The Effect of Face Memorability on Remembering Face-Name Associations: An Ethnically Inclusive Design

Jan Hülsewiesche
S3927547
Department of Psychology, University of Groningen

Honours Bachelor Thesis

Supervisor: dr. M.R. Nieuwenstein

Secondary Supervisor: dr. M.M. Span

July 15, 2022

Abstract

Own-Race-Bias (ORB) is known to reduce the memorability of information if this information is associated with an ethnical in-group instead of an out-group. A previous study in our laboratory found that the memorability of a face transfers to the memorability of an associated name (Van der Wal & Nieuwenstein, 2021). The present study aimed to replicate this result in an ethnically inclusive design. Thereby, we offered an indirect test whether the face memorability scores of Bainbridge (2013) are contaminated by the ORB, which has not been tested before. We conducted a high-powered study with 120 participants to form three groups of white, black, and other ethnicities. A 3 × 2 × 2 mixed ANOVA was conducted with the factors of participant ethnicity, face memorability and match of the participant and face ethnicity. The dependent variable was a face-name recall. We hypothesised that face-name recall would be enhanced for memorable faces. Secondly, an advantage for matching trials was hypothesised. Thirdly, the match effect would be stronger in the white and the black group than in the Other group. Our results showed a clear advantage in face-name recall for memorable faces but ambiguity for the effect of matching faces. We conclude that face memorability transfers to associated information and that face memorability seems universal across ethnicities. The transfer of face and image memorability to associated information could be applied to manipulate neural encoding strength in cognitive neuroscience and the advertising industry.

Keywords: face memorability, face-name recall, Own-Race Bias

The Effect of Face Memorability on Remembering Face-Name Associations: An Ethnically Inclusive Design

Oh, he is asking for my name again. Haven't I told him twice already? Why doesn't he remember? Am I that forgettable?

What determines whether we remember a face-name association? Remembering an acquaintance's name is a crucial social skill. Nevertheless, remembering a proper name is a task more difficult than remembering other attributes like occupation because proper names do not convey a semantic framework as rich as other attributes do. Proper names are meaningless, while occupations, hobbies, and similar attributes typically convey a large amount of semantic information, associations, and stereotypes about a person (Cohen, 1990). For example, the occupation *baker* can convey information about a person's habits (getting up early in the morning to bake), about the workplace (a bakery), and possibly even stereotypes (e.g. "bakers are conservative people").

In contrast, the family name *Baker* is nearly meaningless. It conveys nearly no information about a person, besides that the person's native language could be English. This lack of semantic information makes face-name memory less stable because a face-name association has to rely on much fewer associations between a face and a name. It is unsurprising that face-name memory is also disproportionately affected in age-related cognitive decline (Naveh-Benjamin et al., 2004). Finding the determinants of the strength of encoding a face-name association could increase our understanding of associative memory and the tools to study it. The present study explores if the memorability of a face influences face-name recall.

The Concept of *Intrinsic Image Memorability*

Recently, a concept that received increased attention in memory research is the concept of intrinsic image memorability. The term intrinsic memorability was coined by Isola and colleagues (2011), who showed that some photographs were more easily recognised after a single view than other images. Specifically, an extensive database of 2222 photographs was created from images of various semantic categories (e.g. nature scenes, city scenes, buildings, objects, images with people, etc.). The probability that each of these images would be recognised after a single view was estimated with a memory game for Amazon Mechanical Turk workers. The workers were presented with sequences of photographs on a computer screen. Some of these images (the *targets*) appeared twice in the sequence and the participants were asked to detect such repetitions. The results showed that, across participants, the same photographs were more frequently recognised. The average probability that an image is recognised after a single view they called "*memorability*". It was found that the variance in memorability could partially be explained by various semantic attributes such as the presence of certain objects. For example, *images with people with visible faces* were consistently more memorable than other images. Nevertheless, a substantial amount of variance could not be explained by the visual and conceptual content of the images nor by participants' subjective judgements of memorability or aesthetics. Yet memorability differences between images displayed consistency over different retention intervals and across participants. Isola and colleagues suggested that the unexplained variance in memorability is rooted in the yet undefined concept of intrinsic image memorability, that thus far could only be estimated based on the memory performance of human participants (Isola et al, 2011).

Intrinsic *Face* Memorability

With this study, Isola and colleagues (2011) prompted other researchers to investigate the concept of memorability *within a single semantic category*, namely human faces (Bainbridge et al., 2013). Bainbridge and colleagues created a memory game like Isola and colleagues did, but they exclusively utilised human faces as stimuli: Participants were presented with blocks of image sequences and were asked to indicate when an image was presented for the second time. The results mirrored the findings by Isola et al. (2011): Some faces were consistently remembered better across participants than others. In a subsequent analysis, it was examined whether the memorability of face images could be explained using people's judgments of facial attributes such as typicality, emotional stability, kindness, and subjective memorability. These attributes were rated by Amazon Mechanical Turk workers upon seeing one face at a time. Based on several analyses using different combinations of these attributes, Bainbridge et al. concluded that memorability was only to a small degree explained by the personality, social, and memory-related attributes. Bainbridge et al. concluded that memorability is a parameter that is intrinsic to the face and largely independent of subjective facial characteristics.

Intrinsic Face Memorability and Face-Name Recall

Van der Wal and Nieuwenstein (2021) examined whether the memorability of a face would influence the memorability of the corresponding Face-Name association. They selected 20 memorable and 20 non-memorable faces from the image database provided by Bainbridge et al. (2013). Participants were asked to learn face-name associations for both sets, with names presented acoustically. The results showed that names paired with memorable faces were remembered better. Importantly, however, this study was done using only white participants and white faces and therefore the generalizability of this effect remains to be determined.

A potential problem for the generalizability of the results obtained by Van der Wal and Nieuwenstein is the own-race bias in face memory. This bias is a type of the own-category biases, which refer to findings showing that participants remember information about a person better if the person belongs to the same group as the participant (Brewer, 1979; Hills & Pake, 2013; Meissner & Brigham, 2001; Obermeyer et al., 2019; Rhodes & Anastasi, 2012; Slone et al., 2000). For example, participants remember faces better when the face matches the participant's age (Rhodes & Anastasi, 2012; Strickland-Hughes et al., 2020; Ziaei et al., 2019), race (Harvey, 2014; Hills & Pake, 2013; Meissner & Brigham, 2001; Wong et al., 2021), and even experimentally assigned groups identities (Brewer, 1979). Accordingly, an interesting question is whether the face-memorability advantage found for face-name recall in the study by Van der Wal and Nieuwenstein (2021) would also be found when one uses non-white participants and face stimuli.

There is reason to believe that the ORB might invalidate the memorability scores in Bainbridge et al. (2013). Bainbridge and colleagues took the ORB into account by matching the race distribution of the faces in the database with the race distributions of participants: Both the faces and the participants were 80% Caucasian white. Therefore, the memorability scores for all faces, including those of faces with an ethnicity other than white, were evaluated by 80% white participants, while memorability scores for other ethnicities such as "South Asian" were almost exclusively scored by members of an out-group. Arguably, the memorability of a face, and hence the difference in memorability between faces, could depend on whether memorability is based on the recognition performance of in or out-group members. This could happen if some explanations for the in-group bias are correct. As Hills and Pake (2013) investigated, one such explanation is the following. Identification of another individual is mediated by *physiognomic diagnosticity*. One uses those aspects of the other individual's face for identification that vary most between individuals. For example, a white person who grew up among other whites knows that hair colours vary greatly. Some people have brown hair, some have black hair, some are blond, and some are red. With only very few other features, this person can quickly identify any person he regularly encounters. For instance, he may immediately identify the person with red, curly hair as "Jack". In contrast, when he visits a predominantly black neighbourhood, his usual identification cues become useless. All people he encounters have very similar hair colours. When they identify each other, they may pay more substantial attention to other cues, as the nose (Hills & Pake, 2013).

Suppose such qualitative differences exist between ethnic groups. In that case, the memorability of a black face for a white participant may be virtually independent of the memorability for a black participant. Not only would black participants create higher hit rates, but even which faces are counted as memorable or non-memorable would change dramatically.

The Problem of the ORB from a Neurocognitive Perspective

Several theories have been proposed to explain the process of face-name recall. Among those are the Sequential Stage Model (Bruce & Young, 1986), the Representational Model (Cohen, 1990), as well as the Interactive Activation and Competition Model (Burton & Bruce, 1992), and Pattern Completion (Liu et al., 2016). All these models build on the idea that better face-name recall results from more, stronger, and more direct the associations between a network of activated neurons representing the face and another network representing the name. Neuroimaging has shown that both the neural representation of the name and the face depend on structures in the medial temporal lobe and the fusiform gyrus, located in the cortex of the basal temporal lobe (Bainbridge & Rissman, 2018).

A memorable image, possibly through its more pronounced facial features (e.g. a big nose, asymmetrical eyes), or other presently unknown mechanisms, seems to cause the brain to form more and stronger connections between the face and the name. These stronger associations may lead to easier activation of the name, in response to encountering a face.

Yet, research has shown that the ORB is a "top-down" factor that impacts activity in the medial temporal lobe and in the fusiform gyrus, the brain regions that mediate face-name memory (Chiroro & Valentine, 1995). Activity in the same brain areas, namely the fusiform gyrus and the medial temporal lobe are also affected by memorability (Bainbridge & Rissman, 2018). The ORB is thought to reduce encoding strength in these areas, while memorability enhances it. As both mechanisms exert their effects on the same neural structures, the ORB moderating the effect of memorability is plausible but not necessary. Before researchers employ memorability as an independent variable, it would be practical for studies to show that the memorability scores of Bainbridge et al. are universal, so that researchers are not confined to the use of white participants from the United States.

