in Vigilance Repeats for Each Participant. 59

doi:10.6342/NTU202500760

1. Introduction

Our daily lives are filled with a variety of visual inputs. Some are remembered for a long time, while others are quickly forgotten. Researchers use the concept of "memorability," defined as the probability of a novel stimulus being remembered after a single exposure, to quantify how memorable a stimulus is. Isola et al. (2011) were the first to investigate the consistency of memorability across individuals, revealing that people tend to remember the same stimuli despite individuals' diverse and unique experiences. Moreover, this inter-individual consistency has been robustly replicated across a wide range of stimuli, including faces (Bainbridge et al., 2013), words (Xie et al., 2020), objects (Dubey et al., 2015), visualizations (Borkin et al., 2013), and even videos (Cohendet et al., 2019). These findings suggest that memorability may not be purely subjective or personal but could also reflect intrinsic and measurable properties of a stimulus.

Given that the likelihood of remembering a stimulus primarily depends on its inherent properties, a critical question arises: what differences does memorability induce in brain processing? To address this, Bainbridge (2020) proposed two

doi:10.6342/NTU202500760

possibilities: one is that the brain passively prioritizes memorable stimuli, increasing the likelihood of successful encoding; the other is that memorable stimuli actively influence top-down processing, leading to distinct neural mechanisms and enhanced memory retention. Bainbridge (2020) conducted several experiments to explore this question, and the findings support the former. In one experiment, images varying in memorability were presented with followed cues instructing participants to either remember or forget each image. Both intrinsic memorability and cognitive manipulation (i.e., the cue) significantly influenced memory outcomes. However, the effect size of memorability was substantially larger than that of cognitive manipulation, suggesting that top-down cognitive efforts play a limited role in determining whether a stimulus will ultimately be remembered. Additionally, Bainbridge (2020) examined whether different levels of processing, as described by Bower et al. (1974), could override the effect of memorability. The findings revealed that the influence of processing depth was significantly weaker than that of memorability, indicating that the memorability effect is not simply due to an automatic allocation of attentional resources to memorable stimuli. In sum, Bainbridge (2020) suggests that the memorability effect is primarily driven by bottom-up processes rather than top-down mechanisms.

Beyond being resistant to top-down control, some studies suggest that low-level

doi:10.6342/NTU202500760

features play a crucial role in memorability processing. For example, Broers et al. (2018) found that memorable and forgettable images showed differences in memory performance even when the images were presented for only 13 milliseconds. At such a brief exposure duration, it is likely that only low-level features—such as color or shape—can be processed, while high-level information, such as semantic meaning, remains inaccessible (Maguire & Howe, 2016). Additionally, Han et al. (2023) used machine learning methods to demonstrate that shape-based contour features can predict memorability. This provides evidence that, beyond category or conceptual meaning—both of which have been strongly linked to memorability in past research (Kramer et al., 2023) —low-level visual features also contribute to the memorability of a stimulus. Taken together, these findings suggest that memorability information can be extracted from low-level features alone.

Taking all the findings together, these studies suggest that memorability may be an intrinsic property, largely independent of top-down control, and processable through the low-level features of an image. These characteristics of memorability lead to a hypothesis: memorability information can be processed unconsciously. This claim is based on the fact that other image properties with similar characteristics have been shown to undergo unconscious processing. For example, past research has

doi:10.6342/NTU202500760

demonstrated that people consistently recognize six basic emotions (Ekman, 1992). This suggests that emotion is an inherent property, largely independent of individual viewers—similar to the intrinsic nature of memorability. Additionally, emotions have been shown to be processed unconsciously (Faivre et al., 2012; Tamietto & De Gelder, 2010), supporting the idea that a property that mainly depends on image may not require conscious awareness for processing. Likewise, attractiveness has been found to be processed within extremely short exposure durations (Olson & Marshuetz, 2005) and can also be perceived unconsciously (Hung et al., 2016; Nakamura & Kawabata, 2018). In sum, given that other properties sharing key characteristics with memorability have been shown to undergo unconscious processing, we hypothesize that memorability can also be processed unconsciously.

In this study, we examined the possibility of unconscious processing of imaging memorability by assessing visibility threshold. Specifically, we tested whether stimuli with different levels of memorability will reach consciousness at varying thresholds when initially presented unconsciously. In Experiment One, we utilized the Breaking Continuous Flash Suppression (bCFS) paradigm (Stein et al., 2011) as it provides longer suppression times and stronger suppression effects compared to other masking techniques, such as crowding (Pelli et al., 2004), and masking (Ogmen et al., 2003),

doi:10.6342/NTU202500760

with the threshold measured using a staircase procedure. In Experiment Two, we tested whether the race familiarity effect and image transformations from color to greyscale might influence memorability scores, potentially confounding the results of Experiment One. In Experiment Three, we measured both memorability scores and bCFS thresholds for 100 scene images using the same group of participants and examined the correlation between the two scores. Across all three experiments, we found no significant differences in visibility thresholds based on the memorability levels of the stimuli. In Experiment one, high- and low-memorability images showed no differences in their bCFS thresholds. Experiment two successfully replicated memorability scores at the group level; however, statistical differences disappeared when analyzing only the subset of images used in Experiment one. Finally, Experiment three revealed no significant correlation between memorability scores and bCFS thresholds. Together, these findings provide no evidence to support the hypothesized unconscious processing of image memorability.

doi:10.6342/NTU202500760

2. Materials and Methods

2.1 Participants

All participants were recruited from public Facebook groups, primarily composed of students and alumni of National Taiwan University. All participants were naive to the purpose of the study. Written informed consent was obtained from all participants before the experiment, and they were reimbursed NT$50 for every 15 minutes of participation.

For Experiment One, 23 participants (10 males; age range: 18-25, mean: 21.96, SD: 1.90) were recruited, with two participants failing to complete the experiment. For Experiment Two, 38 participants were recruited (18 males; age range: 18-29, mean: 21.60, SD: 2.58). For Experiment Three, 25 participants were recruited (12 males; age range: 18-24, mean: 20.96, SD: 1.59). All participants who completed Experiment Three also participated in Experiment Two. All participants were native to Taiwan.

2.2 Stimuli

2.2.1 Experiment One

The 10k US Adult Faces Database (Bainbridge et al., 2013) was used in Experiment One. The memorable and forgettable groups were categorized directly

doi:10.6342/NTU202500760

based on the scores (hit rates in their experiment) provided by the database, without re-measuring the memorability scores on local subjects. This approach was chosen because memorability is considered a universal property, as claimed by the authors and other memorability studies (e.g., Borkin et al., 2013; Cohendet et al., 2019; Dubey et al., 2015; Xie et al., 2020). Additionally, Bainbridge (2020) demonstrated that memorability scores from this database remain consistent regardless of the experimental paradigm used. Jeong (2023) replicated the database results with Korean participants and concluded that memorability is consistent across cultures.

The criteria for the memorable and forgettable groups were based on scores in the top 25% and bottom 25%, respectively, following the methodology in Bainbridge (2020). Although Jeong (2023) affirmed cross-cultural consistency in memorability, race may still confound memory performance (Feingold, 1914) and bCFS thresholds (Stein, 2012). Therefore, to investigate the potential influence of race, both Asian and White faces were included in Experiment One.

Three of the most memorable (highest memorability scores) and three of the most forgettable Asian faces were selected, as only three Asian faces were available in the bottom 25% of the database. Corresponding White faces were selected based on two criteria: (1) the White face had the same or a nearly identical memorability score as

doi:10.6342/NTU202500760

each Asian face, and (2) if multiple White faces met the first criterion, the face with the closest false alarm rate was chosen.

In total, 12 images were selected as stimuli for Experiment One, with three images for each condition in the 2x2 design. These 12 images were cropped from the original images to remove the white background and replace it with a black background. They were then processed using MATLAB's SHINE_color toolbox (Dal Ben, 2023) to control brightness and spatial frequency. The processing was not perfect; the 12 images used in the experiment had a brightness range of 93.48-108.78 (mean: 102.66, SD: 5.09). After this, the images were converted to grayscale using Python by averaging the RGB values for each pixel.

2.2.2 Experiment Two

Experiment Two also used the 10k US Adult Faces Database. All 63 Asian faces from the database were used. Each Asian face was paired with a corresponding White face following the same rules as in Experiment One. All 126 images were converted to grayscale using the same method as in Experiment One.

2.2.3 Experiment Three

The MemCAT database (Goetschalckx & Wagemans, 2019) was used in Experiment Three. This database was chosen instead of the one used in prior

doi:10.6342/NTU202500760

experiments due to an imbalance in the number of images representing different races, particularly a lack of Asian faces sufficient to control for race familiarity and oddball effects. Additionally, the MemCAT database demonstrated a larger effect size on memorability in Jeong (2023), which conducted memory experiments using both databases on the same group of participants. Images were selected from the Scene category only. This decision was based on prior research (Dubey et al., 2015) suggesting that the memorability of an image is highly correlated with the most memorable object within it. To ensure that the memorability scores in our experiment reflected the overall image rather than being disproportionately influenced by a single object, only images from the Scene category were used, as they lack prominent main objects. Out of the 2,000 images in the Scene category, 100 images were randomly selected across memorability percentiles. Specifically, one image was randomly selected from the top 20 most memorable images, another from the 21st-40th most memorable images, and so on. This method was chosen over a completely random selection within the category to achieve a near-uniform distribution of memorability in the memory experiment. Additionally, 176 extra images were selected from the remaining 1,900 images in the Scene category for use in the memory experiment. The sizes of the images were kept the same as the original in the memory experiment, while the images used in the bCFS

doi:10.6342/NTU202500760

experiment were resized, with their longer side set to 300 pixels. This size is approximately (slightly smaller than) the maximum dimension of the dichoptic device's field on the screen.

2.3 Procedure

2.3.1 Experiment One

The visual field was split into two parts with the dichoptic device. As in Figure 1, each trial started with a fixation presented for 0.5-1 second. Then Mondrian images were presented on the dominant eye, which had been tested right before the experiment started, and face images were presented on the non-dominant eye for 1 second. The contrast of Mondrian images was set to 1, and the flashing rate was 10 Hz. The contrast of face images ramped up from 0 to its maximum within 0.5 seconds and remained at maximum contrast until the image had been presented for 1 second. Mondrian images continued flashing for an additional 0.3 seconds to prevent afterimages. Participants were instructed to press "1" on the number keypad whenever they saw anything that was not a Mondrian image.

The maximum contrast of each trial was determined using a 1-up-1-down staircase procedure (Dixon & Mood, 1948). The maximum contrast decreased whenever participants responded (i.e., in the next trial of that staircase, it would be more difficult

doi:10.6342/NTU202500760

to perceive the target if the participant saw it) and increased if participants did not respond. Each image had two staircases, with 30 trials per staircase. For the two staircases of each image, the initial contrast values were 0.1 and 0.5, respectively. The step size of the staircases was 0.1 initially and reduced to 0.04 after two reversals in that staircase. The floor and ceiling levels of contrast were 0.02 and 1, respectively.

There were 20 catch trials. In 10 trials, nothing was presented to the non-dominant eye, so participants should not press the key. In the other 10 trials, an extra face image (not one of the 12 target images) was presented to the non-dominant eye at full contrast, requiring participants to see it and press the key. All of trials were mixed. The 740 trials in total were divided into 4 blocks. Participants were allowed to take breaks between blocks and resume the experiment at their own pace.

2.3.2 Experiment Two

The paradigm used in Experiment Two followed the paradigm used in Jeong (2023) instead of the original paradigm in Bainbridge (2013). This adjustment was made because one of the purposes of this experiment was to measure memorability while avoiding the effect of an imbalanced number of faces between different races. However, the 10k US Adult Faces Database does not contain enough Asian-labeled faces to maintain the same number between two races if we use their paradigm.

doi:10.6342/NTU202500760

The selected 126 faces were divided into two sets using the following method: The faces were numbered according to their rank within their race (i.e., 1-63 for both races). Faces with odd numbers in one race were grouped with faces with even numbers in the other race. For example, the set containing odd-numbered Asian faces (Set 1) included Asian1, White2, Asian3, White4, and so on. As a reminder, faces with the same number were "corresponding", as describe in **Stimuli**. As a result, the distributions of memorability between the two sets were almost identical. This method was designed to keep the number and the distribution of memorability of the two races similar within and between the two sets.

The experiment contained three sessions: study, break interval, and test. In the study session (Figure 2, upper panel), one of the two sets was used, with the set counterbalanced between subjects. Each image was presented for 1 second, followed by a fixation presented for 1 second. The fixation color changed to red in 15 (out of 63) trials to maintain participants' attention. Participants were instructed to remember the faces in the study session and press the space bar when the fixation changed to red.

