Psychonomic Bulletin & Review (2022) 29:1387–1396
https://doi.org/10.3758/s13423-022-02080-6

BRIEF REPORT

Can intentional forgetting reduce the cross-race effect in memory?

Huiyu Ding[1] · **Jonathon Whitlock**[1] · **Lili Sahakyan**[1,2]

Accepted: 4 March 2022 / Published online: 4 April 2022

Abstract
Across three studies, we utilized an item-method directed forgetting (DF) procedure with faces of different races to investigate the magnitude of intentional forgetting of own-race versus other-race faces. All three experiments shared the same procedure but differed in the number of faces presented. Participants were presented with own-race and other-race faces, each followed by a remember or forget memory instruction, and subsequently received a recognition test for all studied faces. We obtained a robust cross-race effect (CRE) but did not find a DF effect in Experiment 1. Experiments 2 and 3 used shorter study and test lists and obtained a significant DF effect along with significant CRE, but no interaction between face type and memory instruction. The results suggest that own-race and other-race faces are equally susceptible to DF. The results are discussed in terms of the theoretical explanations for CRE and their implications for DF.

Keywords Directed forgetting · Cross-race effect · Face recognition

Introduction

In our daily life, the ability to intentionally forget outdated, incorrect, or distressing information serves a critical role. Often, we need to engage in processes and behaviors that limit access to unwanted memories, known as *intentional forgetting* (e.g., Bjork et al., 1968). Sometimes the need to forget is implicit and self-inflicted and sometimes more explicitly instructed by others. For example, in the courtroom, jurors are often asked to disregard certain information (Caretta & Moreland, 1983), and studies have shown that jurors can disregard information if they believe it is irrelevant or untrustworthy (Kassin & Sommers, 1997; Mallard & Perkins, 2005). Successful intentional forgetting has real-world implications. For example, Scully and Hupbach (2020) reported that participants could intentionally forget negative behavior of specific individuals, resulting in more positive judgments of those individuals. Sell (2016) demonstrated that participants who were instructed to forget the stories involving transgressions were more likely to later forgive the antagonist, indicating successful intentional forgetting can promote forgiveness. Furthermore, Xie et al. (2021) reported that participants who intentionally forget negative social feedback exhibited more positive emotional evaluations towards the providers of that feedback. Overall, intentional forgetting of an individual could later impact how that individual is evaluated and treated.

The current investigation involved intentional forgetting of identity of faces that differ in race. The critical question was whether intentional forgetting might reduce the *cross-race effect* (CRE) – a phenomenon referring to superior recognition of own-race than other-race faces (for reviews, see Meissner & Brigham, 2001; Young et al., 2012). Imagine you happen to be in the middle of a street fight that breaks out between members of two racial groups, your own race and a different race, that results in an own-race member injuring an other-race member. If you were questioned by police about the event, correct identification of the aggressor might vary depending on your race and the race of the aggressor, and the assumptions you made while you were witnessing the fight. For example, if you assumed that the aggressor was from your own race, it may be more advantageous to intentionally forget the members of the other-race group. Doing so may more efficiently allocate your cognitive resources by focusing on the own-race group to subsequently facilitate an accurate identification of the aggressor from innocent by-standers. Alternatively, if later investigation

✉ Huiyu Ding
hding12@illinois.edu

[1] Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St, Champaign, IL 61820, USA

[2] Beckman Institute for Advanced Science and Technology, Urbana, IL, USA

revealed that the aggressor was actually from the other-race group that you had originally intentionally downregulated, it raises the question of whether your ability to differentiate the true aggressor from the innocent bystanders would be affected by previous downregulation. The CRE literature indicates that people have difficulty differentiating other-race faces, and 36% of wrongful convictions result from cross-race misidentifications (Dwyer et al., 2003). Therefore, the impact of intentional forgetting on own-race and other-race individuals should be thoroughly investigated, as it could have important real-world implications.

Directed forgetting procedure

In the laboratory, the ability to downregulate or forget unwanted information has been studied with the use of several techniques. The directed forgetting (DF) procedure is one of the widely used techniques to study intentional forgetting (Bjork et al., 1968). In the DF procedure, items are followed by instructions to either remember (R) or forget (F) those items, and these memory instructions are presented either after an entire list has been studied (list-method) or following each item on a trial-by-trial basis (item-method). Both procedures result in impaired memory for F items compared to R items, a finding known as the *directed-forgetting effect* (for reviews, see MacLeod, 1998; Sahakyan, 2022; Sahakyan et al., 2013). In the current studies, we utilized the item-method procedure to explore intentional forgetting of own-race and other-race individuals.

The majority of DF studies used verbal stimuli, including syllables (e.g., Bjork, 1970), words (e.g., Bjork & Geiselman, 1978), and sentences (e.g., Geiselman, 1974). Other studies utilized nonverbal stimuli, such as simple line-drawings (e.g., Basden & Basden, 1996), complex visual scenes (e.g., Hauswald & Kissler, 2008; Quinlan et al., 2010), and emotional pictures (e.g., Nowicka et al., 2011). The results indicated that DF occurs with all sorts of stimuli, although non-verbal materials were found to have a smaller DF effect compared to verbal materials (e.g., Quinlan et al., 2010).