The Present Study

To sum up, there are reasonable arguments that the memorability scores by Bainbridge et al. (2013) could be contaminated by own-race bias: The memorability scores for minority faces were derived from a sample of predominantly white participants. The phenomenon of the ORB and the mechanism described by Hills and Pake (2013) allow us to question the universality of the memorability scores obtained by Bainbridge et al. (2013). Yet, despite the presence of a plausible mechanism for the contamination of the memorability scores, thus far there has been *no evidence* that the memorability scores are indeed contaminated by the own race bias. Hence, it remains to be determined whether a memorable face for a white person also constitutes a memorable face for a non-white person.

Van der Wal and Nieuwenstein (2021) established an effect of memorability on face-name recall in white participants while disregarding faces and participants from other ethnicities. A successful, ethnically inclusive replication would provide support for the generalizability of this finding, and, hence, the generalizability of face memorability. Alternatively, if we only replicate the effect of face memorability for white faces in white participants, then this would suggest that face memorability is subject to the effects of whether the ethnicity of a face matches the ethnicity of the participant.

To examine these questions, we used the faces of people whose ethnicity was judged to be "white", "black", or "other". Similarly, we sampled participants who designated their own ethnicity as white, black, or any other ethnicity. The native language of the participants was English. Hence, we assumed them to be familiar with the presented names and their pronunciation. We hypothesised that differences in face memorability are universal. Therefore, the effect of face memorability on face-name recall should hold regardless of whether there is a match between the face's ethnicity and the participant.

To summarise, the memorability scores of Bainbridge 2013 could be contaminated by the Own-Race Bias (ORB). Hence, the purpose of the present study is to provide further evidence for the transfer of face memorability to an associated name, in a ethnically inclusive design. Thereby we also indirectly test the universality of Bainbridge et al.'s memorability scores. We hypothesise that names paired with memorable faces will be remembered better than those paired with non-memorable faces (hypothesis 1). Based on the likely presence of an Own-Race Bias, we expect that names will be recalled better if the ethnicity group of the face and the ethnicity of he participant match (hypothesis 2). We anticipated that the own-race effect might interact with the participant ethnicity group because the participants and faces included in the Other group comprise a mixture of different ethnicities other than white or black. Hence, the ethnicity of the participant and faces will not always match for the Other group, resulting in a reduced own-race effect for participants in the Other group (Hypothesis 3).

Methods

Participants

Participant Characteristics

Participants were recruited via Prolific.com. We aimed to collect the data of 120 of US-Americans between 20 and 45 years. Of those 120 participants, we planned to sample 40 participants who indicated their ethnicity as white, 40 who indicated it as black, and 40 who defined their ethnicity as any of any other category. Our final sample deviated from these specifications, we only sampled 39 participants in the white and the black group each. The Other consisted of only 38 participants (the group name *Other* is capitalised, to signal that we are referring to the group and not to the plain language word *other*). We also allowed participants to participate who indicated to reside in Canada or the United Kingdom. Participation was rewarded with £2.5 for an experiment that took 20 minutes to complete. All participants were asked for their consent to participate before the study. They also had the option to withdraw their data after the debriefing.

Sample Size and Power

We used the effect size of our previous study (Van der Wal & Nieuwenstein, 2021) as an estimate of the effect size and utilised Gpower (Faul et al., 2007) to compute the sample size required to achieve sufficient power to replicate this effect. The previous study displayed a difference in name recall between memorable and non-memorable faces of an effect size $dz = .59$. At an alpha level of .05 we would need 20 participants for 80% power to replicate the effect of face memorability on face-name memory. However, the previous study was exclusively based on *white* participants (Van der Wal & Nieuwenstein, 2021). As we had reason to believe that the memorability scores from Bainbridge are more reliable among white participants, we decided to use a larger sample. In addition, we were interested in interaction effects in a mixed ANOVA model, which would require a larger sample for sufficient power., Finally, we decided to use 40 participants per group. This sample size of 120 participants in total yields more than 99.99% power to detect the difference we found in our previous study for all participants, and 97.83% power to detect the effect in each group. Hence, the probability of missing the effect in any of the three groups was $1 - .9783^3 = .0636$, so 6.4% assuming the effect size of the previous study.

Materials

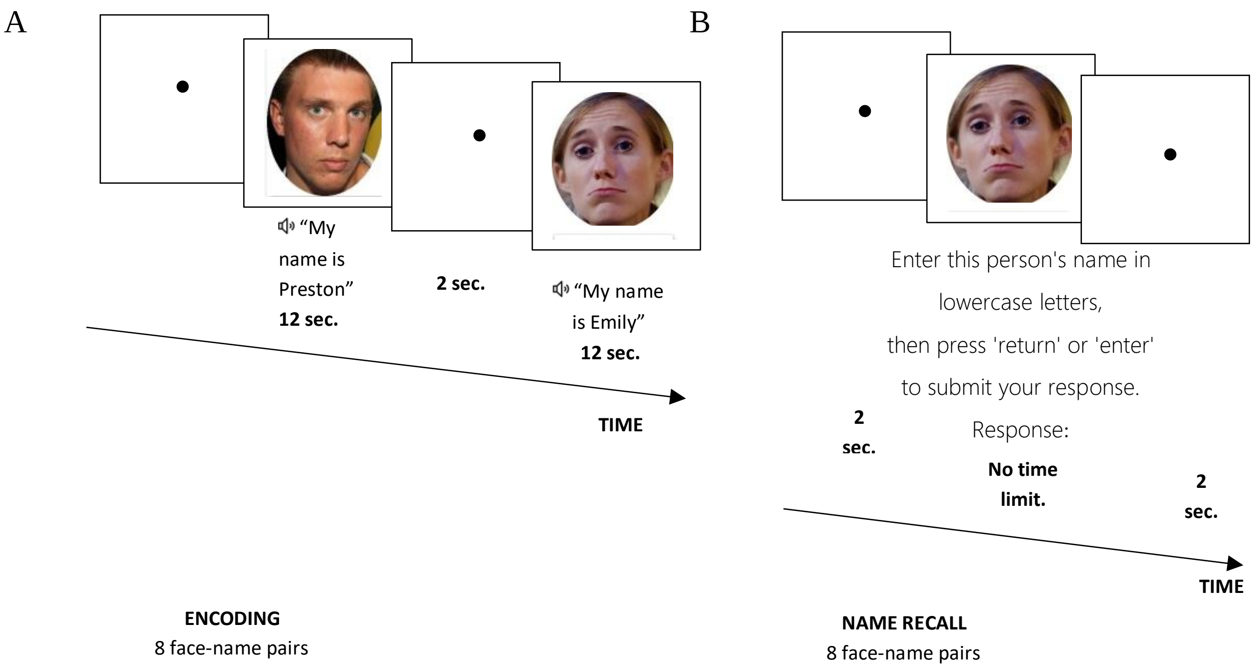
The experiment was programmed in OpenSesame 3.3 (Mathot et al., 2012). The display resolution was set to 1366 x 768 pixels. Participants performed the experiment through a web browser, using a laptop or desktop computer. The frequentist aspect of the data analysis was conducted in SPSS 27 and 28, the Bayesian analysis was conducted in JASP (JASP Team, 2022).

Face-Name Association Task

The Face-Name Association task (FNA) was based on the paradigm of Strickland-Hughes et al. (2020). The task consisted of 6 blocks with 8 face-name pairs. Every block included an encoding and a recall phase. In the encoding phase, each face was shown for 12 seconds. After 2 seconds, an audio recording was played that spoke the sentence "My name is ....". In the recall phase, participants were shown the faces again in a random order, and they were asked to type in the name of each face. Figure 1 illustrates the FNA task used in our study.

Figure 1

Example stimuli illustrating the encoding phase (Figure 1A) and the recall phase (Figure 1B) of the FNA-task.



Note. The faces displayed in this example were not used in the actual study. For licensing reasons, we could only show freely available faces in this example.

Voice stimuli

We chose to generate the audio stimuli using the voices of Mike and Mary in Microsoft's text-to-speech (TTS) service from 1998 (Online Microsoft Sam TTS Generator. n.d.) to create an ethnicity-neutral robotic voice.

Face Stimuli

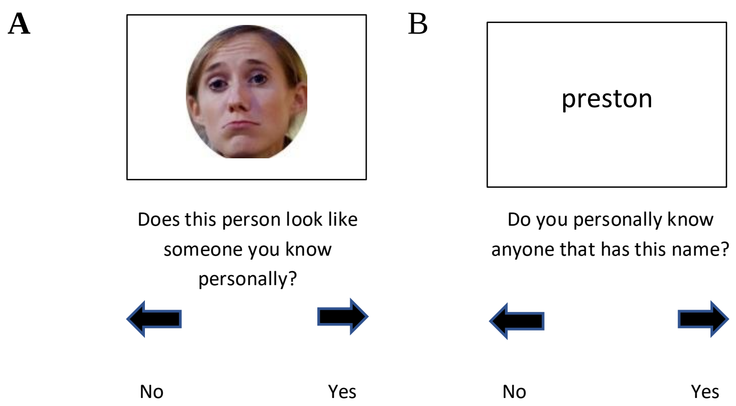
Face stimuli were selected from the 10K US Faces Database (Bainbridge et al., 2013). We selected a total of 48 faces. This included 16 faces of people which were judged by participants in Bainbridge et al. to be "white", 16 which were judged to be "black" and 16 which were judged to be of a different ethnic origin (the Other category). The last category comprised faces judged to be of East Asian (4 faces), South Asian (3 faces), Hispanic (8 faces), and Middle Eastern (1 face) ethnicity. For each subset, half of the faces were found to be memorable, whereas the others were found to be non-memorable in the study by Bainbridge et al. (2013). The average hit rates (HR) for the memorable and non-memorable faces in the white, black, and Other ethnicity categories are listed in Table 1.

Familiarity Task

Next to the FNA task, we also asked the participants whether the faces looked similar to someone they knew personally. After that, we also asked whether they personally knew people with the names used in the experiment. Figure 2 displays the familiarity task.