In the break interval session, the study session instructions were displayed on the screen for 30 seconds. Participants were not required to perform any task during this time.

doi:10.6342/NTU202500760

In the test session (Figure 2, lower panel), all 126 images from both sets were presented. Participants were instructed to answer whether the presented image had appeared in the study session by pressing "y" for yes and "n" for no. There was no reaction time limitation in the test session, and no feedback was provided.

2.3.3 Experiment Three

There are two sub-experiments in Experiment Three: the bCFS experiment and the memory experiment.

2.3.3.1 bCFS Experiment. In Experiment Three, threshold measurement was determined by averaging reaction times across trials instead of using the staircase procedure employed in Experiment One. This approach was chosen because the staircase procedure requires more trials, which is time-consuming. As in Experiment One, participants were instructed to press “1” on the number keypad whenever they saw anything that was not a Mondrian image.

As in Figure 3, each trial began with a fixation presented for 0.8-1.2 seconds. Following this, Mondrian images were displayed to the dominant eye, while target images were presented to the non-dominant eye. The contrast of the Mondrian images was set to 1, with a flashing rate of 10 Hz. The contrast of the target image ramped up over one second, following an exponential progression rather than a linear one. This

doi:10.6342/NTU202500760

adjustment, based on Alais et al. (2023) reflects the logistic relationship between physical contrast and visual sensitivity. Specifically, the target contrast increased from approximately 0.016 ($10^{-1.8}$) to 1 (10^{0}) by incrementally adding 0.03 to the power for each frame (with a screen refresh rate of 60 Hz). After reaching a contrast of 1, the target image remained at that level for the rest of its presentation. Each trial ended when participants responded or after 3 seconds. There were 10 trials for each images. Additionally, there were 8 catch trials in which no image was presented to the non-dominant eye. All of trials were mixed. The 1008 trials in total were divided into 6 blocks. Participants were allowed to take breaks between blocks and resume the experiment at their own pace.

Before the main experiment began, participants completed practice trials. These trials had the same settings as the main experiment, except that they lasted for 10 seconds instead of 3 seconds. The target image used in the practice trials was not included among the 100 images used in the main experiment. Practice trials repeated indefinitely until participants demonstrated familiarity with the task, defined as achieving reaction times under 3 seconds for at least three consecutive trials. Once this criterion was met, the experimenter informed participants that they seemed ready and could stop when they felt they had practiced enough.

doi:10.6342/NTU202500760

2.3.3.2 Memory Experiment. The paradigm and settings for the memory experiment were almost identical to those described in original study (Goetschalckx & Wagemans, 2019). Each block consisted of 200 images, with two blocks in total. During each trial, an image was presented for 600ms, followed by a fixation for 800ms (Figure 4). Within each block, there were 50 target repeat pairs, 12 vigilance repeat pairs, and 76 filler images. Target repeat pairs were composed of the 100 target images (50 per block), each appearing twice with intervals ranging from 19 to 149 intervening images. Vigilance repeat pairs were created using images randomly selected from the additional pool of 176 images. These vigilance images appeared twice with intervals ranging from 0 to 6 intervening images. Filler images were the remaining images from the pool, each appearing only once. The sequence of images and the lengths of the repeat intervals were randomized. Participants were instructed to press the spacebar whenever they recognized an image as repeated. They were informed that images would not repeat between blocks. No feedback was provided during the experiment.

2.4 Apparatus

2.4.1 Hardware

The stimuli were presented on a 27-inch IPS monitor (AORUS AD27QD) with a resolution of 1920×1080 pixels and a refresh rate of 60 Hz. A numeric keypad was used

doi:10.6342/NTU202500760

to collect responses in the bCFS experiments, while a standard keyboard was employed in the other experiment. A chin rest was utilized to maintain a fixed viewing distance of 58 cm.

2.4.2 Dichoptic-Viewing Device

A custom-built dichoptic viewing device was used to constrain participants' visual fields in the bCFS experiments. The interior surfaces were coated with light-absorbing materials to minimize internal reflections, and additional optical lenses were incorporated to facilitate binocular fusion.

2.4.3 Software

Stimuli presentation and data collection were implemented using Python (version 3.6.2). All functions were derived from the PsychoPy library(Peirce et al., 2019).

2.5 Data Analysis

All data in this study were analyzed using custom scripts implemented in R.

2.5.1 Experiment One

One participant was excluded from the analysis due to failing 8 catch trials (40%) in the experiment. The estimated threshold was defined as the last four reversal points for each staircase. All staircases from the remaining 20 participants had more than four reversal points. However, four staircases (out of 480 across all participants) had fewer

doi:10.6342/NTU202500760

than six reversal points, introducing larger bias in threshold estimation because the step size before the first two reversals was larger than that afterward.

2.5.1.1 Analysis 1-1. For the 20 participants, one was excluded because their average threshold across conditions was two standard deviations higher than the group average. A two-way repeated measures ANOVA (memorability: high/low; race: White/Asian) was conducted for Experiment One, with every threshold from each included participant used as a data point. Since the conditions (memorability and race) were constrained by the images, the error mean square in the ANOVA was calculated from the images rather than the participants. Pairwise comparisons were conducted for the significant interaction in the ANOVA.

2.5.1.2 Analysis 1-2. After completing Analysis 1-1 and visualizing the data points, we observed that few participants exhibited trends that might heavily influence the statistically significant results. We speculated that the significant results of Analysis 1-1 might be driven by these outliers. Consequently, participants were excluded if their thresholds in any condition were two standard deviations higher than the average for that condition. Four participants met this criterion, leaving data from 16 participants for reanalysis using the same ANOVA as in Analysis 1-1.

doi:10.6342/NTU202500760

2.5.2 Experiment Two

The overall d' was calculated for each participant as the cumulative density of the total hit rate (the number of "yes" responses to studied images divided by the total number of studied images) minus the cumulative density of the total false alarm rate (the number of "yes" responses to non-studied images divided by the total number of non-studied images) in the standard normal distribution. One participant was excluded from further analysis due to a negative d' (i.e., his/her hit rate was lower than their false alarm rate). The images were divided into three memorability conditions (high/medium/low) based on their original memorability scores, following the proportions used in Jeong (2023): 2:5:2 (14, 35, and 14 images per race, respectively).

2.5.2.1 Analysis 2-1. Three two-way repeated measures ANOVAs (memorability: high/medium/low; race: White/Asian) were conducted for three dependent variables: d', hit rate, and false alarm rate. These scores were analyzed because they provided distinct information. D' represented participants' overall performance, hit rate reflected memorability, and false alarm rate indicated familiarity. For each participant, the average score within each condition was calculated and used as a data point in the ANOVAs. Pairwise comparisons were conducted between each combination of the three conditions among memorability.

doi:10.6342/NTU202500760

2.5.2.2 Analysis 2-2 In addition to analyzing group-level data, the scores (d', hit rate, and false alarm rate) for the 12 images used in Experiment One were also analyzed. For these images, the average score across all participants was calculated and subjected to a two-way repeated measures ANOVA.

2.5.3 Experiment Three

2.5.3.1 Participant Exclusion. For the memory experiment, d' was calculated using the same formula as in Experiment Two. Hit rate was defined as the proportion of "yes" responses to repeated images (including both target and vigilance repeats), and false alarm rate was the proportion of "yes" responses to images shown for the first time (including fillers and first presentations of repeat pairs). Three participants were excluded: one with a negative d' and two whose d' values were more than two standard deviations below the group average.

For the bCFS experiment, trials without responses (i.e., not broken) were assigned a reaction time of 3 seconds (the response limit). Invalid trials, defined as those with reaction times below 0.3 seconds (estimated physical reaction time) or lower than three standard deviations from the participant's mean, were excluded. Participants with fewer than 95% valid trials (i.e., fewer than 950 valid trials) were excluded, following conventional criteria in statistical hypothesis testing. Four participants were excluded

doi:10.6342/NTU202500760

based on this criterion. No additional participants were excluded for failing catch trials or having outlier average reaction times. The excluded participants did not overlap between the two experiments, leaving data from 18 participants for analysis.

2.5.3.2 Analysis 3-1. Pearson correlations were conducted between the bCFS threshold and the memorability scores for 100 images. The bCFS threshold was calculated as the average reaction time across all valid trials for each image. Memorability scores were defined as the hit rate across all participants (where each participant contributed one hit or one miss per image).

2.5.3.3 Analysis 3-2. Upon visualizing the data, we observed that participants were highly conservative in providing positive responses, suggesting that the experimental paradigm may not have been suitable for our local participants. Consequently, additional Pearson correlations were conducted between the bCFS threshold and two types of memorability scores from the original database: uncorrected (hit rate) and corrected (hit rate minus false alarm rate).

2.5.3.4 Analysis 3-3. We also explored the relationship between memorability and visibility thresholds at the individual level. For each participant, images were categorized as either "memorized" or "forgotten" based on whether the participant responded to them during the memory experiment. The average bCFS reaction time for

doi:10.6342/NTU202500760

images in each category was calculated for every participant, and paired t-tests were conducted to compare reaction times between memorized and forgotten images.

doi:10.6342/NTU202500760

3. Results

3.1 Experiment One

3.1.1 Analysis 1-1

The two-way ANOVA revealed that the main effect of memorability was not significant, $F(1,18) = 1.65$, $p = .21$, $\eta_p^2 = .083$. The main effect of race was significant, $F(1,18) = 9.80$, $p = .0058$, $\eta_p^2 = .35$, indicating that Asian faces had a lower threshold than White faces. However, the interaction effect was not significant, $F(1,18) = 2.66$, $p = .12$, $\eta_p^2 = .13$ (Table 1 and Figure 5).

3.1.2 Analysis1-2

The two-way ANOVA showed that the main effect of memorability was not significant, $F(1,15) = .025$, $p = .88$, $\eta_p^2 = .0017$. The main effect of race was significant, $F(1,15) = 8.06$, $p = .012$, $\eta_p^2 = .35$, with Asian faces exhibiting a lower threshold than White faces. However, the interaction effect was not significant, $F(1,15) = .11$, $p = .74$, $\eta_p^2 = .0074$ (Table 2 and Figure 6).

3.2 Experiment Two

3.2.1 Analysis 2-1

For the ANOVA of d', the main effect of memorability was significant, $F(2,72) =$

doi:10.6342/NTU202500760

30.88, $p < .0001$, $\eta_p^2 = .46$. The main effect of race was also significant, $F(1,36) = 4.44$, $p = .042$, $\eta_p^2 = .11$, with Asian faces having higher d' than White faces. The interaction was not significant, $F(2,72) = .26$, $p = .77$, $\eta_p^2 = .0072$ (Table 3 and Figure 7). Pairwise comparisons revealed that high memorability faces had higher d' than low memorability faces, $t(72) = 7.27$, $p < .0001$, Cohen's d = .97, and medium memorability faces, $t(72) = 6.22$, $p < .0001$, Cohen's d = .96. No significant difference was observed between medium- and low-memorability faces, $t(72) = 1.05$, $p = .55$, Cohen's d = .18.

For the ANOVA of hit rate, the main effect of memorability was significant, $F(2,72) = 5.16$, $p = .0080$, $\eta_p^2 = .13$, while the main effect of race was not, $F(1,36) = 3.26$, $p = .079$, $\eta_p^2 = .083$. The interaction was also not significant, $F(2,72) = 2.12$, $p = .13$, $\eta_p^2 = .056$ (Table 4 and Figure 8). Pairwise comparisons indicated that high-memorability faces had a higher hit rate than low-memorability faces, $t(72) = 3.14$, $p = .0068$, Cohen's d = .36. No significant differences were found between high- and medium-memorability faces, $t(72) = .98$, $p = .59$, Cohen's d = .13., or between medium and low-memorability faces, $t(72) = 2.16$, $p = .086$, Cohen's d = .28.

For the ANOVA of false alarm rate, the main effect of memorability was significant, $F(2,72) = 20.60$, $p < .0001$, $\eta_p^2 = .36$. The main effect of race, $F(1,36) = 1.23$, $p = .27$, $\eta_p^2 = .033$, and the interaction, $F(2,72) = 1.79$, $p = .18$, $\eta_p^2 = .047$, were

doi:10.6342/NTU202500760

not significant (Table 5 and Figure 9). Pairwise comparisons showed that high-memorability faces had a lower false alarm rate than both low-memorability faces, $t(72) = -6.09$, $p < .0001$, Cohen's d = .88, and medium-memorability faces, $t(72) = -4.80$, $p < .0001$, Cohen's d = .89. No significant difference was observed between medium- and low-memorability faces, $t(72) = 1.30$, $p = .40$, Cohen's d = .18.