Directed forgetting of faces

A handful of studies have examined DF of faces and showed that participants can intentionally forget faces when instructed to do so (e.g., Goernert et al., 2011; Metzger, 2011; Quinlan & Taylor, 2014; Reber et al., 2002; Tay & Yang, 2017; Wang et al., 2019). Interestingly, some facial attributes were shown to moderate the degree to which faces are susceptible to DF (e.g., Metzger, 2011; Quinlan et al., 2010; Tay & Yang, 2017). For example, angry faces were found to be more resistant to DF than happy faces (Tay & Yang, 2017), although another study reported inconsistent results (Quinlan & Taylor, 2014). In addition, Metzger (2011) found that distinctive faces (i.e., faces receiving a high score on a "distinctiveness" scale) are more resistant to intentional forgetting compared to typical faces.

Overall, the limited findings of DF with faces indicate that faces are susceptible to DF, and that certain attributes might moderate the degree to which they are susceptible to DF. The focus of the current study is on the racial identity of the faces, and its potential interaction with DF. Fitzgerald et al. (2013) manipulated racial attributes of studied faces and found the magnitude of the DF effect to differ between Asian and Black faces. However, since only the other-race faces were presented and tested (i.e., no Caucasian faces were presented in the study, whereas the participants were Caucasian), direct comparison between DF of own- and other-race faces could not be assessed. Therefore, even though this study found that faces of some races were more susceptible to DF than others, the study does not address whether DF is greater or smaller for own-race than other-race faces. Our studies aimed to address the gap in the literature by directly examining DF for own-race and other-race faces.

Theoretical accounts of cross-race effect (CRE) and predictions

Predictions regarding how own-race and other-race faces might be differentially susceptible to DF could be derived from theoretical views of the CRE. Contemporary views of the CRE encompass two broad classes: social-cognitive views and perceptual expertise views. According to the social-cognitive views of the CRE, people process other-race faces as "out-group" members and subsequently cognitively "disregard" them (e.g., Rodin, 1987; Sporer, 2001). People may disregard other-race faces by allocating less processing resources to them after deeming them socially irrelevant or by processing commonalities for other-race faces at the expense of individuating information that focuses on the diagnostic differences between other-race faces (Levin, 2000; Maclin & Malpass, 2003; Meissner et al., 2005). As a result, they may form less well-encoded, weaker representations of other-race faces. Previous DF studies have examined the relationship between the strength of encoding and DF. Studies employing depth of processing manipulations indicate that more elaborate processing reduces the DF effect (Dulaney et al., 2004; Geiselman & Bagheri, 1985; Lee, 2013). Furthermore, DF is also diminished with materials that produce stronger encoding (Earles & Kersten, 2002; Hauswald & Kissler, 2008; Quinlan et al., 2010; Sahakyan & Foster, 2009). Finally, studies have also shown that DF is larger when the learned information is incongruent with the learner's perspective compared to congruent (Waldum & Sahakyan, 2012). Given that stronger encoding as well as congruent perspective can reduce the DF effect, own-race faces may be less susceptible to DF compared to other-race faces. In other words, a greater DF may be observed for other-race than own-race faces.

Alternatively, one could derive an opposite prediction from perceptual expertise views of the CRE, according to which people are deficient at effectively encoding other-race faces because of limited expertise with other-race faces (i.e., people have a biased knowledge base due to insufficient or superficial contact with other-race faces). Perceptual views suggest that people engage in a holistic processing of own-race faces that enables efficient construction of a face into a "unified" object, compared to a more piecemeal processing of other-race faces that focuses on individual facial features in isolation of one another (Michel et al., 2006; Rhodes et al., 1989; Tanaka et al., 2004). Therefore, it is possible to make an argument that a Forget cue may be easier to implement on a more unitized representation because it provides a clearer "target" for downregulation relative to a more disjointed representation. Thus, one might expect a greater magnitude of DF of own-race than other-race faces from the perceptual expertise view.

General method

We report three experiments, which shared the same experimental procedures except for the number of stimuli presented in the learning and testing phases, therefore referred to as Long, Medium, and Short Experiments. Experiment 1 (Long) presented participants with 80 faces during learning and 120 faces during testing. Experiment 2 (Medium) included 40 faces during learning and 60 during testing. Experiment 3 (short) consisted of 24 faces during learning and 36 during testing. All facial stimuli across three experiments were selected from the same database.

To preview the results, Experiment 1, which used longer lists, did not produce a significant DF effect, prompting us to shorten the lists in subsequent experiments to ensure compliance with the memory instruction.