Figure 2

An Example illustrating the familiarity task for faces (Figure 2A), and for names (Figure 2B)



Note. The faces displayed in this example were not used in the actual study. For licensing reasons, we could only show freely available faces in this example.

Table 1:

The average HR uncorrected for false alarms, for the memorable and non-memorable faces in the white, black, and Other categories, as estimated by Bainbridge et al (2013).

	HR		
	Non-memorable	Memorable	Standardised difference (Cohen's d)
White	.32	.61	-3.7***
Black	.41	.67	-3.4***
Other	.39	.63	-3.1***
Total	.37	.64	-3.3***

Note. *<.05, one-tailed. **p<.01, one-tailed. ***p<.001, one-tailed.

We ensured an equal number of male and female faces in selecting the faces. Thus, for each combination of the face-sex category and face-ethnicity category, four memorable and four non-memorable faces were selected, yielding a total of 48 faces for which the factors of face-sex, memorability, and ethnicity group were factorially crossed.

In selecting these faces, we first filtered the database for faces of males and females whose eyes and faces were oriented towards the camera/viewer, with estimated ages between 20 and 40. For the remaining set, we aimed to ensure that our selection of memorable and non-memorable faces would yield a large difference in memorability (as defined by HR, Bainbridge et al., 2013), while showing non-significant differences on attributes that were found to explain a significant amount of variance in HR in the study by Bainbridge et al. We chose the predictors to control by selecting those with the highest correlations with HR. Hence, we controlled for the variables emotUnstable, interesting, irresponsible, kind, unattractive, and unhappy. In each race category, the HRs of memorable and non-memorable faces differed strongly (Table 1). The face images had a resolution of 192 x 256 pixels. The size of the faces, as displayed during the experiment, was 288-by-384 pixels.

Name Stimuli

The names were selected from a database of ethnically neutral names in the US (Sisense, 2022), except for the name "Kevin". Some names from the list of ethnically neutral names were adapted to enable a clear pronunciation by the TTS app. Among those was "Fatima", which was adapted to "Faddima", to eliminate an acoustic artefact. "Faddima" nevertheless mimicked the American pronounciation of the name "Fatima". Further, "Eva" was changed to "Eve".

Procedure

The study was approved by the Ethics Committee of the Faculty of Behavioural and Social Sciences at the University of Groningen. Participants who elected to participate via Prolific received a link to open the study in their browser. There, participants were informed about the purpose of the research, procedure, and treatment of personal details, but they were not informed that the purpose of the study was to determine whether a match between the ethnicity of a face and that of the participant influences face-name recall. We then collected the data using the FNA task and the familiarity task described above. The hypotheses were displayed during the debriefing at the end of the study, and participants then were given the opportunity to provide consent or to request that their data would not be used for analysis. A total of 10 participants chose to revoke their data and these participants were replaced.

Data Analysis

The dependent variable was face-name recall, that is, the proportion of faces for which the associated name was correctly recalled upon seeing the corresponding face. We applied two scoring methods to determine whether a recalled name was counted as correct. Under the first, the liberal scoring method, we counted recalled names as correct if *the recalled name differed at maximum by one phoneme from the correct name*. For example, the name *Hannah* was frequently recalled as *Anna,* and was counted as correct under the liberal scoring method. Also*, short forms* of the correct names we counted as correct under this method, such as *Mike* for *Michael*. Also, spelling errors we allowed, such as *Johnathan* for *Jonathan*. The purpose of this scoring method was to maximise construct validity. We aimed to measure whether a participant recalled the name that they encoded upon encountering the face. The robotic voice of the TTS app may create difficulties for participants to recognise every single phoneme of the name. A participant may encode *Hannah* as *Anna*, because he simply overheard the "h". Recalling the name *Anna* during the recall phase of the experiment would therefore comprise a correct recall of the previously encoded name. Short forms were scored as correct for a similar reason. In everyday life, most English speakers may rarely refer to a *Michael* as *Michael,* but instead apply his short form, *Mike*. Finally, spelling errors were probably created by lack of knowledge of the correct spelling, or simply by typing errors, in both cases the *recall* of the name was still correct. Therefore, we otherwise matching names with spelling errors as correct.

Under the second, the strict scoring method we only counted recalls as correct if the recalled name was *phonetically identical* with the correct name. Here, recalling *Hannah* as *Anna* and *Michael* as *Mike* was counted as false. Still, we allowed spelling errors. This scoring method left the least room for interpretation by the raters.

The data from the FNA task was analysed using a 3 (participant ethnicity: black, white, Other) × 2 (Match vs non-Match between the ethnicity of the participant and the face) × 2 (low vs high memorability of the face) mixed ANOVA. All statistical analyses were performed using both NHST and Bayes Factors.

Results

We analysed the data for the strict and the liberal scoring methods separately. for each scoring method, we conducted both frequentist and Bayesian repeated-measures ANOVAs with the factors participant ethnicity (between: white, black, Other) × memorability of face (within: memorable, non-memorable) × match between participant and face ethnicity (within: match, non-match). We reported appropriate *F*-tests, and we applied a significance criterion of $p < .05$. For the Bayes factors, we applied the interpretation that a BF of 1-3 is considered anecdotal, 3-10 moderate, 10-30 strong, 30-100 very strong, and BFs above 100 as extremely strong evidence for or against the presence of a tested effect.

Recall Accuracy Liberally Scored

Assumption Checks

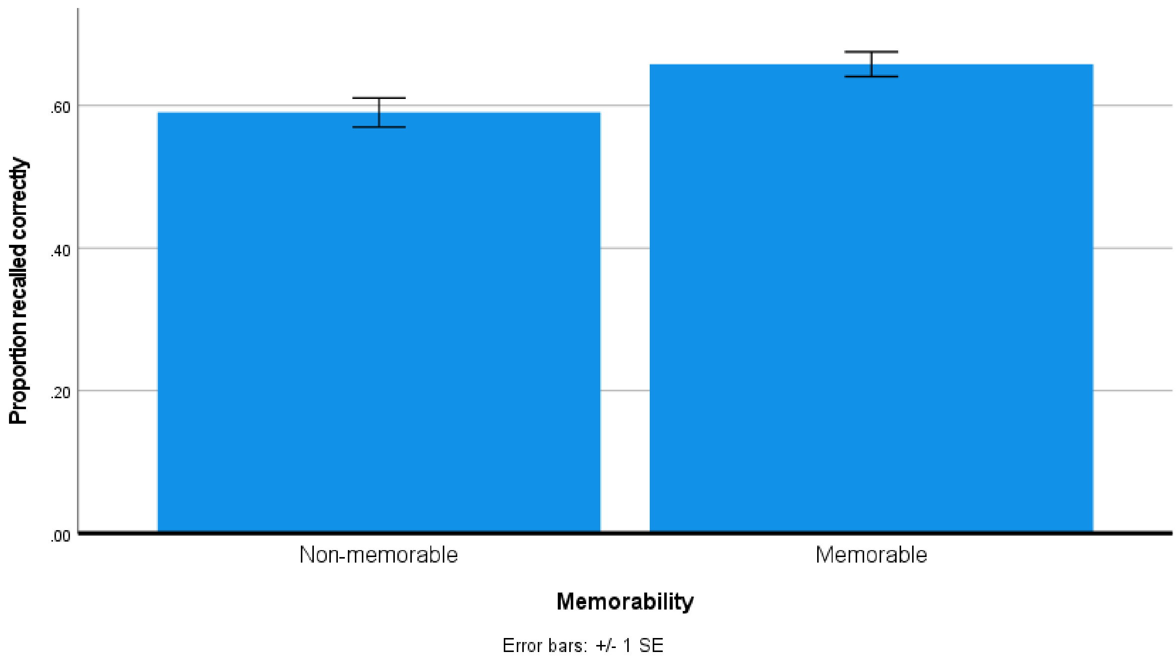
RM-ANOVA requires the assumption of independent observations, a normal distribution of residuals, and sphericity. The independence assumption is met, because each data point used in the analysis comprises an average performance per condition for each participant. As illustrated by Figure A1, the residuals of for each combination of the within-group factors are approximately normally distributed. Sphericity could not be violated because our within-factors only comprised two levels each.

The Effect of Memorability: Were Names Paired with Memorable Faces Recalled Better?

First, we examined the effect of face memorability on face-name recall. Figure 3 shows the accuracy of face-name recall with non-memorable and memorable faces. In accordance with our main hypothesis, the effect of memorability was significant, with memorable faces resulting in superior name recall ($F[1, 113] = 29.38$, $p <.001$, $\eta_p^2 = .21$, $BF_{10} = 496488.71$) and the two-way interaction of memorability and match, the two-way interaction of memorability and participant ethnicity, and the three-way interaction did not reach statistical significance. Thus, the advantageous effect of memorability was comparable across the different combinations of the three participant groups and the three ethnicity categories of the faces. In addition, the Bayesian analysis consistently showed more evidence for the absence of any interaction with memorability. The corresponding *F*-tests yielded the following results $F(2, 113) = .20$, $p =.85$, $\eta_p^2 = .003$, $BF_{10} = .42$ for the interaction between memorability and participant ethnicity, $F(1, 113) = 3.04$, $p =.08$, $\eta_p^2 = .026$, $BF_{10} = .06$ for the interaction between memorability and match, and $F(2, 113) = .2.27$, $p =.11$, $\eta_p^2 = .039$, $BF_{10} = .33$ for the three-way interaction of memorability, participant ethnicity, and match. In short, the results under the liberal scoring method are consistent with our hypothesis that the names paired with memorable faces would be recalled better than those paired with non-memorable faces and that this effect would be consistent across participants and faces of different ethnicities.

Figure 3

Face-name recall in the non-memorable and memorable conditions.



The Effect of Ethnicity Match: Did Participants Recall Names Better, if the Face Matched their own Ethnicity?