3.2.2 Analysis 2-2

For the ANOVA of d', the main effect of memorability was significant, $F(1,8) = 12.95$, $p = .0070$, $\eta_p^2 = .62$, with the high-memorability group showing higher d' than low-memorability group. The main effect of race, $F(1,8) = .18$, $p = .68$, $\eta_p^2 = .022$, and the interaction, $F(1,8) = .017$, $p = .90$, $\eta_p^2 = .0021$, were not significant (Table 6 and Figure 10).

For the ANOVA of hit rate, the main effect of memorability, $F(1,8) = 4.93$, $p = .057$, $\eta_p^2 = .38$, the main effect of race, $F(1,8) = 3.58$, $p = .095$, $\eta_p^2 = .31$, and the interaction, $F(1,8) = .002$, $p = .96$, $\eta_p^2 = .00033$, were not significant (Table 7 and Figure 11, left bars).

For the ANOVA of false alarm rate, the main effect of memorability was significant, $F(1,8) = 8.76$, $p = .018$, $\eta_p^2 = .52$, with the high-memorability group exhibiting a lower false alarm rate than the low-memorability group. Neither the main

doi:10.6342/NTU202500760

effect of race, $F(1,8) = .23$, $p = .65$, $\eta_p^2 = .028$, nor the interaction, $F(1,8) = .066$, $p = .80$, $\eta_p^2 = .0082$, were significant For the ANOVA of hit rate, the main effect of memorability, $F(1,8) = 4.93$, $p = .057$, $\eta_p^2 = .38$, the main effect of race, $F(1,8) = 3.58$, $p = .095$, $\eta_p^2 = .31$, and the interaction, $F(1,8) = .002$, $p = .96$, $\eta_p^2 = .00033$, were not significant (Table 8 and Figure 11, right bars).

3.3 Experiment Three

3.3.1 Analysis 3-1

The Pearson correlation analysis indicated no significant relationship between bCFS threshold and memorability, $r = -.070$, $t(98) = -.70$, $p = .49$ (Figure 12).

3.3.2 Analysis 3-2

The Pearson correlation between uncorrected memorability and bCFS threshold was not significant, $r = .18$, $t(98) = 1.76$, $p = .081$ (Figure 13). Similarly, the correlation between corrected memorability and bCFS threshold was not significant, $r = .16$, $t(98) = 1.62$, $p = .11$ (Figure 14).

3.3.3 Analysis 3-3

The t-test found no significant difference in bCFS thresholds between memorized and forgotten images, $t(17) = -.61$, $p = .55$, Cohen’s d = .045 (Figure 15).

doi:10.6342/NTU202500760

4. Discussion

4.1 Experiment One

The results of the two analyses were consistent, indicating that the findings in Analysis 1-1 were not driven solely by few outliers. Both revealed no main effect of memorability on visibility thresholds, suggesting that differences in image memorability did not significantly influence them. The main effect of race showed that Asian faces exhibited lower thresholds than White faces, aligning with the race familiarity effect observed in previous study (Stein, 2012). The significant main effect of race familiarity provides positive evidence that our experimental setup did not have critical flaws that would have led to a failure in detecting the effect of memorability.

Several limitations in our experiment may have confounded the results. First, differences in visibility thresholds between faces of varying memorability might exist but were not detected due to specific methodological constraints. The selection of faces was based solely on controlling memorability across racial groups and maximizing the effect size of memorability. Consequently, other variables, such as gender and viewpoint, were not controlled between groups, potentially introducing noise and diminishing the observed effect of memorability on visibility thresholds. Additionally,

doi:10.6342/NTU202500760

irrelevant elements, such as hair or background, may have further contaminated the results.

Second, while memorability is generally considered an intrinsic property of visual stimuli—a claim supported by prior studies (e.g. Bainbridge et al., 2013; Borkin et al., 2013; Cohendet et al., 2019; Dubey et al., 2015; Xie et al., 2020)—the memorability scores used in this study may not fully reflect true memorability. Several factors could contribute to this discrepancy. Image preprocessing, including controlling low-level features and converting images to greyscale, may have altered not only the physical but perceptual properties of the stimuli. For instance, brown and blue eyes with similar luminance might contribute differently to a face's memorability, but this distinction could be lost in greyscale transformation. Furthermore, the original database used to derive memorability scores was flawed. Of the 2,222 faces measured for memorability and had race labels, there was a significant imbalance between White (1,836) and Asian (36) faces. The remaining 7,946 filler images were unlabeled but appeared predominantly White, potentially creating a contextual "odd-ball" effect that artificially inflated the memorability scores of Asian faces due to their minority status.

Cultural differences may have also influenced the results. Although Jeong (2023) reported cross-cultural consistency in memorability, this does not preclude cross-

doi:10.6342/NTU202500760

cultural differences between the White participants in the original database and our local participants in Taiwan. Cultural interactions could lead to two opposing effects: the "other-race effect" (Feingold, 1914), which suggests superior recognition of own-race faces, and a racial "odd-ball" effect, where faces of rarely seen races are perceived as novel and thus more memorable.

Additionally, our visibility measurement method may have been suboptimal. The raw data revealed that some participants consistently detected suppressed faces even when the contrast reached floor levels, which likely reduced the sensitivity of our measure and obscured potential differences between conditions. Furthermore, the results may have been influenced by insufficient statistical power. Jeong(2023) reported that the effect size of memorability differences among face images from Bainbridge et al. (2013) was smaller than that for scene images from Goetschalckx and Wagemans (2019). Specifically, Jeong found effect sizes of $\eta_p^2 = .15$ for the former and $\eta_p^2 = .61$ for the latter, indicating that our null results might be attributable to a lack of statistical power.

To address these limitations and explore the influence of preprocessing and contextual effects on memorability measurements, we conducted Experiment Two. This experiment aimed to replicate the memorability results reported by Bainbridge et al.

doi:10.6342/NTU202500760

(2013) using an equal number of White and Asian face images, including those used in Experiment One, all transformed to greyscale.

4.2 Experiment Two

For Analysis 2-1 (group-level analysis), we successfully replicated the difference in hit rates, with an effect size (η_p^2 = .13) similar to that reported by Jeong (2023) (η_p^2 = .15). The results of the post-hoc comparisons were also consistent; both studies only found significant differences between high- and low-memorability groups (our Cohen's d = .36; Jeong (2023)'s Cohen's d = .47). The analyses of d' and false alarm rates yielded consistent results, indicating that participants' performance differences across memorability levels reflected their ability to better memorize memorable images rather than higher familiarity or a general tendency to respond "yes" during the experiment. These findings suggest that cultural effects, contextual effects, and color information may not significantly influence memorability, aligning with Bainbridge (2017), which proposed that the memorability of face images is identity-dependent rather than image-dependent.

The results of the main effect of race were inconclusive across the ANOVAs for the three types of scores. While the effect on d' was significant, it was not significant for hit rates or false alarm rates. According to Cohen's (1992) criteria, d' (η_p^2 = .11) and

doi:10.6342/NTU202500760

hit rate (η_p^2 = .086) had medium effect sizes (η_p^2 > .06), whereas false alarm rate (η_p^2 = .033) had a small effect size (η_p^2 > .01). These results suggest that the other-race effect may be weak or offset by other factors (e.g., the loss of color information). Another possibility is that the faces used were not genuinely representative of Asian faces, thus failing to elicit a same-race advantage. In Bainbridge et al. (2013), race was defined based on participant ratings on a 7-point categorical scale, with the mode of responses used to determine race, rather than the actual identity of the faces. In summary, our participants exhibited minimal or no differences in performance across races in Experiment Two.

For Analysis 2-2 (image-level analysis), the main effect of memorability across the ANOVAs for the three scores was inconsistent. Memorability had a significant main effect on d’ and false alarm rates but not on hit rates, which directly represent memorability. According to Cohen’s (1992) criteria, the effect sizes for d’ (η_p^2 = .62), hit rate (η_p^2 = .38), and false alarm rate (η_p^2 = .52) all met the threshold for large effect sizes (η_p^2 > .14). The lack of a significant result for hit rates despite a large effect size may be due to a limited sample size (number of images) or high within-group variance. This suggests that the null results for memorability in Experiment One may have been due to a small difference in memorability between the high- and low-memorability

doi:10.6342/NTU202500760

groups.

The results for the main effect of race were consistent, showing no significant differences in d', hit rates, or false alarm rates. This indicates that the race-familiarity effect observed in Experiment One was likely independent of memorability.

The findings of Experiment Two suggest that memorability may remain consistent after color transformation and across different contexts. However, they do not provide conclusive evidence that the images used in Experiment One accurately represented high- and low-memorability images. These results imply that while the memorability scores provided by the database may be reliable at the group level, they may not be as reliable when only a few images are selected. This could be due to selection bias or the inherent randomness of the measurement process. Consequently, we decided to examine the relationship between memorability and visibility thresholds on a larger scale (100 images) and to measure the memorability of these images using the same group of participants to minimize the potential influence of individual differences.

4.3 Experiment Three

In our memory experiment, participants were much more conservative in their responses compared to those in the database (average hit rate: .076; memCat: .76). Looking at d', our participants also performed worse (average d': .66; memCat: 2.50).

doi:10.6342/NTU202500760

Furthermore, as shown in Figure 16, this tendency also rendered the vigilance repeat ineffective in determining whether participants were paying attention to the task. Counterintuitively, most participants did not achieve 100% accuracy even when the interval was 0 (i.e., when the repetition occurred immediately). These results suggest that the repeat-detection paradigm may not be an appropriate method for measuring memorability in our local participant sample. One possible reason for their conservative responses is that participants may have assumed that repeats occurred only a few times. A potential solution could involve modifying the instructions. In our current instructions, participants were simply told to press a button whenever an image was repeated, following the description in Goetschalckx and Wagemans (2019), without being informed about the number of repeats in each block. To address this, we could encourage participants to respond more confidently when uncertain by explicitly stating that multiple repeats occur and that intervals between repeats can vary.

In the bCFS experiment, we acknowledge several design flaws. First, each trial ended immediately when participants gave a response. This design might have inadvertently encouraged participants to rush through the experiment to complete it faster and receive compensation, potentially compromising the data quality. Additionally, the response time was limited to three seconds, which may not have been

doi:10.6342/NTU202500760

sufficient. Some participants were unable to detect the target within this time frame. Although we considered these issues, we opted to retain the current design to manage the experiment's duration. Based on past experience, participants tend to feel fatigued and bored after 30–40 minutes, which could render the latter part of the data unreliable. Extending the trial duration or removing the time limit would have risked overburdening participants and compromising data quality further.

Second, the low-level features of the images were not controlled. This design choice aimed to preserve the potential influence of features such as color, luminance, or shape, which could act as common factors driving the correlation between memorability and visibility thresholds. However, this approach—particularly the lack of shape control—introduced its own challenges. For instance, an image with an elongated or slender shape might be memorable due to its distinct proportions. When resized to a fixed long side of 300 pixels, such an image might appear smaller, reducing its visibility strength. As a result, a highly memorable image could exhibit a high visibility threshold, contrary to our prediction.

The results of Analysis 3-1 provided no evidence of a correlation between memorability and visibility thresholds. As discussed earlier, one possible explanation for these null results is the inadequacy of our memorability measurement. To further

doi:10.6342/NTU202500760

investigate, we conducted two Pearson correlation analyses (Analysis 3-2) between bCFS thresholds and the memorability scores from the database. The results were not statistically significant, although a weak trend was observed (r = .16 for uncorrected memorability and r = .17 for corrected memorability). However, this trend was in the opposite direction of our prediction. We had hypothesized that memorable images would be easier to perceive, leading to a negative correlation between memorability scores and bCFS reaction times.

In Analysis 3-3, we found no significant difference in bCFS reaction times between memorized and forgotten images. However, we acknowledge a strong imbalance in the number of images between the two groups. Consequently, the mean reaction times were averaged from differing numbers of trials, which may have increased the risk of violating the t-test assumption of equal variance. Overall, the results of Experiment Three suggest that the relationship between memorability and visibility thresholds cannot be detected, even with a large number of images.