Participants

Participants were mostly undergraduate students at the University of Illinois at Urbana-Champaign who participated for course credit. In addition, 30 volunteers from the same age group were recruited from mainland China in Experiment 1. The data collection for this study took place when the sample size determination for repeated-measures analysis of variance (ANOVA) was less well understood. The data were collected during the Fall of 2015, and Spring and Fall semesters of 2016. Our sample size was based on studies that investigated DF with another variable of interest (e.g., Metzger, 2011; Otani et al., 2012; Quinlan & Taylor, 2014). Those studies typically recruited participants ranging in number from 20 to 40. We aimed to collect more than what previous studies had used and as much data as we could obtain within a semester. The sensitivity analysis using G*power 3.1.9.6 (Faul et al., 2009) indicated that with sample sizes collected in the current study, 80% power and $\alpha = .05$, we could detect a small-to-medium effect for all three experiments ($f = 0.20$, $f = 0.18$, and $f = 0.18$, respectively).

All participants were tested individually, in the lab, before the COVID-19 pandemic. The studies were approved by the Institutional Review Board of the University of Illinois. To be eligible for the studies, student participants from UIUC had to self-identify themselves as Caucasian or Asian in the departmental mass-screening that took place at the start of each semester. Students who identified themselves as international students from Asia, who were in their freshman or sophomore academic year, and who had stayed in the USA for less than 2 years were regarded as Asian participants. Students who identified themselves as White in the pre-screening or questionnaire were regarded as Caucasian participants.

A total of 51 participants took part in Experiment 1, which included 30 mainland Chinese participants with a mean age of 20 years. In Experiment 2, a total of 65 students participated in the study, which consisted of 17 Asian and 48 Caucasian participants. In Experiment 3, a total of 60 students participated in the study, which consisted of eight Asian and 52 Caucasian participants.

Materials

The stimuli consisted of human faces in black and white, from the neck up, all with neutral facial expressions (Fig. 1). The stimuli were selected from the database used in Tullis et al. (2014). The complete dataset used in the current study includes 120 faces, consisting of 30 Asian male, 30 Asian female, 30 Caucasian male, and 30 Caucasian female faces. Each face had an equal chance of being assigned a Remember or Forget instruction in the learning phase, as well as an equal chance of being targets or lures in the testing phase. Studies were programmed in E-prime 2.0.

Procedures

All three experiments consisted of a learning phase, immediately followed by a testing phase after participants were informed of the testing instruction. The learning phase involved presenting faces, each followed by a Remember or Forget instruction. Participants were told that they would only be tested on faces followed by a Remember instruction, whereas faces followed by a Forget instruction would not, and therefore they should try to forget the faces followed by Forget instructions.

Fig. 1 Examples of stimuli presented during the learning and testing phase in the experiments

In the learning phase, participants first viewed a fixation cross for 1 s, followed by an image of a face for 6 s, which was then followed by either a Remember or a Forget instruction for 2 s. After learning all of the faces, they immediately proceeded to the test phase. At test, the Forget instruction was "canceled," wherein participants were told to indicate if they remember seeing the presented face, regardless of whether it was previously assigned a Forget or a Remember instruction. The test list randomly intermixed old and new faces for each participant, who were instructed to indicate whether they had seen the face during the learning phase using a 4-point scale that combined old/new responses with confidence judgements, where 1 = sure OLD, 2 = maybe OLD, 3 = maybe NEW, and 4 = sure NEW.

Analytic plan

We combined "sure old" and "maybe old" responses that were given to studied items to obtain the hit rate and the same category of responses given to unstudied items to obtain the false alarm rate. The raw hit rates and false alarm rates across the experiments, memory instruction, and face type are reported in Table 1. For each level of race (own-race and other-race), we calculated recognition accuracy (*d*'), which is a typical measure in item-method DF studies (e.g., Metzger, 2011; Quinlan & Taylor, 2014). Hits and false alarms were transformed using a loglinear correction before the calculation of *d*' (Hautus, 1995; Stanislaw & Todorov, 1999). In each Experiment, *d'* was analyzed with repeated-measures ANOVA, using Instruction (R vs. F) and Face Type (other-race vs. own-race) as within-subject factors. The results are summarized in Fig. 2. In addition to *d'*, we also analyzed the results by fitting receiver-operating characteristic (ROC) curves specific to each participant and then computing the area under the curve (AUC). The AUC measure accounts for both the hit rates and false-alarm rates across the confidence levels and it is an established measure of memory sensitivity (e.g., Macmillan & Creelman, 2005; Yonelinas & Parks, 2007). It is also an established measure in item-method DF studies (e.g., Chiu et al., 2021; Wang et al., 2019). AUC was computed for each participant in each condition, and then AUC across participants was tested with repeated-measures ANOVA, using memory Instruction (R vs. F) and Face Type (other-race vs. own -race) as within-subject factors. The results are summarized in Fig. 3.