We hypothesised that trials would result in a superior face-name recall if the participant's ethnicity and ethnicity of the face matched. We also anticipated moderation of the effect of match, such that the advantage of matching faces would be stronger in the black and white groups than in the Other groups for whom the faces did not necessarily match the participants' ethnicity. Figure 4 displays the averages for face-name recall with matching and non-matching ethnicities. As predicted by the ethnicity-match hypothesis, trials with matching faces had higher face-name recall than non-matching trials (F[1, 113] = 13.53, p <.001, η_p^2 = .11, BF_{10} = 135.59). In addition, the interaction of ethnicity match and participant ethnicity was significant (F[2, 113] = 5.7, p = .005, η_p^2 = .09, BF_{10} =14.16). Importantly, however, the nature of this interaction was not what we expected as a single effect analysis revealed that face-name recall was higher in the matching condition compared to the non-matching condition only in the *black* group, t(38) = 4.90, one-sided p < .001, Cohen's d = .78, BF_{10} =

411.41. Both the white and the other group showed no significant advantage for the matching faces, with $t(38) = 1.258$, one-sided $p = .108$, Cohen's $d = .20$, and $BF_{10} = .13$ for the white group and $t(37) = .182$, one-sided $p = .428$, Cohen's $d = .03$, and $BF_{10} = .11$ for the other group. Figure 5 displays the interaction of match and Participant ethnicity.

To summarise, the data under the liberal scoring method did not entirely support our hypothesis that there would be superior recall for matching faces in the *white* and black groups, and a smaller advantage for matching faces in the *Other* group. In contrast, the data yielded a statistically significant advantage of match *only* in the *black* group.

Figure 4

Face-name recall in the non-matching and matching conditions (liberal scoring method).

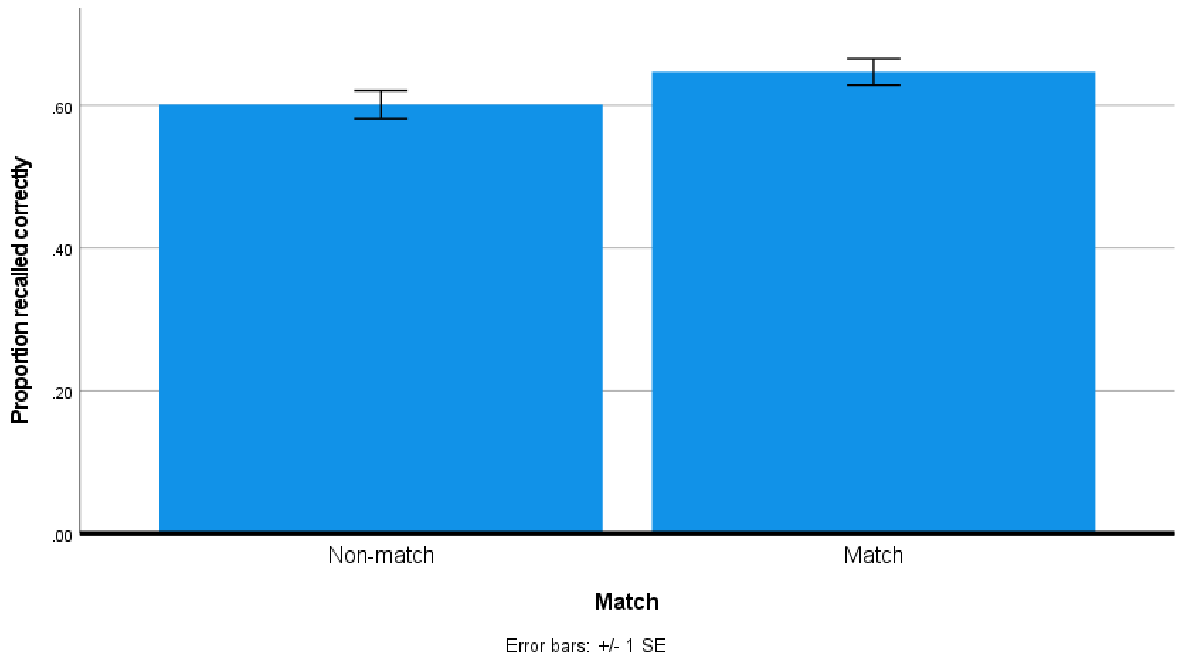
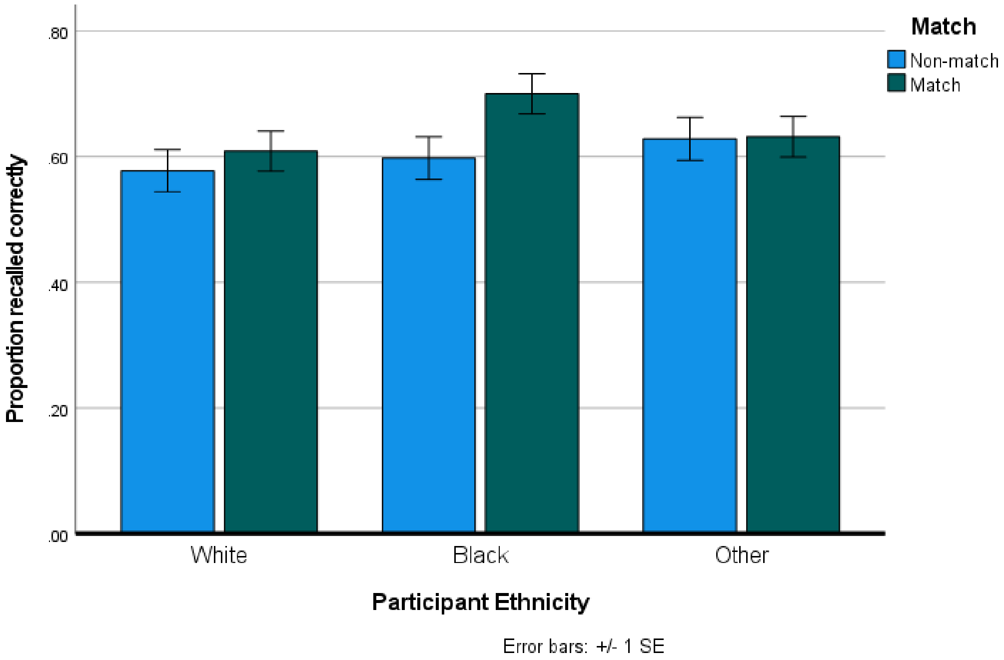


Figure 5

Face-name recall in the matching and the non-matching condition for each ethnicity group


Match
Non-match
Match
.80
.60
.40
.20
.00
Proportion recalled correctly
White
Black
Other
Participant Ethnicity
Error bars: +/- 1 SE

Recall Accuracy Strictly Scored

Assumption Checks and Descriptive Statistics

As for the liberal scoring method, we detected no meaningful violation of the RM-ANOVA assumptions. Assumption checks for normality are included in Figure A1.

The Effect of Memorability: Were Names Paired with Memorable Faces Recalled Better?

Memorable faces resulted in higher face-name recall than non-memorable faces (F[1, 113] = 33.14, p <.001, η_p^2 = .227, BF_{10} = 1.53e +6). However, contrary to the liberally scored accuracy data, the analysis for the strictly-scored accuracy values showed that memorability was involved in several statistically significant interactions. Specifically, memorability had a statistically significant interaction with match (F[1, 113] = 4.654, p = .033, η_p^2 = .040), due to a seemingly larger effect of memorability for matching trials (Figure 6), which was not corroborated by a Bayesian analysis which showed anecdotal evidence in favour of the null hypothesis for this interaction (BF_{10} = .78). Nevertheless, we conducted a simple effects analysis to determine whether the effect of memorability was significant for the matching and non-matching conditions. The results of these analyses showed that memorable faces were recalled better in both conditions with t(115) = 3.4 p < .001, Cohen's d = .31, BF_{10} = 81.40 for non-matching faces and t(115) = 5.1 p < .001, Cohen's d = .48, BF_{10} = 5442.69 for matching faces. Those results confirmed that memorability leads to superior face-name recall, although this effect was greater for matching faces.

Finally, also the three-way interaction of memorability, match, and participant ethnicity was statistically significant according to its p-value, but with only anecdotal evidence according to the Bayes factor (F[2, 113] = 4.733, p =.011, η_p^2 = .077, BF_{10} = 2.27). The results for follow-up tests comparing recall with memorable and non-memorable faces in all combinations of match and participant ethnicity are displayed in Table 5. All conditions with memorable faces had higher face-name recall than their non-memorable counterparts. Still, the size of effect of memorability differed widely, from Cohen's ds of .15 to .91 (Table 5, Figure 7). Also, the effect of memorability only reached statistical significance in some of the conditions: In the non-matching trials in the white group and the matching trials in the black group, we failed to find the hypothesised effect according to the p-values, with the Bayes factors yielding anecdotal evidence for the presence or absence of these effects.

To summarise, we did not find the hypothesised effect of memorability in all combinations of the factors match and participant ethnicity under the strict scoring method. Among white participants, we only found a significant advantage for memorable faces in the matching condition. The black group only showed the advantage of memorability in the *non-*

matching condition. Only in the Other group we found a significant advantage of memorable faces in both conditions of match.

Table 5.