4.4 General Discussion

In our experiments, we failed to detect an effect of memorability on visibility thresholds. Here, we propose a potential explanation: the effect may exist, but we were unable to detect it due to limitations in how memorability is typically measured. The

doi:10.6342/NTU202500760

current mainstream approach to measuring memorability may not fully represent the concept of memorability as it is commonly understood. In everyday life, when we describe a visual stimulus as "memorable" or "forgettable," we typically refer to its retention in long-term memory rather than in working memory. However, most memorability studies (e.g., Bainbridge et al., 2013; Borkin et al., 2013; Cohendet et al., 2019; Dubey et al., 2015; Xie et al., 2020) have assessed memorability using tasks that rely on working memory rather than long-term memory. Notably, Cohendet et al. (2019) found only a moderate correlation between short-term and long-term memorability, suggesting that they may not be equivalent constructs. It is possible that the relationship between memorability and visibility thresholds would emerge if memorability were assessed based on long-term memory rather than working memory.

4.5 Conclusion

In conclusion, our findings provide no evidence that memorability is correlated with visibility thresholds. This suggests that differences in memorability during cognitive processing may primarily occur at the conscious level.

doi:10.6342/NTU202500760

Tables

Table 1

Analysis 1-1: Repeated Measures ANOVA Table for Experiment One (Excluding Participants Based on General Performance)

Item	Sum.Sq	Mean.Sq	NumDF	*F*	*p*
Memorability	.0058	.0058	1	1.65	.21
Residual	.064	.0059	18		
Race	.058	.058	1	9.80	.0058
Residual	.11	.0059	18		
Interaction	.0098	.0098	1	2.66	.12
Residual	.066	.0037	18		

Note. Including 19 participants.

doi:10.6342/NTU202500760

Table 2

Analysis 1-2: Repeated Measures ANOVA Table for Experiment One (Excluding Participants Based on Performance Within Each Condition)

Item	Sum.Sq	Mean.Sq	NumDF	*F*	*p*
Memorability	1.7×10^{-5}	1.7×10^{-5}	1	.025	.88
Residual	.010	.00070	15		
Race	.013	.013	1	8.06	.012
Residual	.024	.0016	15		
Interaction	.00020	.00020	1	.11	.74
Residual	.027	.0018	15		

Note. Including 16 participants.

doi:10.6342/NTU202500760

Table 3

Analysis 2-1: Repeated Measures ANOVA Table for d' in Experiment Two (Group-Level Analysis)

Item	Sum.Sq	Mean.Sq	NumDF	*F*	*p*
Memorability	44.91	22.46	2	30.88	< .0001
Residual	52.35	.73	72		
Race	2.37	2.37	1	4.44	.04
Residual	19.23	.53	36		
Interaction	.34	.17	2	.26	.77
Residual	47.14	.65	72		

doi:10.6342/NTU202500760

Table 4

Analysis 2-1: Repeated Measures ANOVA Table for Hit Rates in Experiment Two (Group-Level Analysis)

Item	Sum.Sq	Mean.Sq	NumDF	*F*	*p*
Memorability	.31	.15	2	5.16	.0080
Residual	2.13	.03	72		
Race	.063	.063	1	3.26	.079
Residual	.69	.020	36		
Interaction	.10	.051	2	2.12	.13
Residual	1.72	.024	72		

doi:10.6342/NTU202500760

Table 5

Analysis 2-1: Repeated Measures ANOVA Table for False Alarm Rates in Experiment Two (Group-Level Analysis)

Item	Sum.Sq	Mean.Sq	NumDF	*F*	*p*
Memorability	1.21	.61	2	20.60	< .0001
Residual	2.12	.0029	72		
Race	.021	.021	1	1.23	.27
Residual	.62	.017	36		
Interaction	.054	.027	2	1.79	.18
Residual	1.09	.015	72		

doi:10.6342/NTU202500760

Table 6

Analysis 2-2: Repeated Measures ANOVA Table for d' in Experiment Two (Image-Level Analysis)

Item	Sum.Sq	Mean.Sq	NumDF	*F*	*p*
Memorability	7.55	7.55	1	12.95	.0070
Race	.11	.11	1	.18	.68
Interaction	.010	.010	1	.017	.90
Residual	4.66	.58	8		

doi:10.6342/NTU202500760

Table 7

Analysis 2-2: Repeated Measures ANOVA Table for Hit Rates in Experiment Two (Image-Level Analysis)

Item	Sum.Sq	Mean.Sq	NumDF	*F*	*p*
Memorability	.094	.094	1	4.93	.057
Race	.068	.068	1	3.58	.095
Interaction	< .0001	< .0001	1	.0020	.96
Residual	.15	.019	8		

doi:10.6342/NTU202500760

Table 8

Analysis 2-2: Repeated Measures ANOVA Table for False Alarm Rates in Experiment Two (Image-Level Analysis)

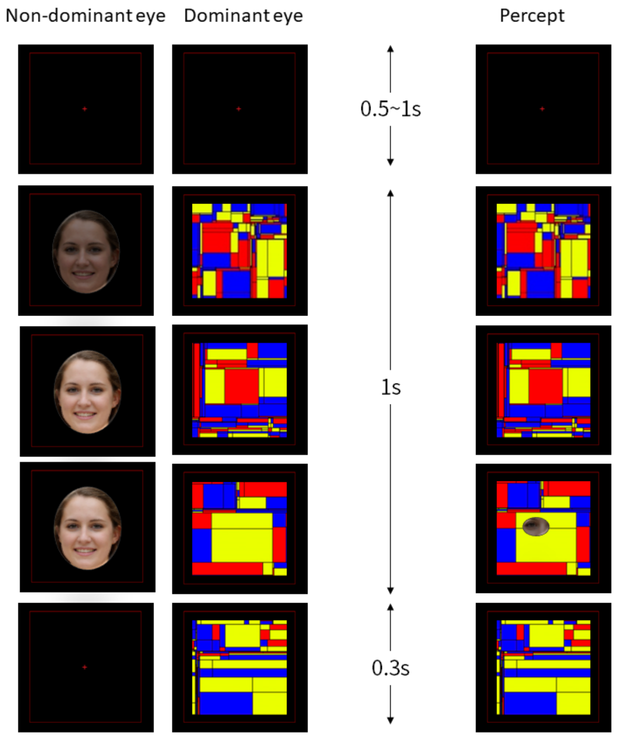
Item	Sum.Sq	Mean.Sq	NumDF	*F*	*p*
Memorability	.12	.12	1	8.76	.018
Race	.0030	.0030	1	.23	.65
Interaction	.00090	.00090	1	.066	.80
Residual	.11	.013	8		

doi:10.6342/NTU202500760

Figures

Figure 1

Experiment One Procedure

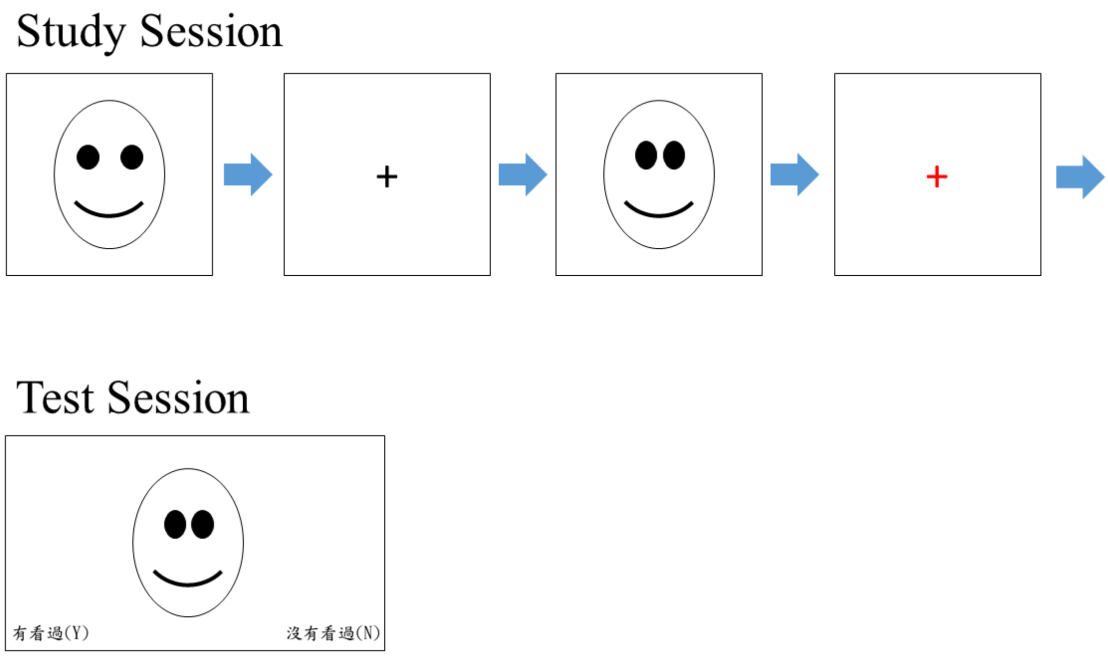


Note. Each trial started with a fixation presented for 0.5-1 second. Then Mondrian images were presented on the dominant eye and face images were presented on the non-dominant eye for 1 second. The contrast of face images ramped up from 0 to its maximum within 0.5 seconds and remained at maximum. Mondrian images continued flashing for an additional 0.3 seconds. Participants responded whenever the CFS broke. Face image in this figure were generated using a random face generator (https://thispersondoesnotexist.com/)

doi:10.6342/NTU202500760

Figure 2

Experiment Two Procedure

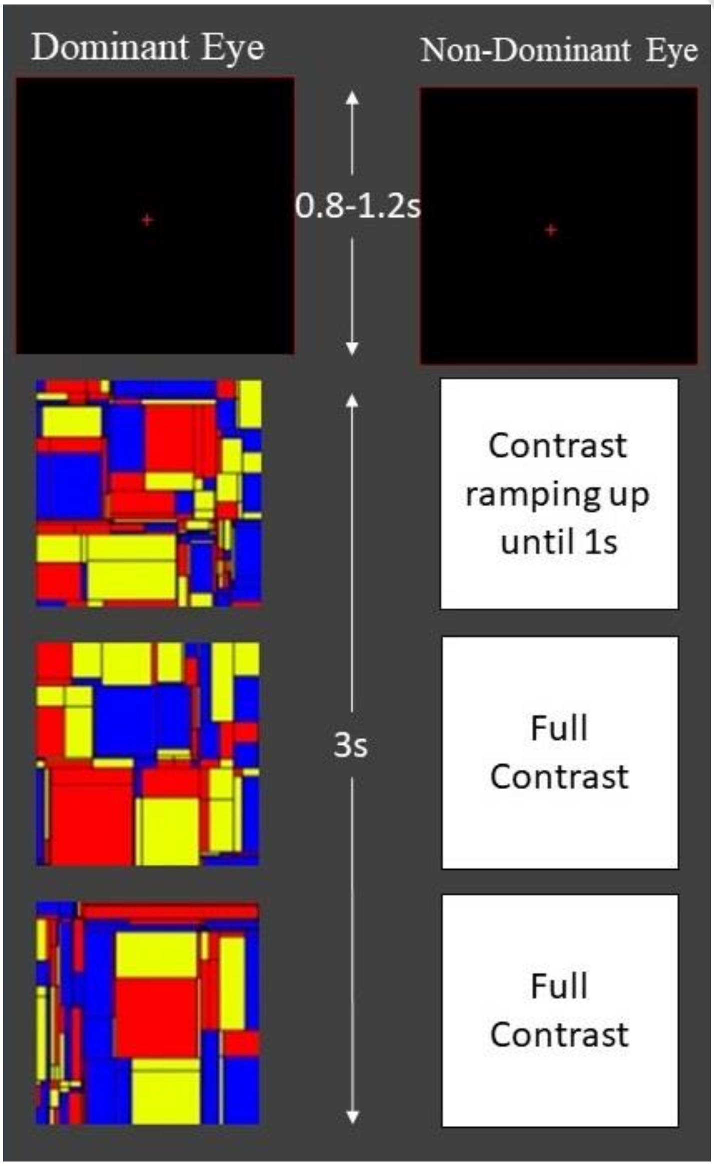


Note. During the study session, participants were instructed to memorize the presented faces and respond whenever the fixation turned red. In the test session, participants indicated whether the presented faces had been shown previously. Instructions for the designated key were displayed on the screen.

doi:10.6342/NTU202500760

Figure 3

Experiment Three Procedure (bCFS Experiment)