Results

The data from two participants in Experiment 1 and one participant in Experiment 3 were excluded from the analyses because of the technical errors that resulted in corrupted/incomplete data files. To preview our results, we did not find any differences between Asian and Caucasian

Table 1 Untransformed hits and false alarms across Instruction, Face Type, and Experiment

	Experiment 1 (long lists)		Experiment 2 (medium lists)		Experiment 3 (short lists)	
	own-race	other-race	own-race	other-race	own-race	other-race
Remember hits	.70 (.16)	.67 (.11)	.78 (.17)	.75 (.17)	.87 (.16)	.82 (.16)
Forget hits	.70 (.12)	.65 (.15)	.74 (.17)	.70 (.16)	.79 (.17)	.77 (.17)
False alarms	.16 (.14)	.23 (.15)	.09 (.12)	.20 (.15)	.09 (.16)	.22 (.23)

Values are given as means and standard deviations (in parenthesis)

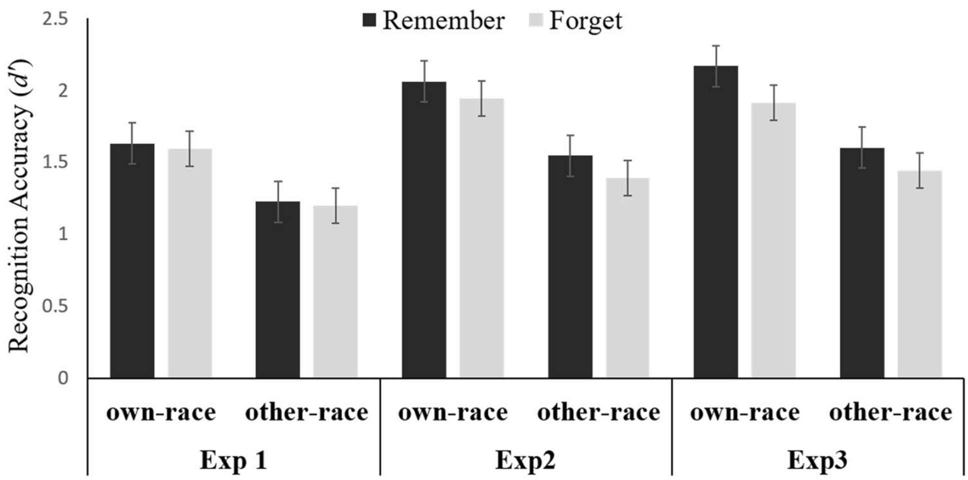

Fig. 2 Recognition accuracy (d') across Instruction, Face Type, and Experiment. The error bars reflect 95% confidence intervals

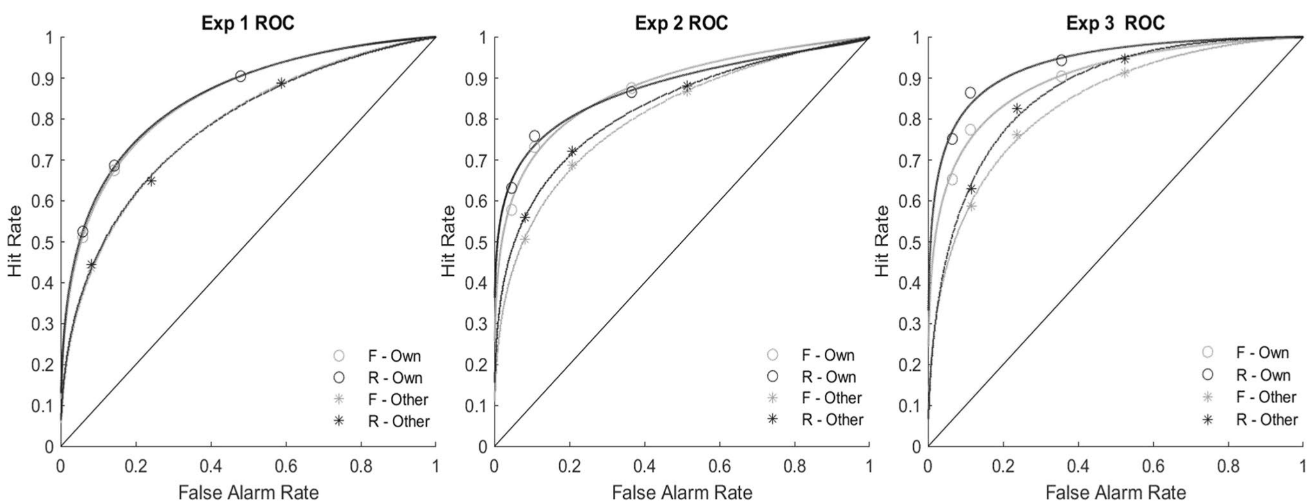

Fig. 3 Receiver operating characteristic (ROC) curves by Instruction and Face Type across Experiments 1, 2, and 3

participants in terms of overall memory performance, DF effect, and CRE.[1]