Results of paired samples t*-tests for the lower recall of non-memorable faces, in all combinations of the factors match and participant ethnicity.*

Condition	*t*	df	Cohen's *d*	One-sided *p*	BF_{10}
White					
Non-match	.95	38	.15	.170	1.03
Match	5.71	38	.91	< .001	959.78
Black					
Non-match	2.90	38	.46	.003	6.36
Match	1.38	38	.22	.088	.82
Other					
Non-match	1.84	37	.30	.037	1.72
Match	2,67	37	.43	.006	115.10

Figure 6

The effect of memorable on face-name recall in the matching and non-matching conditions

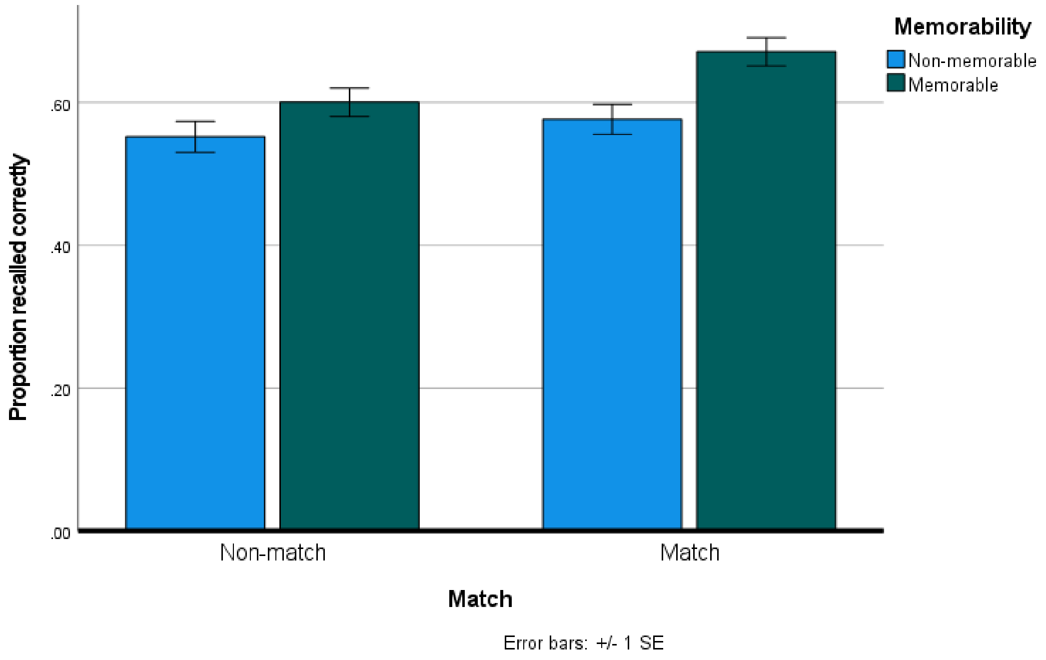
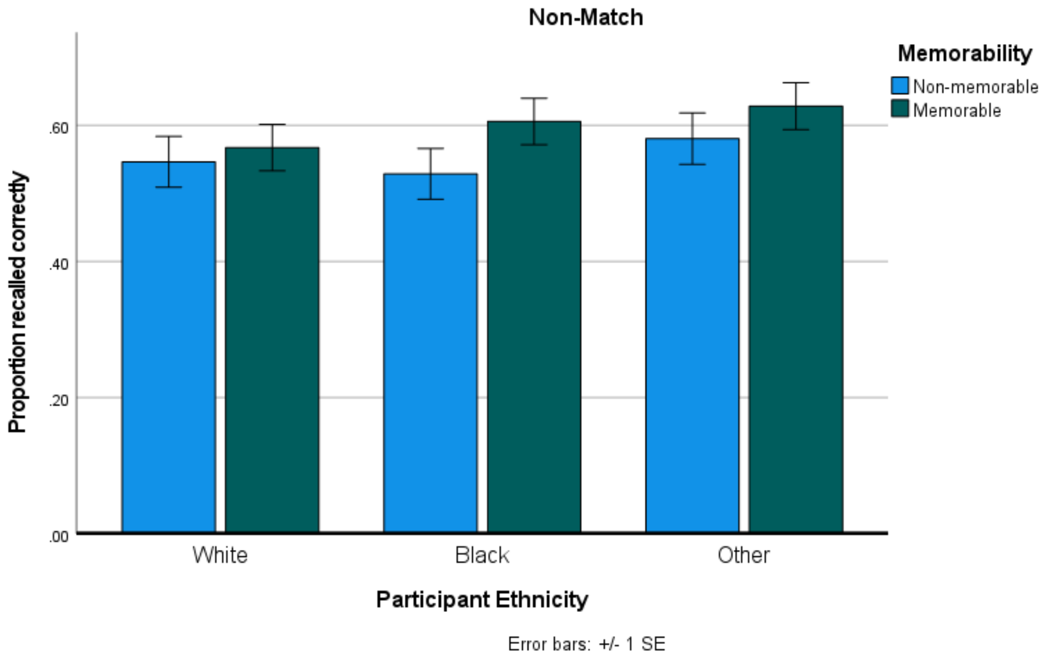


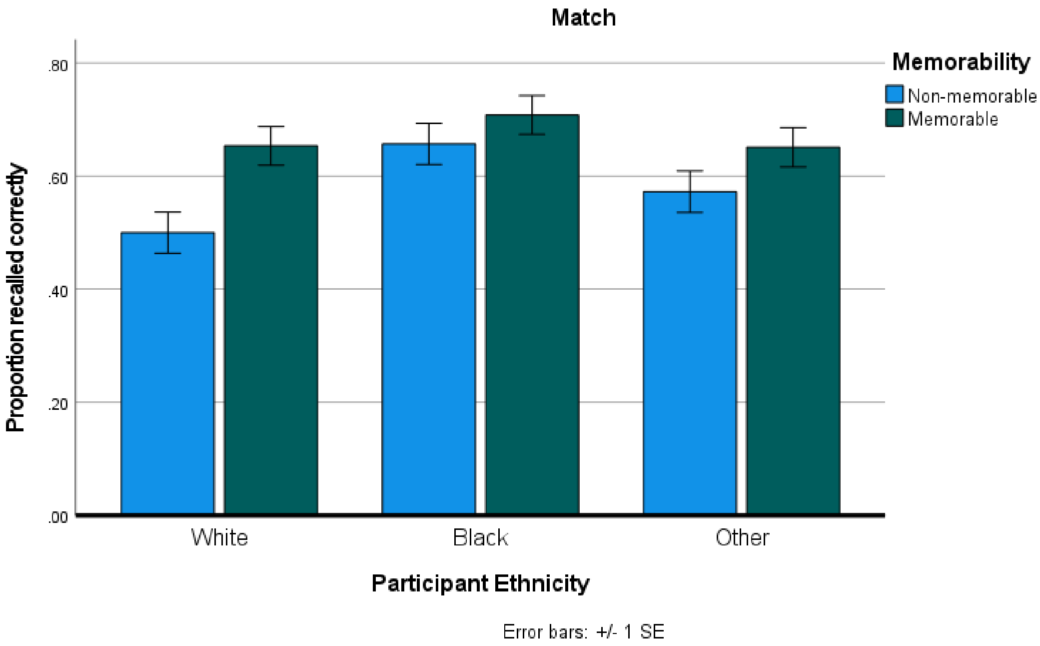
Figure 7

The effect of memorability on face-name recall in the matching conditions (Figure 7A) and in the non-matching conditions (Figure 7B).

A



B



The Effect of Match: Did participants recall names better if the face matched their own ethnicity?

We hypothesised that trials would result in superior face-name recall if the participant's ethnicity and ethnicity of the face matched. This effect was expected to be more pronounced in the white and black groups than in the Other group. Under the liberal scoring method, the results showed a pattern different from these hypotheses, as we only found a statistically significant advantage for matching trials in the black group. Under the strict scoring method, the main effect of match also turned out significant (F[1, 113] = 14.66, p <.001, η_p^2 = .115, BF_{10} = 162.94). As hypothesised, the two-way interaction of match and participant ethnicity was significant (F[1, 113] = 7.538, p <.001, η_p^2 = .115, BF_{10} = 59.92). Table 6 displays the simple effects of match per participant ethnicity: The names of matching faces were recalled significantly better only in the black group. In the black group, the alternative hypothesis also received extremely strong support in the Bayesian analysis. In both the white and the Other group, the effect of race was not significant and the BF_{10} favoured the null hypothesis.

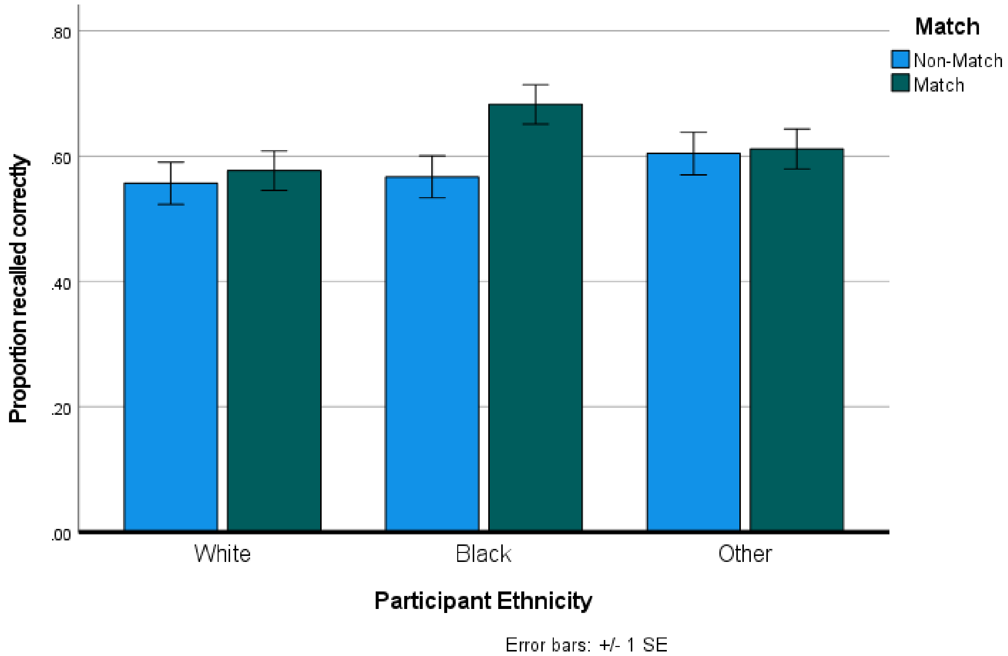
Table 6.