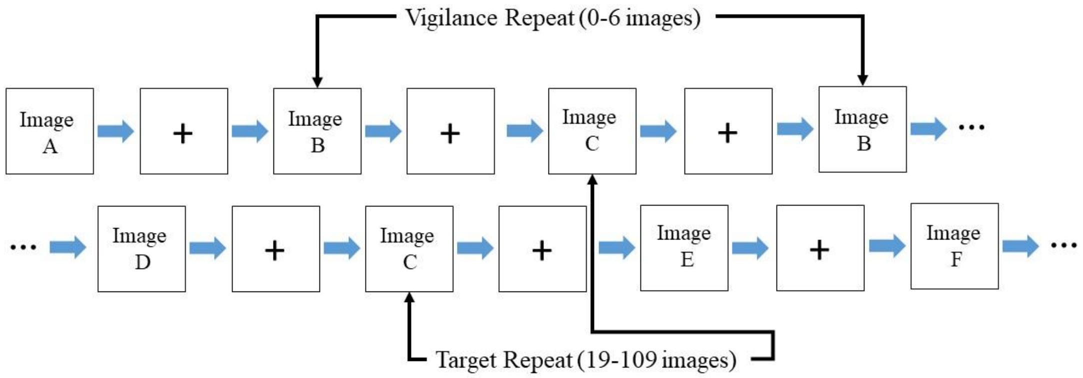


Note. Each trial started with a fixation presented for 0.8-1.2 second. Then Mondrian images were presented on the dominant eye and face images were presented on the non-dominant eye. The contrast of face images ramped up from 0 to 1 within 1 second and remained at 1. Each trial ended when participants responded or after 3 seconds.

doi:10.6342/NTU202500760

Figure 4

Experiment Three Procedure (Memory Experiment)

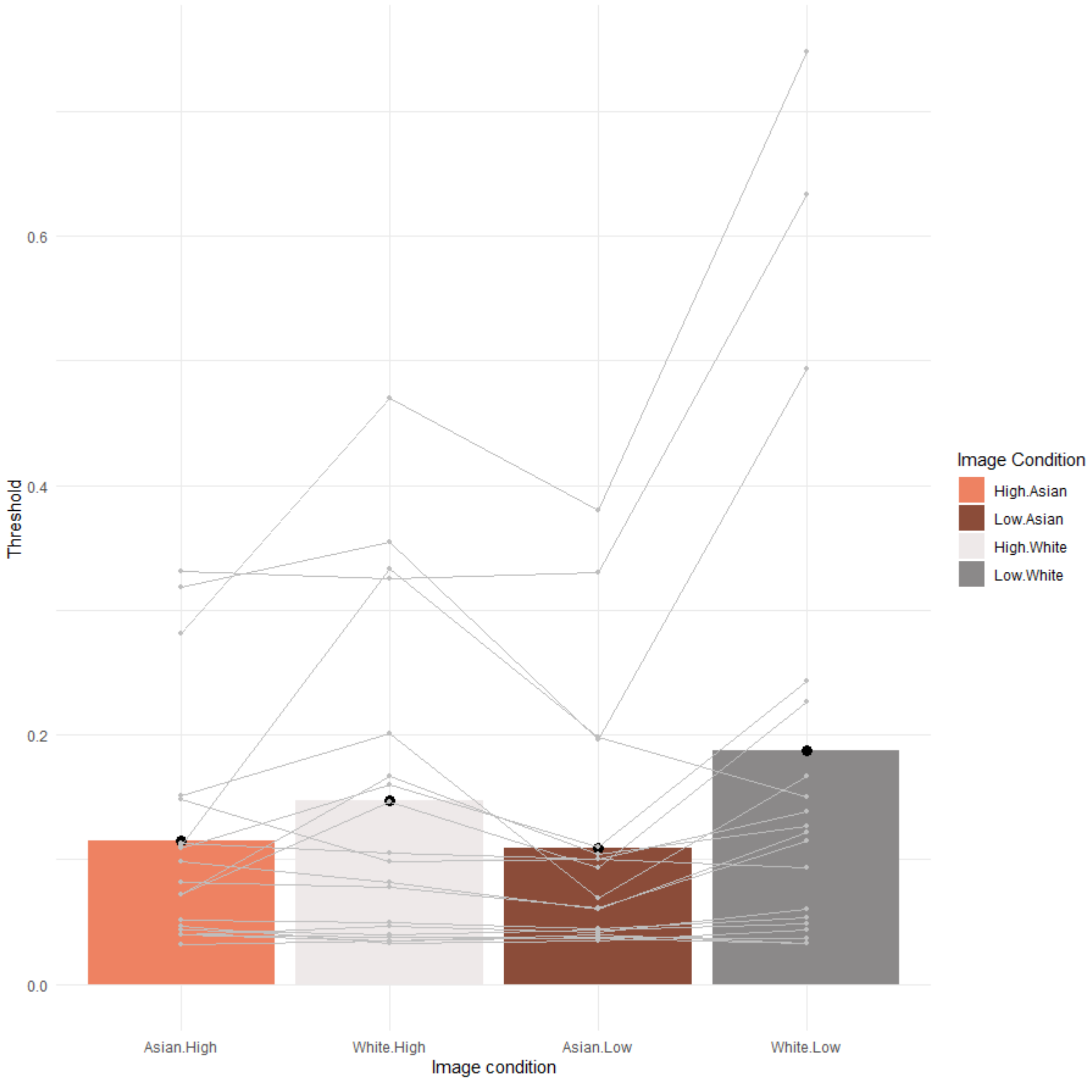


Note. Each image was presented for 600 ms, followed by a fixation lasting 800 ms. Participants were instructed to respond whenever an image repeated. Vigilance repeats occurred with an interval of 0–6 images, while target repeats occurred with an interval of 19–109 images.

doi:10.6342/NTU202500760

Figure 5

Analysis 1-1: Results (Excluding participants according to average performance)

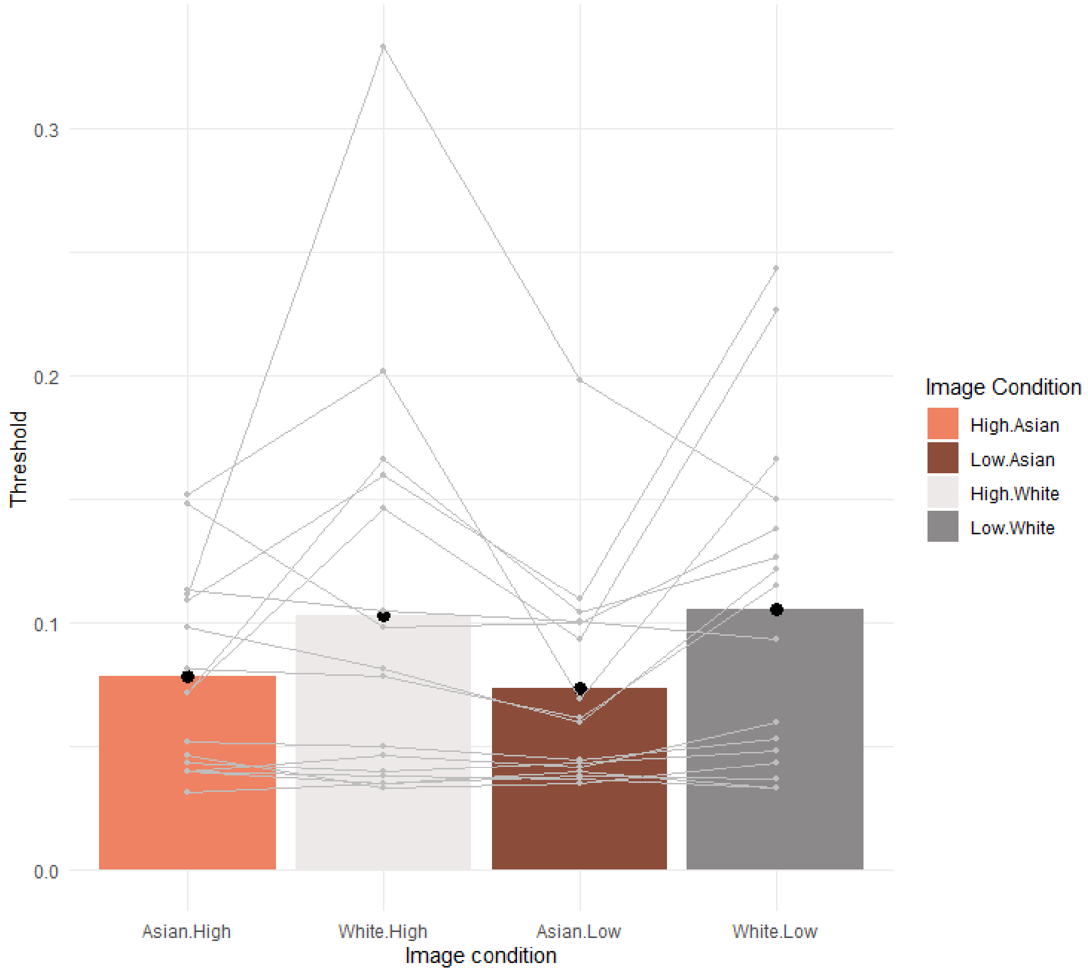


Note. The black dots represent the average for each condition. The linked gray lines represent the data from individual participants. 19 participants were included in this analysis.

doi:10.6342/NTU202500760

Figure 6

Analysis 1-2: Results (Excluding participants according to performance within condition)

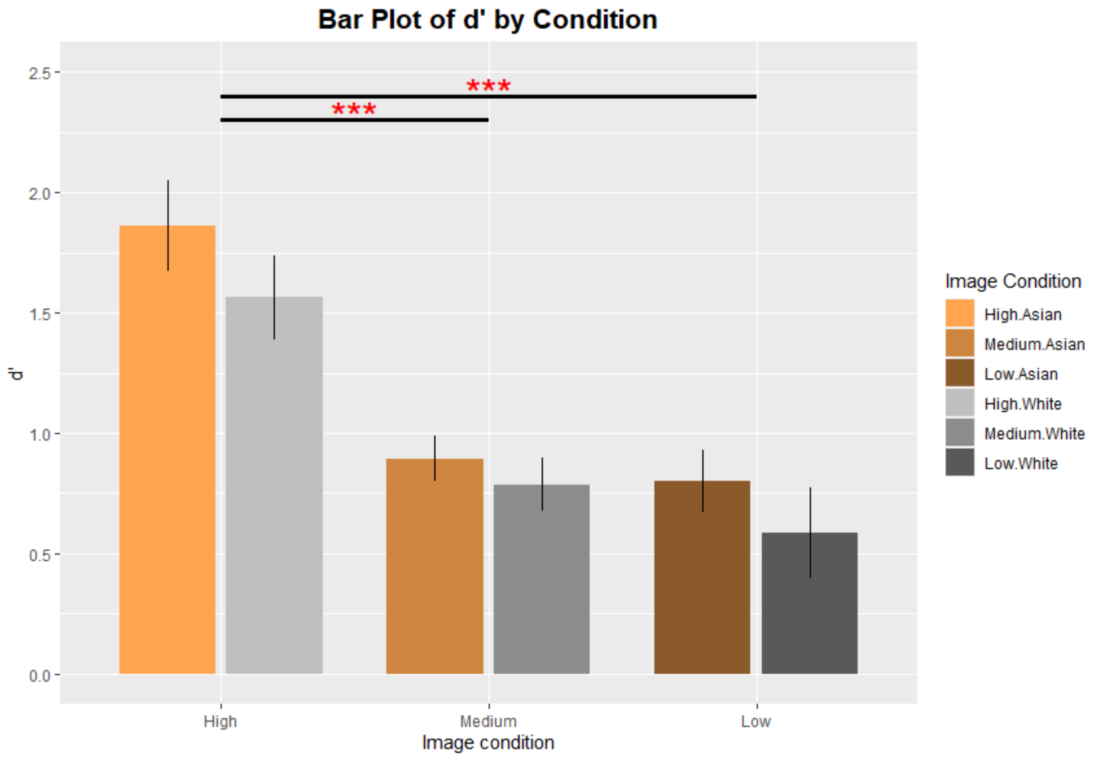


Note. The black dots represent the average for each condition. The linked gray lines represent the data from individual participants. 16 participants were included in this analysis.

doi:10.6342/NTU202500760

Figure 7

Analysis 2-1: Results of d' Analysis (Group Level)