Experiment 1: Long lists

The analyses of *d'* revealed a significant main effect of Face Type, $F(1, 48) = 33.35$, $MSE = .24$, $p < .001$, indicating better recognition accuracy for own-race ($M = 1.61$, $SD = .59$) than other-race faces ($M = 1.21$, $SD = .44$), confirming the CRE effect, $d_z = 0.83$. However, there was no significant main effect of Instruction, $F(1, 48) = .45$, $MSE = .11$, $p = .504$, indicating that there was no significant DF effect in Experiment 1, $d_z = 0.10$ (Remember: $M = 1.43$, SD = .41; Forget: $M = 1.40$, SD = .46). There was also no interaction of Instruction with Face Type, $F < 1$. These results were replicated in the AUC analyses. Namely, there was a significant main effect of Face Type, $F(1, 48) = 30.41$, $MSE = .005$, $p < .001$, indicating that overall own-race faces were remembered better ($M = 0.82$, $SD = 0.09$) than other-race faces ($M = 0.77$, $SD = 0.09$), confirming a significant CRE effect, $d_z = 0.79$. However, we did not observe a significant main effect of Instruction, $F(1, 48) = .073$, $MSE = .003$, $p = .79$, suggesting no differences between recognition of Remember ($M = 0.80$, $SD = 0.10$) and Forget (M

[1] An independent-sample t-test on overall accuracy did not show a significant difference between Asian and Caucasian participants as well as between participants in China and US Asian participants in Experiment 1. Also, we did not observe any significant interactions between the Race of participants and DF or the Race of Participants and CRE in any of the three experiments.

= 0.79, SD = 0.09) faces, d_z = 0.04. Furthermore, there was no interaction, $F(1, 48) = .82$, $MSE = .003$, $p = .37$.

Experiment 2: Medium lists

The analyses of d' revealed that there was a significant main effect of Face Type, $F(1, 64) = 34.07$, $MSE = .55$, $p < .001$, confirming the superior recognition for own-race faces (M = 2.00, SD = .67) over other-race faces (M = 1.47, SD = .60), indicating a CRE, d_z = .73. In addition, a significant main effect of Instruction was observed, $F(1, 64) = 7.09$, MSE = .17, $p = .01$, with better recognition of Remember (M = 1.80, SD = .59) compared to Forget faces (M = 1.67, SD = .52), indicating a DF effect, d_z = 0.33. However, no significant interaction of Instruction with Face Type emerged, $F(1, 64) = .21$, $MSE = .14$, $p = .65$. All of these findings were replicated in the AUC analyses. There was a significant main effect of Face Type, $F(1, 64) = 29.18$, $MSE = .009$, $p < .001$, indicating better recognition for the own-race faces (M = 0.88, SD = 0.10) than other-race faces (M = 0.81, SD = 0.11), d_z = .67. In addition, there was a significant main effect of Instruction, $F(1, 64) = 5.81$, $MSE = .004$, $p = .019$, indicating better recognition of Remember (M = 0.86, SD = 0.10) than Forget (M = 0.84, SD = 0.11) faces, confirming a significant DF effect, d_z = 0.30. However, there was no interaction, $F(1, 64) = .97$, $MSE = .004$, $p = .33$.

Experiment 3: Short lists

The analyses of d' again revealed a significant main effect of Face Type, $F(1, 58) = 23.80$, $MSE = .66$, $p < .001$, confirming the CRE effect, d_z = .64, on own-race faces (M = 2.04, SD = .71) and other-race faces (M = 1.52, SD = .67). We also obtained a significant main effect of Instruction, $F(1, 58) = 13.26$, $MSE = .19$, $p = .001$, indicating a DF effect, d_z = .48, with better recognition of Remember (M = 1.89, SD = .60) than Forget faces (M = 1.68, SD = .60). However, there was no interaction of Instruction with Face Type, $F < 1$. In terms of the AUC analyses, we again replicated a robust CRE, d_z = 0.51, with higher accuracy for own-race (M = .89, SD = .13) than other-race faces (M = .83, SD = .13), $F(1, 58) = 14.22$, $MSE = .019$, $p < .001$. There was also a significant main effect of Instruction, confirming a significant DF effect, d_z = .34, with better recognition of Remember faces (M = .87, SD = .11) than Forget faces (M = .85, SD = .12), $F(1, 58) = 7.20$, $MSE = .005$, $p < .01$. Finally, there was no interaction, $F(1, 58) = .63$, $MSE = .005$, $p = .43$.

Bayesian analyses

Given that in all of the experiments we obtained null interactions, we also included Bayes factors to assess the evidence for or against the null hypothesis of no Instruction × Face Type interaction (using default priors in JASP, JASP Team, 2020). The evidence for a model that includes only the main effects of Instruction and Face Type is evaluated against an alternative model that additionally includes the Instruction × Face Type interaction. The Bayes factor BF_{10} may be interpreted as the ratio of evidence in favor of a model that includes the interaction in contrast to a "null" model that includes only the main effects of Instruction and Face Type.

Bayesian analysis was conducted based on both d' measure and AUC measure. Based on d' measure, in Experiment 1, comparing the evidence for a model that includes only the main effects of Instruction and Face Type against an alternative model that additionally includes the interaction term yielded $BF_{10} = 0.20$, which suggests that the data were five times more likely to be in favor of null interaction. In Experiment 2, $BF_{10} = 0.18$, suggesting that the data were 5.55 times more likely to be the outcome of a model that does *not* include an interaction. Finally, in Experiment 3, $BF_{10} = 0.22$, indicating that the data were 4.55 times more likely to be in favor of null interaction.