Results of paired samples t-*tests for the superior recall of matching faces, for each category of race*

Condition	*t*	df	Cohen's *d*	One-sided *p*	BF_{10}
White			[illegible]		
Black			[illegible]		[illegible]
Other			[illegible]		

Figure 7

Face-name recall match and non-match for each ethnicity group



Discussion

The first purpose of this study was to replicate the result that face memorability enhances face-name memory (Van der Wal and Nieuwenstein, 2021). In an ethnically inclusive design, we provided an indirect test of whether the face memorability scores of Bainbridge are universal or contaminated by the Own-Race-Bias (ORB). We hypothesised that face-name associations with memorable faces would be remembered better than those with non-memorable faces. We also hypothesised the presence of an Own-race Bias: Trials in which the ethnicity of the participant and the ethnicity of the face matched would show higher face-name recall. To examine these hypotheses, we used a $2 \times 2 \times 3$ design with participant ethnicity (between: white, black, Other) face memorability (within: memorable, non-memorable) and match of the ethnicity of the participant and face (within: match, non-match). Face-name recall was measured with an adapted version of the face-name association (FNA) task of Strickland-Hughes et al (2020). In this task, participants first were presented with a face visually and a name acoustically. Later, they were asked to recall the name upon seeing the face. Our analyses displayed some differences between a strict and a liberal scoring method for recall accuracy, as well as between the frequentist and the Bayesian analysis.

As there was some disagreement between analysis and scoring methods, we decided to base our conclusion on the liberal scoring method and the Bayesian analysis. The liberal method possesses more construct validity for the concept of *name recall* because it also scores a recall response as correct when the response resulted from *encoding* a slightly incorrect name, which was then *correctly recalled.* The phenomenon that acoustically presented names are encoded as *different* yet phonetically *similar* versions has influenced the measurement instruments in several studies. For example, Zhao et al (2012) used a recognition test, Strickland-Hughes et al. (2020) applied a liberal scoring method of freely recalled names. Our liberal scoring method is preferred, as it both avoids scoring correct recalls as incorrect while still purely measuring name recall, not recognition. The Bayesian analysis is preferred because its results generally are more conservative, creating less false positives (Wetzels et al., 2011).

Research Results

Our results showed that names paired with memorable faces were recalled better than non-memorable faces. Further, memorability scores for faces seemed universal, as the Bayesian analysis showed extremely strong evidence for the enhancing effect of face memorability on name recall, but only anecdotal evidence for its interactions. In contrast, the effect of Match did not conform to the predicted pattern. Across scoring methods and analysis methods, only the black group showed superior recall for matching faces, while the white group and the Other group did not: In the white and the Other groups, the absence of an effect of match was more likely than its presence.

Study Limitations

We found that the memorability of a face transferred to the associated name, this effect was discovered in van der Wal and Nieuwenstein (2021), and holds in our ethnically inclusive design. Yet, our study was only maximally inclusive within the limits of the face-image data base of Bainbridge (2013). All participants were from predominantly white societies: the United States of America, the United Kingdom, and Canada. Non-white individuals in predominantly white societies learn to remember the faces of white people in daily life, with more interracial contact correlating with lower ORB, as a South-African-Australian study showed (Sadozai et al., 2019). A maximally inclusive replication of van der Wal and Nieuwenstein's (2021) study could have used participants who grew up in not predominantly white countries, like China, Botswana, or Indonesia.

Although this is a valid limitation of our study, we nevertheless suggest that Bainbridge et al.'s (2013) face memorability scores seem universal, and that this memorability can be transferred to an associated name. We draw this conclusion because, thus far, there has not been any indication that the memorability scores by Bainbridge are *not* universal. Also, our data demonstrated strong evidence for a beneficial effect of memorability in *all* ethnic groups sampled.

Another limitation of our study are some significant differences in some attributes when comparing the full set of memorable and non-memorable face for common familiar, forgettable, and interesting. Some of these attributes were significant predictors of memorability (Bainbridge et al., 2013), which we attempted to control for. In other words, we encountered a Simpson's paradox here. The memorable and non-memorable faces we used were successfully equalised on all attributes from Bainbridge et al. (2013), within each of the three race categories. When all combined, however, there were statistically significant differences on some of these attributes between the combined memorable and the non-memorable subset (Appendix 1, Table A1). Nevertheless, we decided that those significant *t*-tests represented no potential confounds for two reasons.

Firstly, when calculating the ANOVA for memorability, the difference between the memorable and non-memorable trials is calculated for each *participant/race combination separately*. Yet, we only found significant differences when we compared *all* memorable faces with *all* non-memorable faces *combined*. Our analysis was conducted in each group separately, both due to the nature of an ANOVA calculation and for the simple effect analysis. On the level of each subset, we successfully controlled for all potential confounds for each race separately. None of the potentials confounds in the white, black, and mixed-race data sets were significant. As the differences in face attributes are only significant for all subsets *combined*, and our analysis was conducted on the level of each *subset*, the Simpson's Paradox should *not* affect our results. Further, we detected no interactions between the effect of memorability and ethnicity group, as we would expect if the differences in attributes would matter.

Secondly, we detected the potential confounds with 52 individual *t*-tests, namely one for each attribute in the dataset collected by Bainbridge et al. (2013). Such a large number of *t*-tests tends to produce coincidental findings (Weisstein, 2004). Hence, we utilised the Bonferroni correction (Weisstein, 2004) to correct the threshold for statistical significance. As the 52 attributes represented 26 distinct constructs, we divided the conventional *alpha* by 26. This resulted in 0.05/26 = 0.00193 as the corrected significance threshold. With the corrected alpha, none of the potential confounds remained significant either.

Finally, the Own-Race Bias (ORB) did not follow the predicted pattern, as we only found the effect of match in the black group. This is surprising, as the Own-race bias is a very well-replicated phenomenon (Meissner & Brigham, 2001). A lack of power is an unlikely explanation missing the effect of match in the white group. With the average effect size of match in the white and the black group combined, we possessed a power of .94 (.93 for the strict scoring method). Further, the observed power for the interaction of participant ethnicity and match is.94 (.85 for the strict scoring method). Also the Bayesian analysis favoured the null hypothesis for the white group. Both the frequentist power analysis and the Bayesian methods, which are less susceptible to error due to a lack of power, agree. We conclude that the surprising pattern of the ORB in our study is unlikely to result from a lack of power. A more plausible explanation for our failure to replicate is that, either the ORB did does not affect associated information in general or the ORB may affect associated information, but this effect has been buffered be confounding effects specific for our study.

Implications of the Present Study and Future Research

Implications for Memorability Research

The memorability of a face enhances the probability that the associated name is recalled. Beyond this everyday implication of our study, we have demonstrated that the memorability of an image transferred to associated information. Further, it would be interesting to test under which conditions memorability transfers to another simultaneously presented stimulus, as those findings could be applied for marketing purposes. To apply image memorability in advertising, we would like to predict the influence of potential moderating factors using theory.

Pattern completion is an established theory on memory storage, and may predict under which circumstances image memorability will enhance the recall of associated information (Liu et al., 2016). Pattern completion explains how the engram of a memory is recalled from a memory cue. The *engram* of a memory is thought to comprise a pattern of neural activation that includes all encoded visual, acoustic, and other sensory experiences that accompanied an event, together with the representations of the relevant abstract concepts and a "plot" describing how these parts related to each other. According to pattern completion, when one subpart of the network is activated by cue, then all the other recorded concepts are activated as well, so that the whole memory is reinstated in the brain (a pattern is completed). The stronger and larger the number of connections between the cue and the rest of the pattern, the more likely the cue is to trigger a full recollection. As memorability represents encoding strength of a memory, a memorable face likely also forms stronger connections with associated stimuli than a forgettable face. For instance, imagine an advertising poster for a toothbrush: One sees a picture of a beautiful woman smiling with shining white teeth, holding a toothbrush, with the brand name "Bloom Brush" next to it (the pattern). Next time one is searching for a new toothbrush, this small subpart (toothbrush), can activate the rest of the pattern: The smiling woman with white teeth and the brand name. Concepts more central in the network for this particular memory are more likely to trigger a full retrieval: The concept of toothbrush relates closely to the white teeth and the brand name. The concept of "hair" would also be part of the network (because the model probably had some hair on her head), but the concept "hair" only has very little importance in the network and has little cross-connections with other concepts in the network, such as the brush.

Based on the pattern completion network, we can derive the following prediction about when the memorability of an image transfers to associated information. Memorability of a face will transfer best to a brand name, the more central the name and the face are in the network representing the memory, and the more direct associations are formed between the name and face. For example, memorability would transfer best if a memorable face *speaks* the brand name, or interacts with it in another way. This hypothesis could be tested in future research.

Implications for Own-Race Bias Research

As stated above, the surprising pattern of the ORB in our study is unlikely to result from a lack of power, but rather from confounding moderators of the ORB specific for our study. The ORB for facial recognition and eyewitness memory is a well-established phenomenon, but the ORB for associated information is a much younger area of research. Two studies found that the recall of associated information was enhanced for same-race faces (Knuycky et al., 2014; Murphy et al., 2022). The two studies that investigated the effect of the ORB on face-name associations found that the ORB enhanced name-recall for the in-group face-name associations (Ramon et al., 2016). In another study, participants required less trials to memorise a face-name association, the face and participant ethnicity matched (Hayward et al., 2017).

Some known moderators of the ORB may shed light on the ambiguous pattern of the ORB in our study: It was found that positive emotions eliminate the ORB (Johnson & Fredrickson, 2005). The FNA-task consisted of various people introducing themselves to the participant, which might have triggered associations of meeting new people at a social event. Social events elicit far more positive emotions than the police suspect lineups do, which are typically used in the study of the ORB (Meissner & Brigham, 2001).