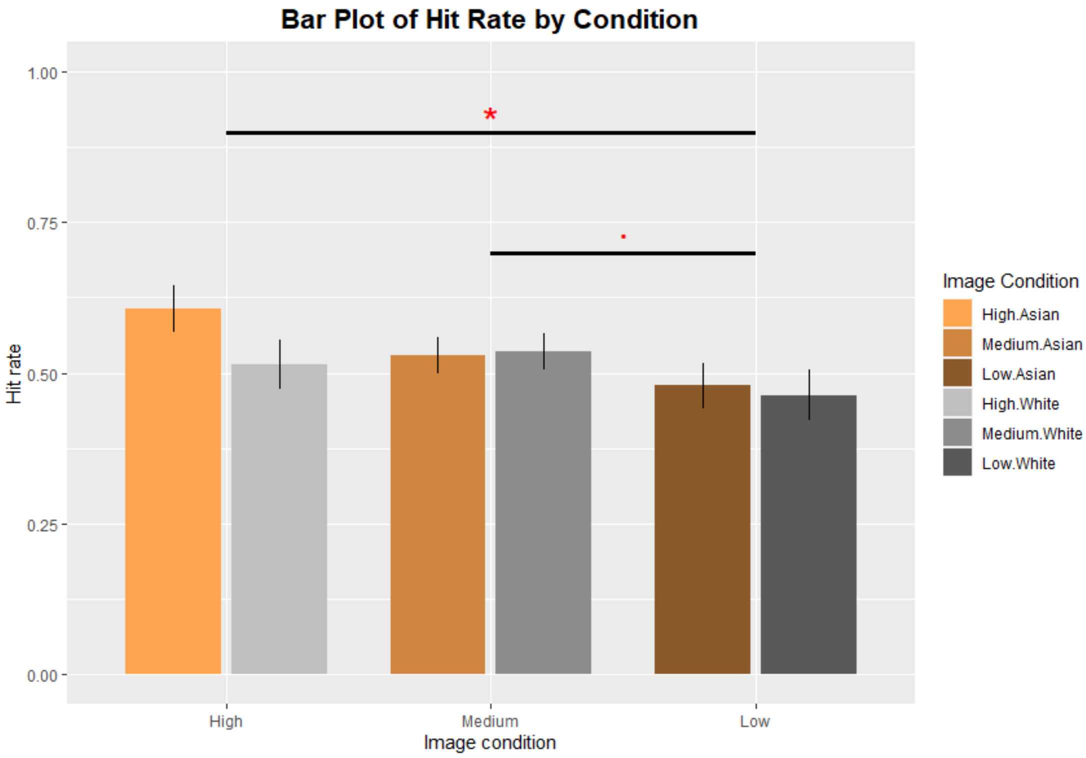


Note: Error bars represent the standard error for each condition. The symbols indicate the significance levels of the pairwise comparisons, following the notation used in R: ***$p < .001$. **$p < .01$. *$p < .05$. • $p < .1$.

doi:10.6342/NTU202500760

Figure 8

Analysis 2-1: Results of Hit Rates Analysis (Group Level)

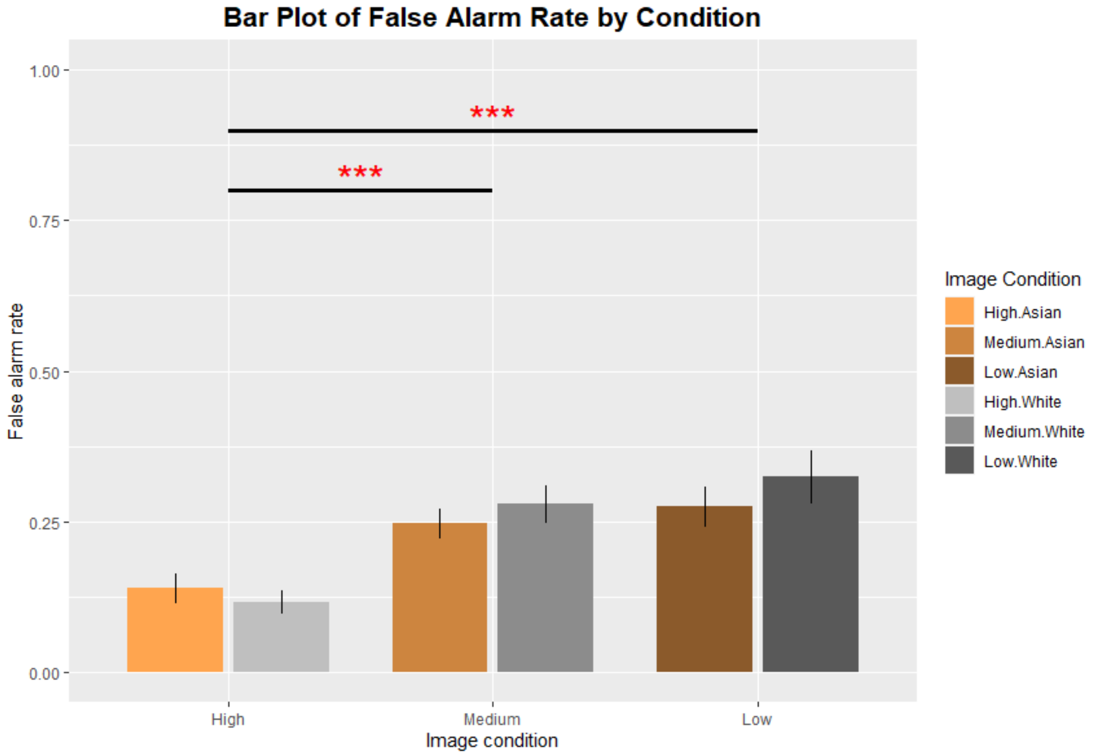


Note: Error bars represent the standard error for each condition. The symbols indicate the significance levels of the pairwise comparisons, following the notation used in R: ***p < .001. **p < .01. *p < .05. ·p < .1.

doi:10.6342/NTU202500760

Figure 9

Analysis 2-1: Results of False Alarm Rates Analysis (Group Level)

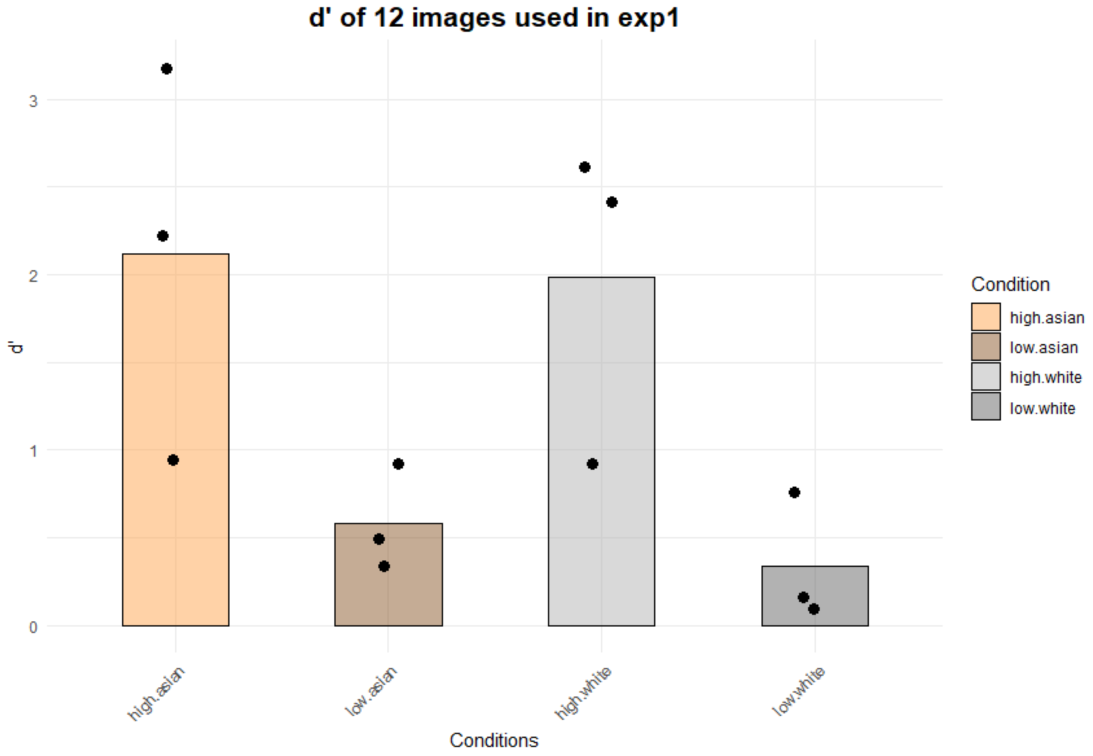


Note: Error bars represent the standard error for each condition. The symbols indicate the significance levels of the pairwise comparisons, following the notation used in R: ***$p < .001$. **$p < .01$. *$p < .05$. ·$p < .1$.

doi:10.6342/NTU202500760

Figure 10

Analysis 2-2: Results of d' Analysis (Image Level)

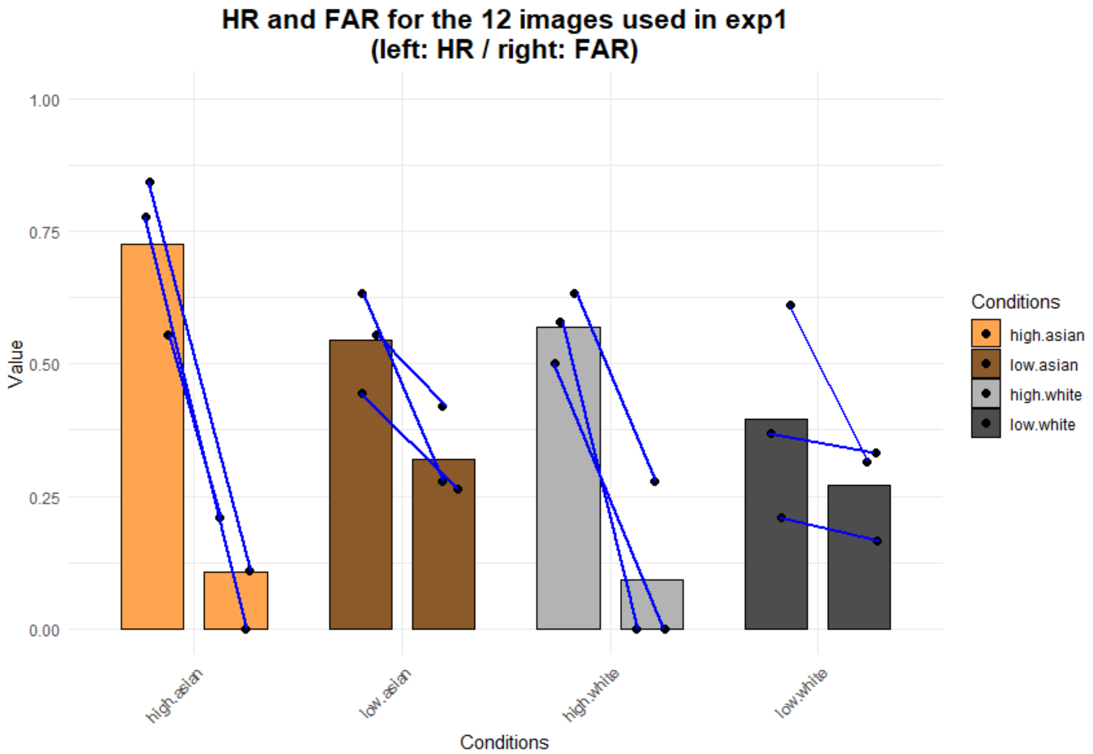


Note. The black dots represent the data of each image.

doi:10.6342/NTU202500760

Figure 11

Analysis 2-2: Results of Hit Rates and False Alarm Rates Analysis (Image Level)

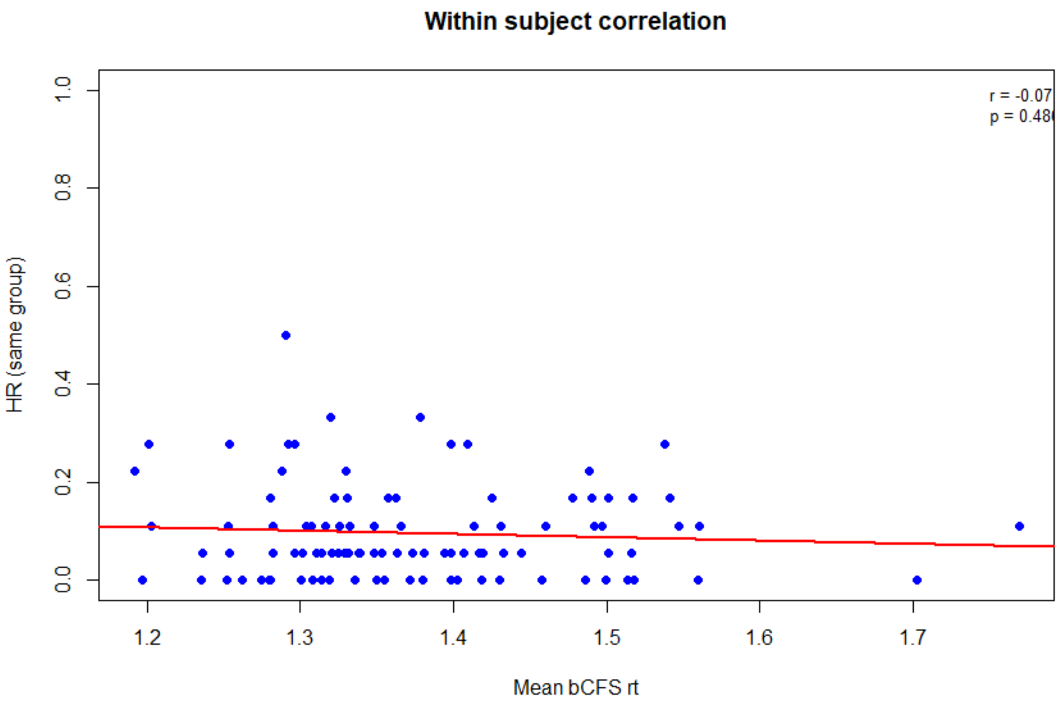


Note. The black dots represent the data for each image, with blue lines connecting data points corresponding to the same images. Within each condition, the left and right bars represent hit rates and false alarm rates, respectively.