The Bayesian analysis of AUC measure yielded similar results. In Experiment 1, we obtained $BF_{10} = .28$ for the interaction of Instruction and Face Type, which suggests the data were 3.57 times more likely to be the outcome of a model that does not include an interaction. In Experiment 2, we obtained $BF_{10} = .25$, suggesting the data were four times more likely to in favor of null interaction. Lastly, in Experiment 3, we obtained $BF_{10} = .22$, indicating the data were 4.55 times more likely to be in favor of null interaction. Overall, the Bayesian analyses on both measures provide support for the null interactions between DF and CRE across all three experiments.

Combined analyses

Across three experiments, we did not observe an interaction between DF and CRE. We combined the data across the studies to have greater power to detect a potential interaction. The DF effect of own-race and other-race faces combined across three experiments is shown in Fig. 4. We analyzed d' across the experiments and included Experiment as a variable in the mixed-factor ANOVA, along with Instruction and Face Type. The aggregated analyses revealed a significant main effect of Instruction, $F(1, 169) = 16.69$, MSE = .160, $p < .001$, a significant main effect of Face Type, $F(1, 169) = 80.52$, $MSE = .499$, $p < .001$, and a significant main effect of Experiment, $F(2, 169) = 7.98$, $MSE = 1.07$, $p < .001$, indicating decreasing accuracy across increasing list lengths. Importantly, however, there were neither two-way nor three-way interactions in the combined analyses. Consistent with individual experiments, Instruction × Face Type interaction was not significant, $F < 1$ ($BF_{10} = .12$), indicating that the results were 8.33 times more likely to

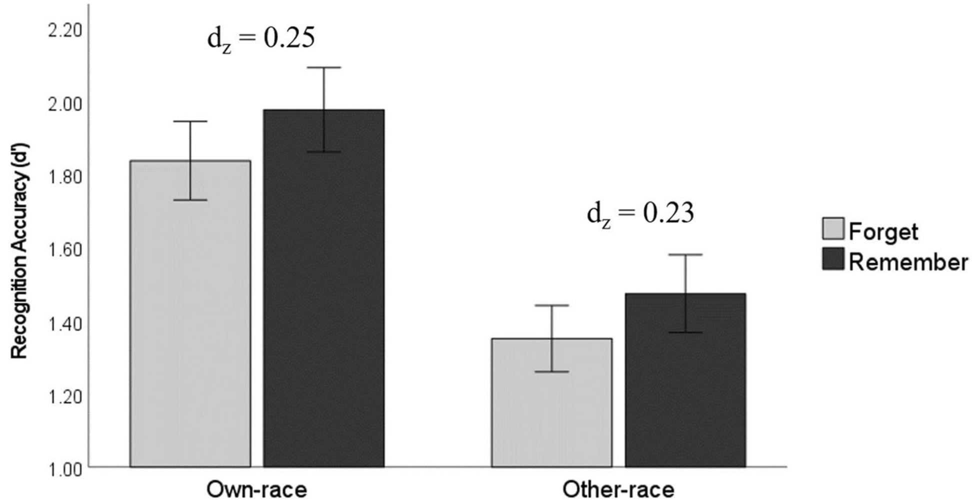

Fig. 4 Directed forgetting effect for own-race and other-race faces combined across three experiments. The error bars reflect 95% confidence intervals

be the outcome of a model that does not include an interaction between DF and CRE. The findings and conclusions hold if the aggregated analyses are conducted on the AUC measure. The Instruction × Face Type interaction was not significant in the combined analyses, $F < 1$, ($BF_{10} = .13$), indicating that the results were 7.69 times more likely to be the outcome of a model that does not include an interaction between DF and CRE.

Discussion

It is firmly established that faces of own-race are better recognized than other-race faces, and that faces can be intentionally forgotten. Our current findings replicate both the DF effect and the CRE. The main question behind this investigation was whether intentional forgetting might moderate the CRE. In three experiments that varied in list length, we never observed the interaction between the two effects. Own-race faces were as forgettable as other-race faces. Also, when the data were pooled across the experiments to increase the power to detect such an interaction, there was still no evidence supporting differential DF for own-race than other-race faces. The null interactions were further confirmed by the Bayesian analyses.

Although the CRE is a robust finding explained by different accounts, several previous studies have reported a reduction or elimination of the CRE using various manipulations. For example, Hills et al. (2011) instructed White participants to focus more on the lower part of faces, which was considered to be the critical area of Black faces. Doing so eliminated the CRE by improving later recognition performance towards Black faces more than towards White faces, whereas instructing participants to attend to the upper portion of the faces left the CRE intact. Also, Shriver and Hugenberg (2010) found that when other-race faces were labeled with occupations as high in power and socioeconomic status, the CRE was attenuated with the results showing equal recognition performance for own-race and other-race faces. Overall, the CRE can be modulated under some circumstances. Against this backdrop, it is important to note that DF does *not* appear to be among those manipulations that can alter the magnitude of the CRE.