Another hidden moderator of the ORB might have been the lack of semantic information that was conveyed by the name. As stated in our introduction, names are difficult to remember, because they convey nearly no semantic information (Cohen, 1990). Yet, the ORB is more pronounced for semantic information and especially if this semantic information follows a (potentially racist) stereotype (Knuycky et al., 2014). Knuycky et al., using white participants, found that the ORB enhanced recall of the profession of a black person better, if this person was labelled a "drug dealer" instead of a "professor" (2014). In the associated information in our study we controlled for semantic cues, stereotypes, or contextual information, possibly reducing the own-race bias.

Conclusion

We aimed to replicate the finding of Van der Wal and Nieuwenstein (2021) that face memorability transfers to an associated name, also in a maximally inclusive design. In this regard, we succeeded. Surprisingly, our data displayed ambiguous results regarding the Own-Race Bias. Future research should focus on the transfer of image memorability to associated information, possibly conducted within a Pattern Completion framework.

References

Bainbridge, W. A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Scientific Reports*, *8*(1), 1-11.

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142*(4), 1323.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*(2), 307.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305-327.

Burton, A. M., & Bruce, V. (1992). I recognise your face but I can't remember your name: A simple explanation? *British Journal of Psychology, 83*(1), 45-60.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 48*(4), 879-894.

Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology, 81*(3), 287-297.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Harvey, A. J. (2014). Some effects of alcohol and eye movements on cross-race face learning. *Memory, 22*(8), 1126-1138.

Hayward, W. G., Favelle, S. K., Oxner, M., Chu, M. H., & Lam, S. M. (2017). The other-race effect in face learning: Using naturalistic images to investigate face ethnicity effects in a learning paradigm. *Quarterly Journal of Experimental Psychology*, *70*(5), 890-896.

Hills, P. J., & Pake, J. M. (2013). Eye-tracking the own-race bias in face recognition: Revealing the perceptual and socio-cognitive mechanisms. *Cognition, 129*(3), 586-597.

Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). Understanding the intrinsic memorability of images. *Advances in Neural Information Processing Systems, 24*

JASP Team (2022). JASP (Version 0.16.3)[Computer software].

Johnson, K. J., & Fredrickson, B. L. (2005). "We all look the same to me" Positive emotions eliminate the own-race bias in face recognition. *Psychological Science, 16*(11), 875-881.

Knuycky, L. R., Kleider, H. M., & Cavrak, S. E. (2014). Line-up misidentifications: When being 'prototypically black'is perceived as criminal. *Applied Cognitive Psychology, 28*(1), 39-46.

Liu, K. Y., Gould, R. L., Coulson, M. C., Ward, E. V., & Howard, R. J. (2016). Tests of pattern separation and pattern completion in humans—A systematic review. *Hippocampus, 26*(6), 705-717.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3.

Murphy, D. H., Silaj, K. M., Schwartz, S. T., Rhodes, M. G., & Castel, A. D. (2022). An own-race bias in the categorisation and recall of associative information. *Memory, 30*(2), 190-205.

Naveh-Benjamin, M., Guez, J., Kilb, A., & Reedy, S. (2004). The associative memory deficit of older adults: further support using face-name associations. *Psychology and Aging, 19*(3), 541.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453.

Online Microsoft Sam TTS Generator. (n.d.). Online Microsoft Sam TTS Generator. Retrieved April 20, 2022, from https://tetyys.com/SAPI4/

Ramon, M., Miellet, S., Dzieciol, A. M., Konrad, B. N., Dresler, M., & Caldara, R. (2016). Super-memorisers are not super-recognisers. *PLoS One*, *11*(3), e0150972.

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological Bulletin, 138*(1), 146.

Sadozai, A. K., Kempen, K., Tredoux, C., & Robbins, R. A. (2019). Can we look past people's race? The effect of combining race and a non-racial group affiliation on holistic processing. *Quarterly Journal of Experimental Psychology*, *72*(3), 557-569.

Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in Whites. *Basic and Applied Social Psychology*, *22*(2), 71-84.

Strickland-Hughes, C. M., Dillon, K. E., West, R. L., & Ebner, N. C. (2020). Own-age bias in face-name associations: Evidence from memory and visual attention in younger and older adults. *Cognition, 200*, 104253.

Van der Wal, D., & Nieuwenstein, M. (2021). *Het effect van de memorabiliteit van een gezicht op het onthouden van gezicht-naam associaties*

Weisstein, E. W. (2004). Bonferroni correction. *Https://Mathworld.Wolfram.Com/,*

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*(3), 291-298.

Wong, H. K., Estudillo, A. J., Stephen, I. D., & Keeble, D. R. (2021). The other-race effect and holistic processing across racial groups. *Scientific Reports, 11*(1), 1-15.

Zhao, K., Wu, Q., Shen, X., Xuan, Y., & Fu, X. (2012). I undervalue you but I need you: the dissociation of attitude and memory toward in-group members. *PLoS One, 7*(3), e32932.

Ziaei, M., Persson, J., Bonyadi, M. R., Reutens, D. C., & Ebner, N. C. (2019). Amygdala functional network during recognition of own-age vs. other-age faces in younger and older adults. *Neuropsychologia, 129*, 10-20.

Appendix

Table A1

Means and SD for attributes between the memorable and non-memorable conditions In the white (Table A1A), the black (Table A1B), the Other group (Table A1C), and the total selection of faces (Table A1D)

A

Attribute	M(SD)		t_{diff}
	Non-memorable	Memorable	
Common	3.19(0.23)	3.00(0.3)	1.40
Familiar	4.92(0.49)	4.71(0.60)	.80
Forgettable	4.94(0.43)	4.35(0.52)	2.410*
Interesting	5.09(0.39)	5.40(0.47)	-1.4

Note. *<.05, two-tailed. **p<.01, two-tailed. ***p<.001, two-tailed.

B

Attribute	M(SD)		t_{diff}
	Non-memorable	Memorable	
Common	3.14(0.28)	3.01(0.28)	0,91
Familiar	5.05(0.76)	4.60(0.46)	0.84
Forgettable	4.90(0.67)	4.60(0.72)	0.79
Interesting	4.96(0.34)	5.30(0.92)	-0.98

Note. *<.05, two-tailed. **p<.01, two-tailed. ***p<.001, two-tailed.

C

Attribute	M(SD)		t_{diff}
	Non-memorable	Memorable	

Common	3.05(0.27)	2.74(0.44)	1.70
Familiar	5.06(0.42)	4.26(0.40)	3.88**
Forgettable	4.88(0.47)	4.20(0.65)	2.42*
Interesting	5.25(0.42)	5.60(0.42)	-1.58

Note. *<.05, two-tailed. **p<.01, two-tailed. ***p<.001, two-tailed.

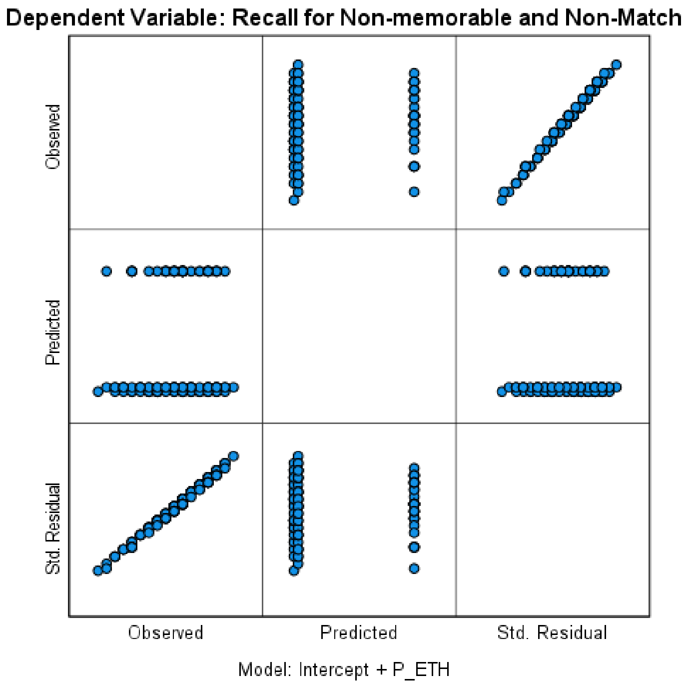
D

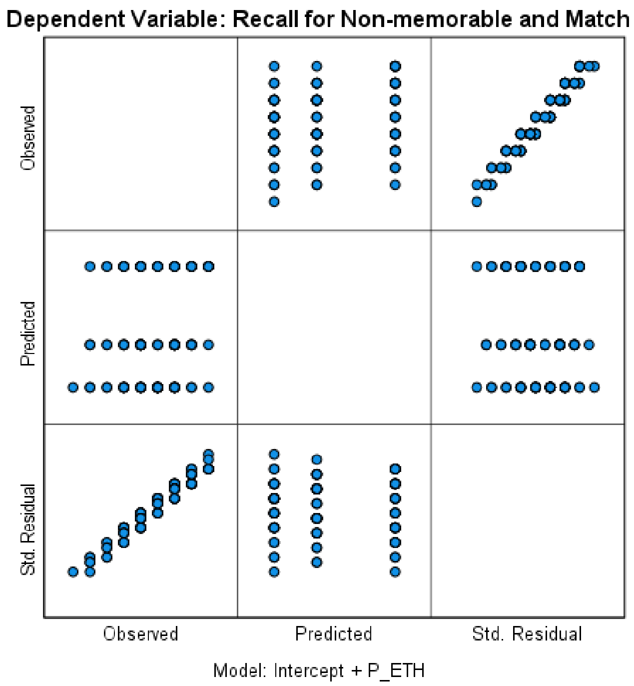
Attribute	M(SD)		t_{diff}
	Non-memorable	Memorable	
Common	3.12 (0.26)	2.92 (0.36)	2.31*
Familiar	5.01(0.55)	4.59(0.52)	2.74**
Forgettable	4.89(0.51)	4.38(0.63)	3.10**
Interesting	5.1(0.39)	5.4(0.62)	-2.18*