doi:10.6342/NTU202500760

Figure 12

Analysis 3-1: Results of Correlation Between Memorability Scores and bCFS Reaction Times (Within Subject)

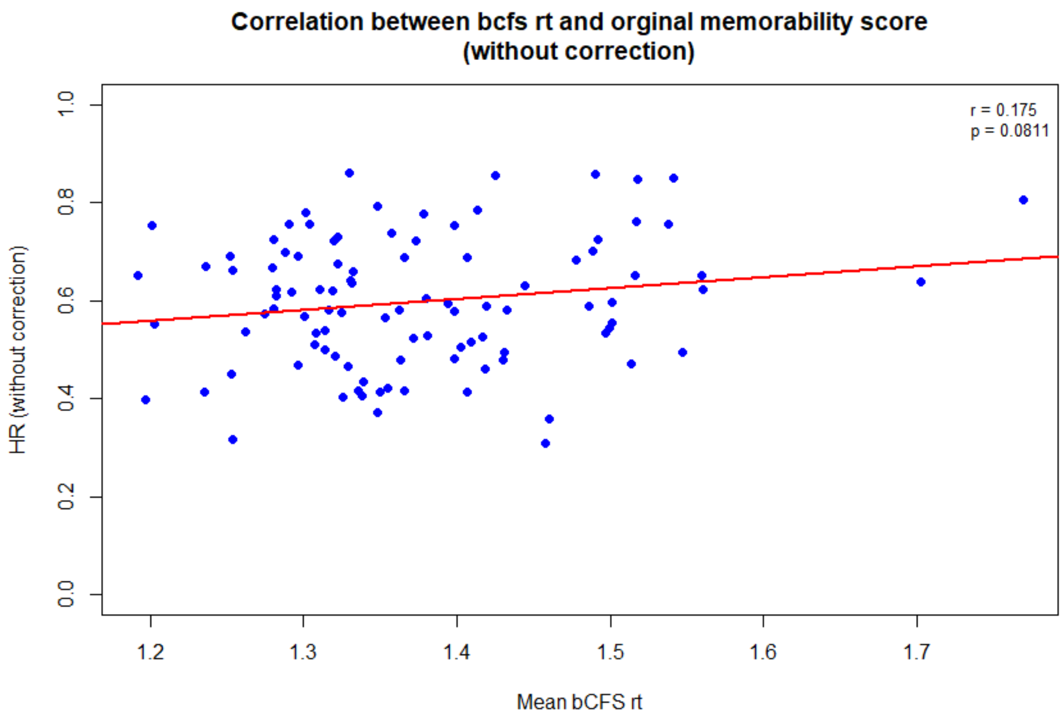


Note. The blue dots represent the data for each image. The red line represents the estimated regression line.

doi:10.6342/NTU202500760

Figure 13

Analysis 3-2: Results of Correlation Between Memorability Scores and bCFS Reaction Times (Without Correction)

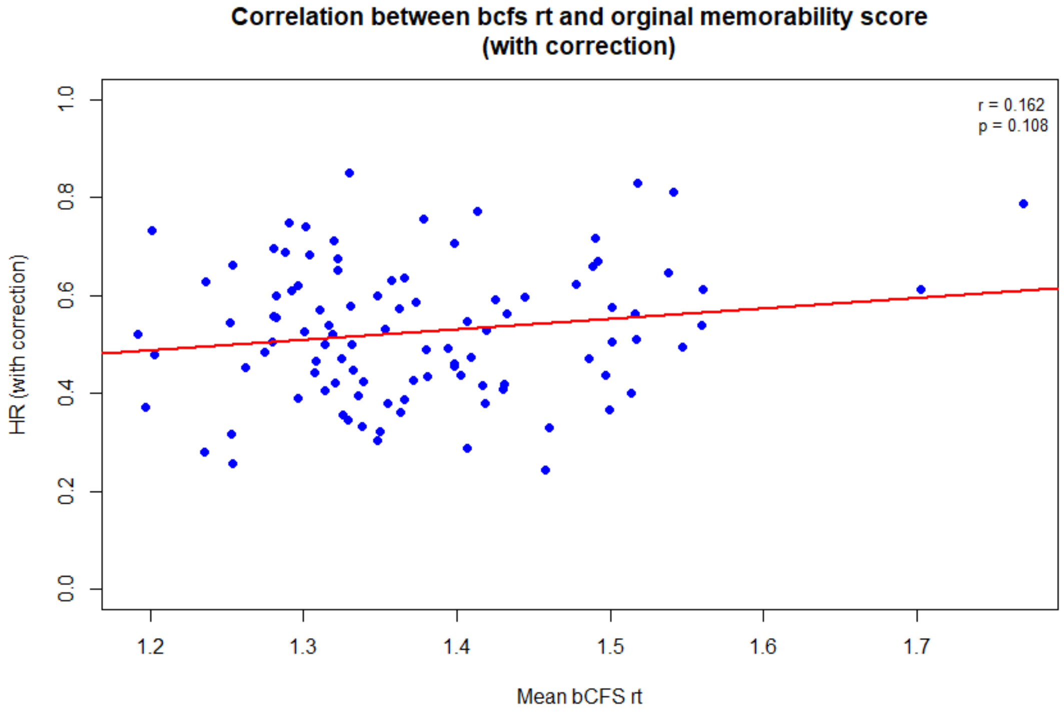


Note. The blue dots represent the data for each image. The red line represents the estimated regression line. The memorability scores (without correction) provided by the database were defined as the hit rates in their experiment.

doi:10.6342/NTU202500760

Figure 14

Analysis 3-2: Results of Correlation Between Memorability Scores and bCFS Reaction Times (with Correction)

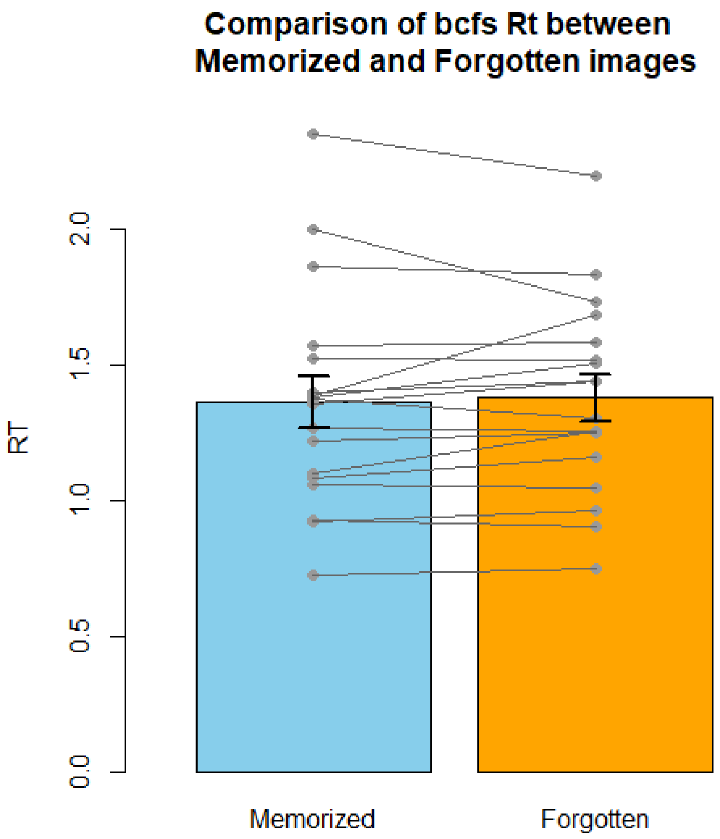


Note. The blue dots represent the data for each image. The red line represents the estimated regression line. The memorability scores (with correction) provided by the database were defined as the hit rates minus false alarm rates in their experiment.

doi:10.6342/NTU202500760

Figure 15

Analysis 3-3: Results of Comparison bCFS Reaction Times Between Memorized and Forgotten Images



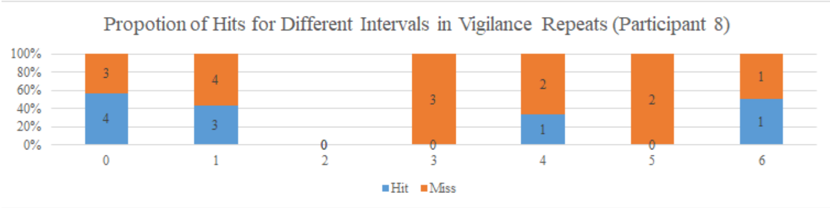
Note. The error bars represent the standard error for each condition. The linked grey lines represent the data from individual participants.

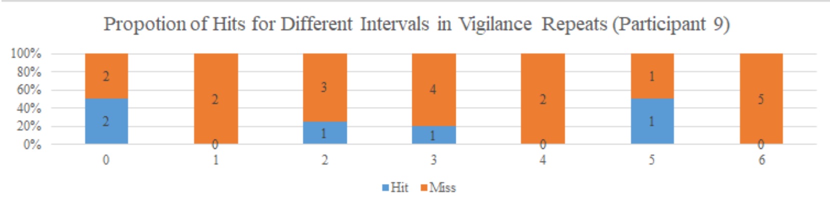
doi:10.6342/NTU202500760

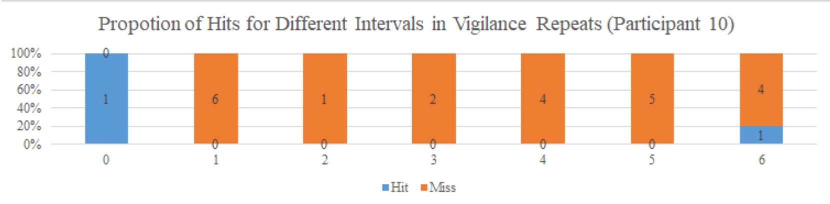
Figure 16

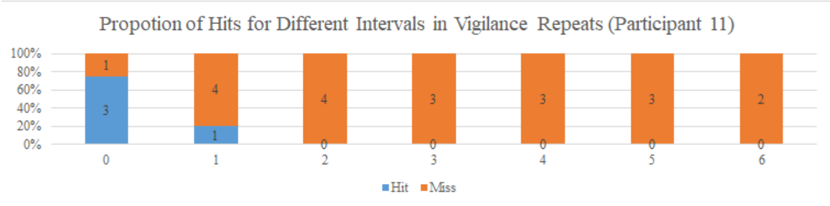
Bar Plots Showing the Hit Rates for Different Intervals in Vigilance Repeats for Each Participant.