Based on perceptual expertise and social-cognitive views of the CRE, we predicted the emergence of an interaction between DF and the CRE, despite their differential predictions on the magnitude of DF for own-race versus other-race faces. We did not obtain such an interaction either in individual experiments or in the combined data, where own-race and other-race faces were equally susceptible to the DF manipulation. The lack of interaction between DF and CRE may indicate that no single theory sufficiently explains the CRE. More views suggest that a mixture of mechanisms may be involved in explaining the CRE (e.g., Hugenberg et al., 2010; Meissner et al., 2005), including the strength of memory representation and unified/disjointed bindings of facial features, and these two views may cancel the opposite effects predicted in our studies, ultimately leading to equivalent DF of own-race and other-race faces.

Our findings suggest race might play a fundamentally different role in intentional forgetting of individuals in comparison with other facial features. As described previously, studies indicated that certain facial characteristics that are related to facial memorability render faces to be differentially susceptible to DF. For instance, a handful of studies have found that emotional expressions impact the memorability of faces, where unfamiliar happy faces are better remembered than angry or neutral faces (e.g., Foa et al., 2000; Kottoor, 1989). In a DF investigation with unfamiliar faces of happy and angry expressions, Tay and Yang (2017) found that happy faces are less resistant to DF than are angry faces. It appears that faces of different memorability resulting from the facial expression interact with DF manipulation. Our findings, on the other hand, indicate that even though own-race faces are more memorable than other-race faces, they do not appear to be differentially susceptible to DF. These different outcomes between the current findings and previous ones may arise from the types of characteristics (i.e., race and expression) in facial recognition. According to Bruce and Young (1986), face processing consists of two distinct routes, one involving recognition of unchangeable (invariant) features such as facial identity and one responsible for processing changeable (dynamic) features such as facial expression. These two pathways were found to involve functionally and neurologically unique systems (for review, see Calder & Young, 2005). Critically, our results suggest that faces varying in unchangeable features including race (or racial identity) may not be differentially susceptible to DF but faces differing in dynamic characteristics including facial expression do, as demonstrated by previous studies. Future studies can extend the current findings and systematically investigate if and how DF may differentially impact these two types of facial information.

The absence of interaction between DF and CRE is also consistent with the facial memorability literature. Memorability is defined as a predictive value of the likelihood that an item will be remembered on a subsequent memory test (Isola et al., 2011). Facial memorability appears to be intrinsic and identity-specific, and is preserved across different facial expressions and viewpoints (Bainbridge, 2017). Importantly, Bainbridge (2020) found that DF was equally likely for faces of different degrees of memorability, which is consistent with the current findings, indicating that the DF effect does not differ between own-race faces (more memorable) and other-race faces (less memorable).

The current studies tentatively suggest that DF of faces might be easier to observe with shorter list lengths than longer lists. Whereas with longer lists in Experiment 1 we did not observe DF, we detected DF in subsequent medium list length and short list length experiments, with an increasing effect size of DF. These observations are consistent with previous studies in the literature. For example, Reber et al. (2002) used 150 faces at encoding and 300 faces at test and failed to observe DF for faces. In contrast, other studies used fewer facial stimuli, presenting participants with 20, 60, and 48 faces during study, respectively, and obtained DF (e.g., Metzger, 2011; Quinlan & Taylor, 2014; Tay & Yang, 2017). We suspect that a large number of faces presented during study may hinder the emergence of DF for faces because participants may stop complying with Forget/Remember instructions due to the increased difficulty of the task. The role of list length in DF needs to be more systematically evaluated in future research given the Experiment × DF interaction did not emerge as being significant in our combined analysis. Unlike the potential impact on DF, list length did not appear to alter the magnitude of the CRE as reflected in approximately similar effect sizes of the CRE, which remained a large size effect across the experiments with different list lengths. Thus, CRE is a robust phenomenon, which unfortunately does not appear to be mitigated by DF instructions.

Author's Contribution Huiyu Ding contributed to study design, collected and analyzed the data, and wrote the initial draft of the manuscript. Experiment 1 and Experiment 2 were part of Huiyu Ding's Honors' Thesis. Jonathon Whitlock assisted with data analyses and contributed to manuscript preparation. Lili Sahakyan supervised all aspects of the project.

Data availability The dataset is available on the Open Science Framework (https://osf.io/uxe32/)

Declarations

Conflicts of interest The authors declare that they have no conflicts of interest.

Ethics approval The studies were approved by the Institutional Review Board of the University of Illinois.

Consent to participate All participants provided informed consent prior to commencing the studies.

Consent for publication Not applicable

References

Bainbridge, W. A. (2017). The memorability of people: Intrinsic memorability across transformations of a person's face. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 706–716.

Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia, 141*, 12.

Basden, B. H. (1996). Directed forgetting: Further comparisons of the item and list methods. *Memory, 4*(6), 633–653.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning &Verbal Behavior, 9*(3), 255–268.

Bjork, R. A., & Geiselman, R. E. (1978). Constituent processes in the differentiation of items in memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*(4), 347–361.

Bjork, R. A., Laberge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science, 10*(2), 55–56.

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

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641–651.

Caretta, T. R., & Moreland, R. L. (1983). The Direct and Indirect Effects of Inadmissible Evidence1. *Journal of Applied Social Psychology, 13*(4), 291–309.