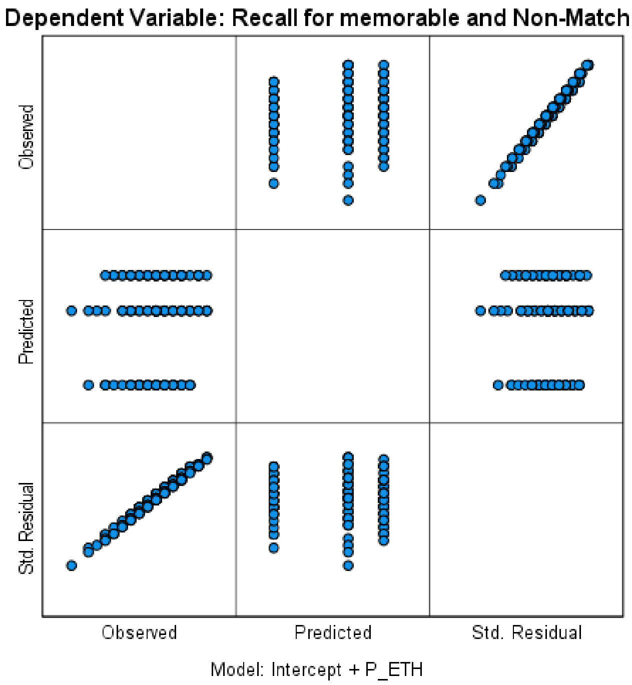
Note. *<.05, two-tailed. **p<.01, two-tailed. ***p<.001, two-tailed.

Figure A1

The residual plots for each combination of the within-group variables (liberal scoring method)


Dependent Variable: Recall for Non-memorable and Non-Match
Observed
Predicted
Std. Residual
Observed
Predicted
Std. Residual
Model: Intercept + P_ETH


Dependent Variable: Recall for Non-memorable and Match
Observed
Predicted
Std. Residual
Observed
Predicted
Std. Residual
Model: Intercept + P_ETH


Dependent Variable: Recall for memorable and Non-Match
Observed
Predicted
Std. Residual
Observed
Predicted
Std. Residual
Model: Intercept + P_ETH

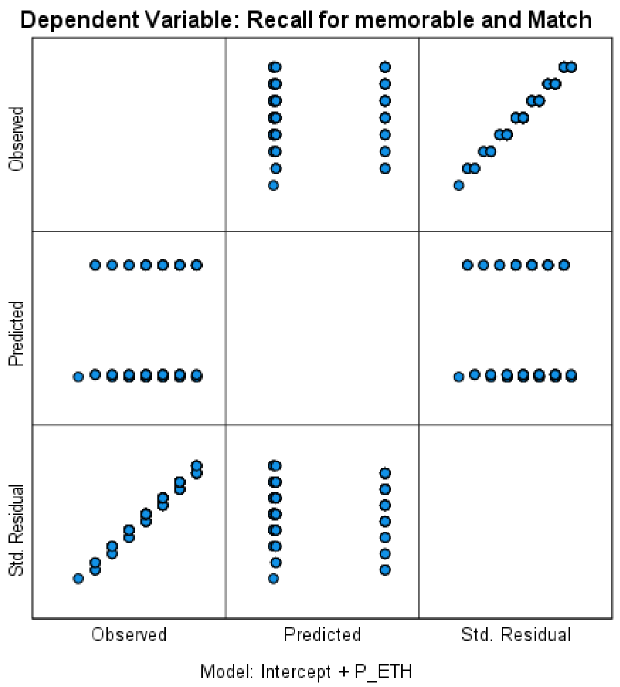
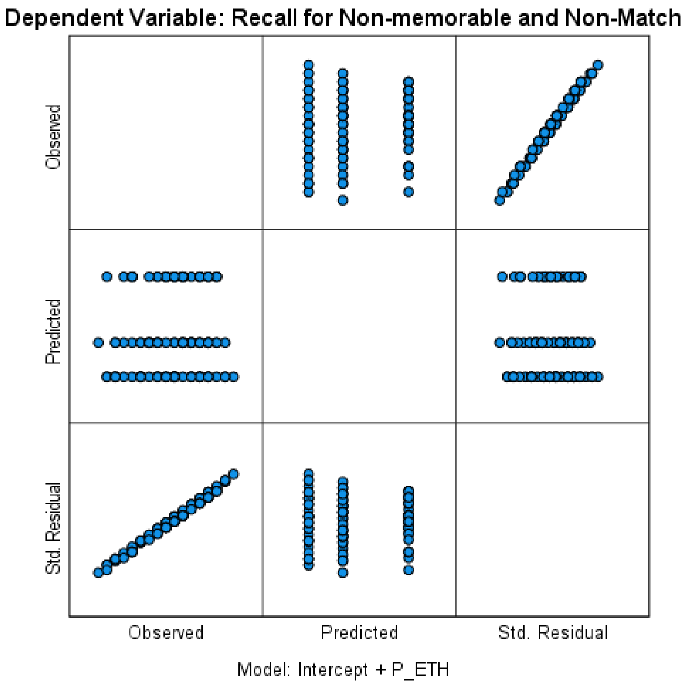
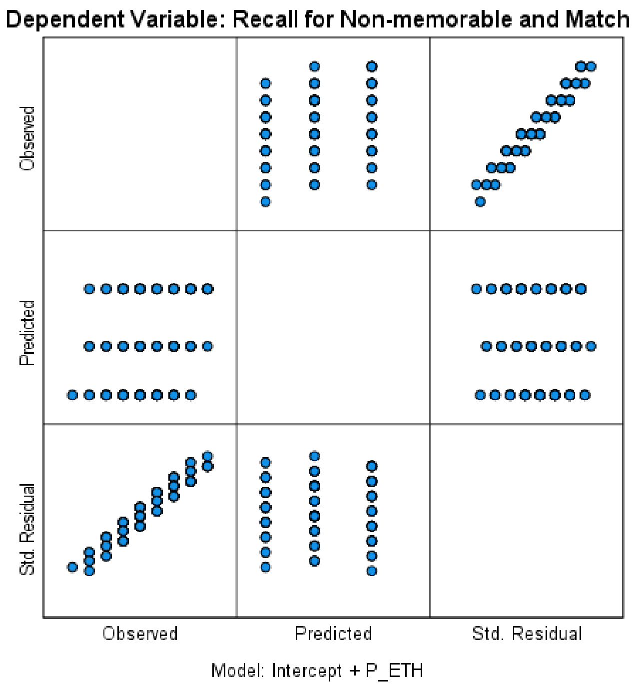
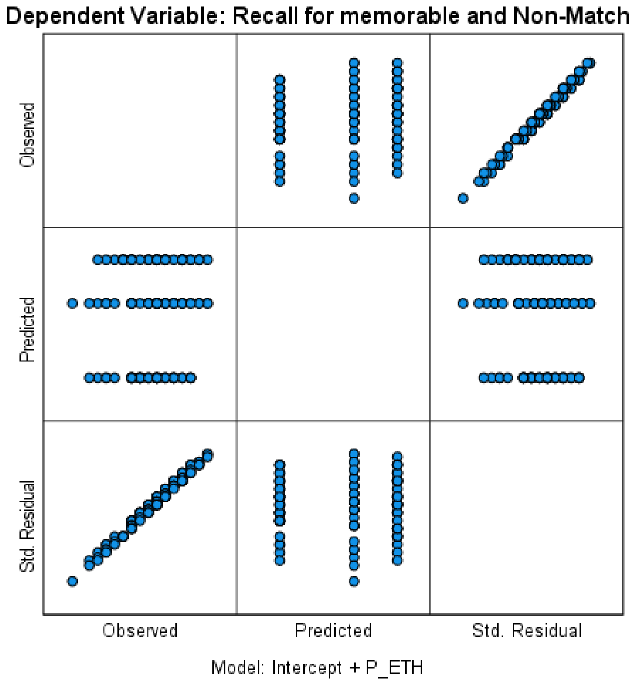

Dependent Variable: Recall for memorable and Match
Observed
Predicted
Std. Residual
Observed
Predicted
Std. Residual
Model: Intercept + P_ETH

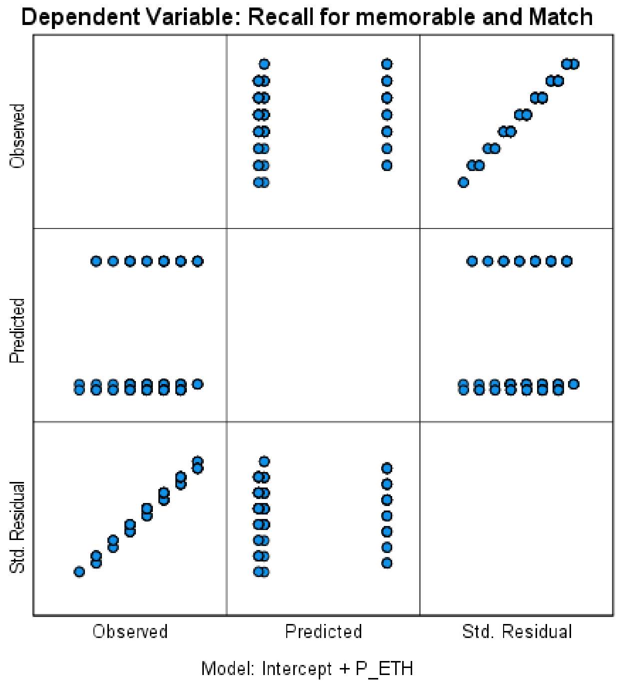
Figure A2

The residual plots for each combination of the within-group variables (strict scoring method)






Dependent Variable: Recall for memorable and Non-Match
Observed
Predicted
Std. Residual
Observed
Predicted
Std. Residual
Model: Intercept + P_ETH


Dependent Variable: Recall for memorable and Match
Observed
Predicted
Std. Residual
Observed
Predicted
Std. Residual
Model: Intercept + P_ETH