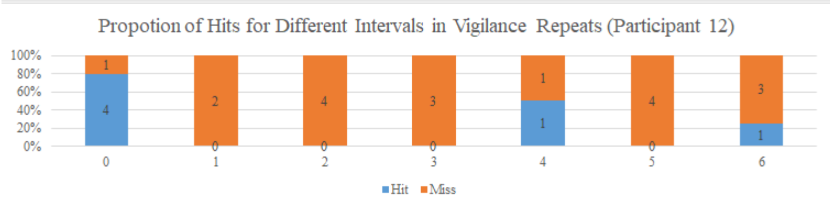
doi:10.6342/NTU202500760

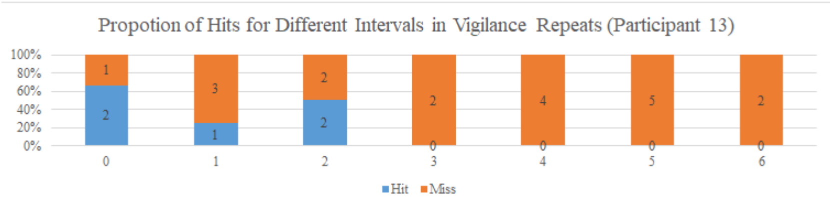


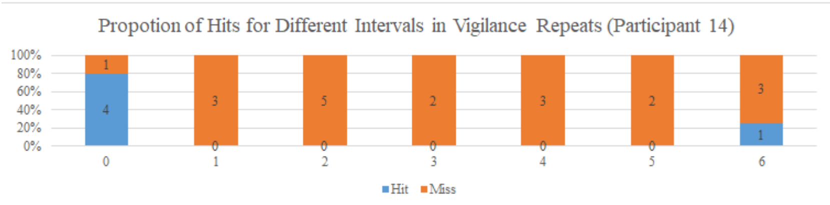


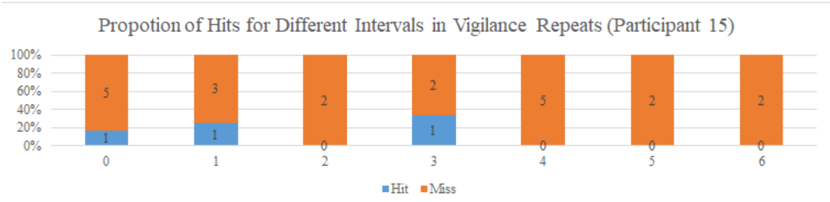


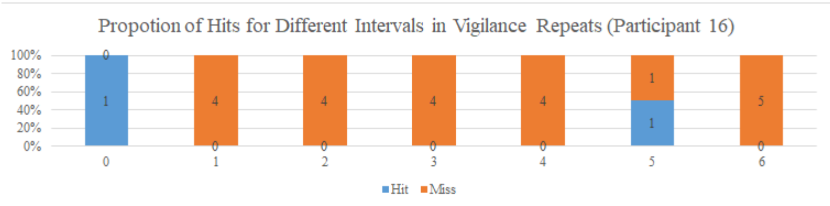




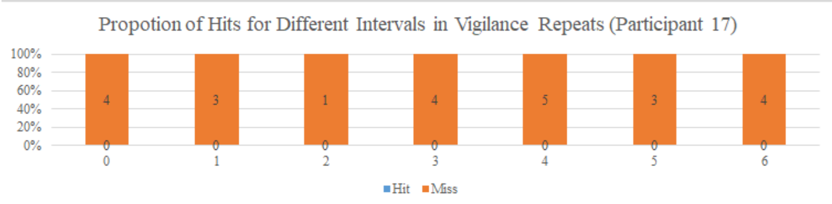


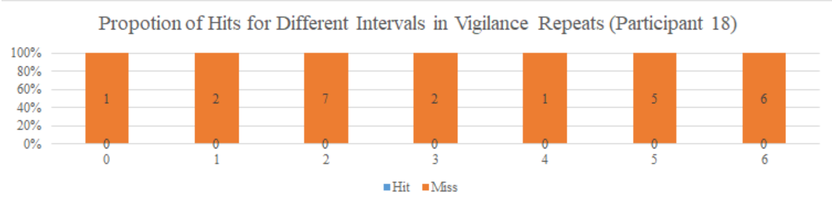


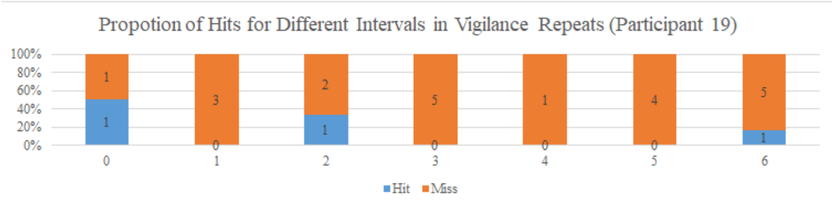


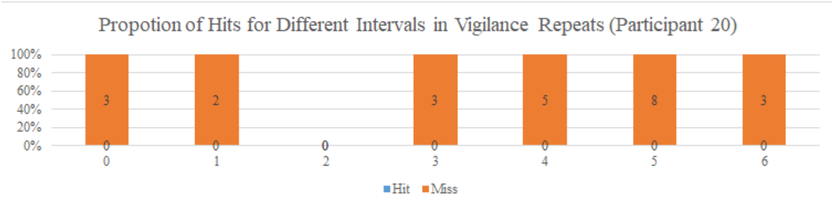


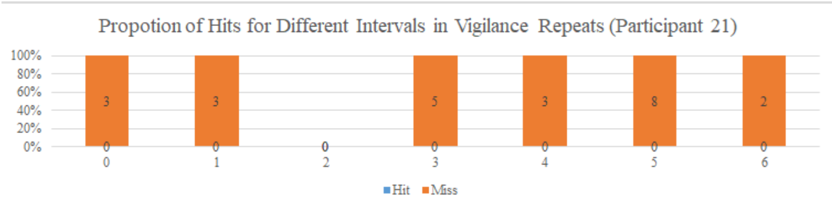
doi:10.6342/NTU202500760

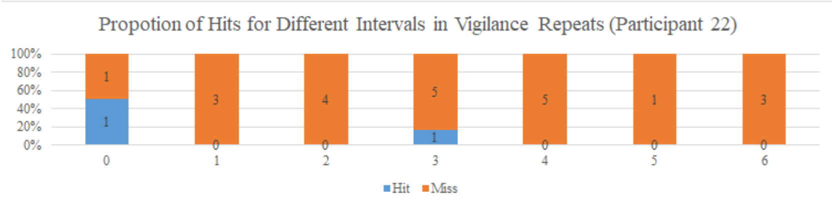


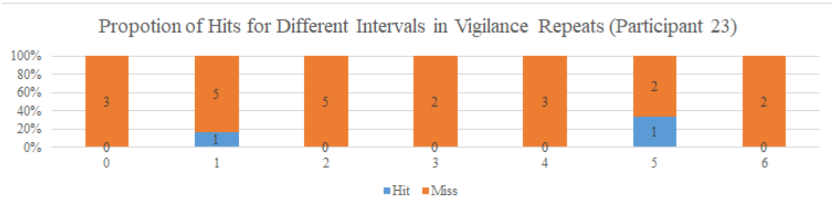


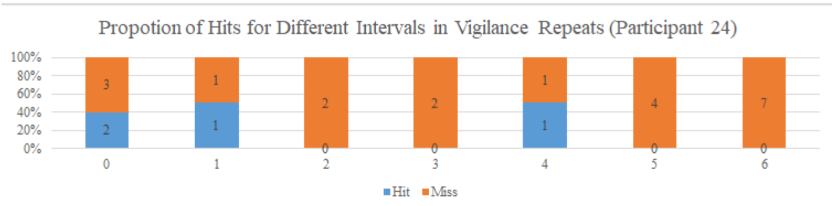


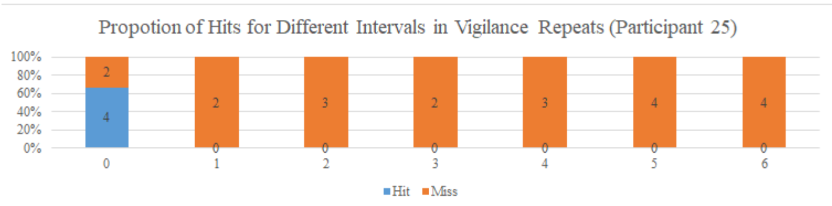












doi:10.6342/NTU202500760

References

Alais, D., Coorey, J., Blake, R., & Davidson, M. J. (2023). A new 'CFS tracking' paradigm reveals uniform suppression depth regardless of target complexity or salience. *ELife*, *12*, 1–27. https://doi.org/10.7554/eLife.91019

Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia*, *141*(773), 1–43. https://doi.org/10.1016/j.neuropsychologia.2020.107408

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, *142*(4), 1323–1334. https://doi.org/10.1037/a0033872

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2306–2315. https://doi.org/10.1109/TVCG.2013.234

Bower, G. H., Karlin, M. B., & Bower, H. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology* (Vol. 103, Issue 4).

Broers, N., Potter, M. C., & Nieuwenstein, M. R. (2018). *Enhanced recognition of*

doi:10.6342/NTU202500760

memorable pictures in ultra-fast RSVP. 1080–1086. https://doi.org/10.3758/s13423-017-1295-7

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Cohendet, R., Demarty, C.-H., Duong, N., & Engilberge, M. (2019). VideoMem: Constructing, analyzing, predicting short-term and long-term video memorability. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2531–2540. https://doi.org/10.1109/ICCV.2019.00262

Dal Ben, R. (2023). SHINE_color: Controlling low-level properties of colorful images. *MethodsX*, *11*(September), 102377. https://doi.org/10.1016/j.mex.2023.102377

Dixon, W. J., & Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, 43(241), 109-126.

Dubey, R., Peterson, J., Khosla, A., Yang, M. H., & Ghanem, B. (2015). What makes an object memorable? *Proceedings of the IEEE International Conference on Computer Vision*, *2015 Inter*, 1089–1097. https://doi.org/10.1109/ICCV.2015.130

Ekman, P. (1992). Facial expressions of emotion: new findings, new questions. *Psychological Science*, *3*(1), 34–38. https://doi.org/10.1111/j.1467-9280.1992.tb00253.x

doi:10.6342/NTU202500760

Faivre, N., Berthet, V., & Kouider, S. (2012). Nonconscious influences from emotional faces: A comparison of visual crowding, masking, and continuous flash suppression. *Frontiers in Psychology*, *3*(MAY), 1–13. https://doi.org/10.3389/fpsyg.2012.00129

Feingold, G. A. (1914). The influence of environment on identification of persons and things. *Journal of the American Institute of Criminal Law and Criminology*, *5*(1), 39. https://doi.org/10.2307/1133283

Goetschalckx, L., & Wagemans, J. (2019). MeMCAT: A new category-based image set quantified on memorability. *PeerJ*, *2019*(12). https://doi.org/10.7717/peerj.8169

Han, S., Rezanejad, M., & Walther, D. B. (2023). Memorability of line drawings of scenes: The role of contour properties. *Memory and Cognition*, *October*. https://doi.org/10.3758/s13421-023-01478-4

Hung, S. M., Nieh, C. H., & Hsieh, P. J. (2016). Unconscious processing of facial attractiveness: Invisible attractive faces orient visual attention. *Scientific Reports*, *6*. https://doi.org/10.1038/srep37117

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *Journal of Vision*, *11*(11), 1282–1282. https://doi.org/10.1167/11.11.1282

Jeong, S. K. (2023). Cross-cultural consistency of image memorability. *Scientific*

doi:10.6342/NTU202500760

Reports, *13*(1), 1–8. https://doi.org/10.1038/s41598-023-39988-5

Kramer, M. A., Hebart, M. N., Baker, C. I., & Bainbridge, W. A. (2023). The features underlying the memorability of objects. *Science Advances*, *9*(17), 1–14. https://doi.org/10.1126/SCIADV.ADD2981

Maguire, J. F., & Howe, P. D. L. (2016). Failure to detect meaning in RSVP at 27 ms per picture. *Attention, Perception, and Psychophysics*, *78*(5), 1405–1413. https://doi.org/10.3758/s13414-016-1096-5

Nakamura, K., & Kawabata, H. (2018). Preferential access to awareness of attractive faces in a breaking continuous flash suppression paradigm. *Consciousness and Cognition*, *65*(March), 71–82. https://doi.org/10.1016/j.concog.2018.07.010

Ogmen, H., Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research*, *43*(12), 1337–1350. https://doi.org/10.1016/S0042-6989(03)00138-X

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion*, *5*(4), 498–502. https://doi.org/10.1037/1528-3542.5.4.498

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203.

doi:10.6342/NTU202500760

https://doi.org/10.3758/s13428-018-01193-y

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12), 1136–1169. https://doi.org/10.1167/4.12.12

Stein, T. (2012). *Visual processing of social information during interocular suppression*.

Stein, T., Hebart, M. N., & Sterzer, P. (2011). Breaking continuous flash suppression: A new measure of unconscious processing during interocular suppression? *Frontiers in Human Neuroscience*, *5*(NOVEMBER), 1–17. https://doi.org/10.3389/fnhum.2011.00167

Tamietto, M., & De Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, *11*(10), 697–709. https://doi.org/10.1038/nrn2889

Xie, W., Bainbridge, W. A., Inati, S. K., Baker, C. I., & Zaghloul, K. A. (2020). Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. *Nature Human Behaviour*, *4*(9), 937–948. https://doi.org/10.1038/s41562-020-0901-2

doi:10.6342/NTU202500760