Chiu, Y., Wang, T. H., Beck, D. M., Lewis-Peacock, J., & Sahakyan, L. (2021). Separation of item and context in item-method directed forgetting. *NeuroImage, 235*, 10.

Dulaney, C. L., Marks, W., & Link, K. E. (2004). Aging and directed forgetting: Pre-cue encoding and post-cue rehearsal effects. *Experimental Aging Research, 30*(1), 95–112.

Dwyer, J., Neufeld, P., & Scheck, B. (2003). *Actual innocence: when justice goes wrong and how to make it right*. New American Library.

Earles, J. L., & Kersten, A. W. (2002). Directed forgetting of actions by younger and older adults. *Psychonomic Bulletin & Review, 9*(2), 383–388.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Fitzgerald, R. J., Price, H. L., & Oriet, C. (2013). Intentionally forgetting other-race faces: Costs and benefits? *Journal of Experimental Psychology: Applied, 19*(2), 130–142.

Foa, E. B., Gilboa-Schechtman, E., Amir, N., & Freshman, M. (2000). Memory bias in generalized social phobia: Remembering negative emotional expressions. *Journal of Anxiety Disorders, 14*(5), 501–519.

Geiselman, R. E. (1974). Positive forgetting of sentence material. *Memory & Cognition, 2*(4), 677–682.

Geiselman, R. E., & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory & Cognition, 13*(1), 57–62.

Goernert, P. N., Corenblum, B., & Otani, H. (2011). Directed forgetting of faces: The role of response criterion. *The Quarterly Journal of Experimental Psychology, 64*(10), 1930–1938.

Hauswald, A., & Kissler, J. (2008). Directed forgetting of complex pictures in an item method paradigm. *Memory, 16*(8), 797–809.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments, & Computers, 27*, 46–45.

Hills, P. J., Werno, M. A., & Lewis, M. B. (2011). Sad people are more accurate at face recognition than happy people. *Consciousness and Cognition: An International Journal, 20*(4), 1502–1517.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*(4), 1168–1187.

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 145–152.

JASP Team (2020). JASP (Version 0.14)[Computer software].

Kassin, S. M., & Sommers, S. R. (1997). Inadmissible testimony, instructions to disregard, and the jury: Substantive versus procedural considerations. *Personality and Social Psychology Bulletin, 23*(10), 1046–1054.

Kottoor, T. M. (1989). Recognition of faces by adults. *Psychological Studies, 34*(2), 102–105.

Lee, Y. (2013). Costs and benefits in item-method directed forgetting: Differential effects of encoding and retrieval. *Journal of General Psychology, 140*(3), 159–173.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition. *Journal of Experimental Psychology: General, 129*, 559–574.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Erlbaum.

MacLin, O. H., & Malpass, R. S. (2003). The ambiguous-race face illusion. *Perception, 32*(2), 249–252.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates Publishers.

Mallard, D., & Perkins, D. P. (2005). Disentangling the Evidence: Mock Jurors, Inadmissible Testimony and Integrative Encoding. *Psychiatry, Psychology and Law, 12*, 289–297.

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), 3–35.

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: A dual-process approach. *Applied Cognitive Psychology, 19*(5), 545–567.

Metzger, M. M. (2011). Directed forgetting: Differential effects on typical and distinctive faces. *Journal of General Psychology, 138*(2), 155–168.

Michel, C., Rossion, B., Han, J., Chung, C., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science, 17*(7), 608–615.

Nowicka, A., Marchewka, A., Jednoróg, K., Tacikowski, P., & Brechmann, A. (2011). Forgetting of emotional information is hard: An fMRI study of directed forgetting. *Cerebral Cortex, 21*(3), 539–549.

Otani, H., Libkuman, T. M., Goernert, P. N., Kato, K., Migita, M., Freehafer, S. E., & Landow, M. P. (2012). Emotion, directed forgetting, and source memory. *British Journal of Psychology, 103*(3), 343–358.

Quinlan, C. K., & Taylor, T. L. (2014). "I never forget a face, but in your case I'll be glad to make an exception": Intentional forgetting of emotional faces. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 68*(3), 212–221.

Quinlan, C. K., Taylor, T. L., & Fawcett, J. M. (2010). Directed forgetting: Comparing pictures and words. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 64*(1), 41–46.

Reber, P. J., Siwiec, R. M., Gitelman, D. R., Parrish, T. B., Mesulam, M., & Paller, K. A. (2002). Neural correlates of successful encoding identified using functional magnetic resonance imaging. *The Journal of Neuroscience, 22*(21), 9541–9548.

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80*(3), 313–331.

Rodin, M. J. (1987). Who is memorable to whom: A study of cognitive disregard. *Social Cognition, 5*(2), 144–165.

Sahakyan, L. (2022). Current perspectives on directed forgetting. In A. Wagner & M. Kahana (Eds.), *Oxford handbook of human memory: Foundations and applications*.

Sahakyan, L., & Foster, N. L. (2009). Intentional forgetting of actions: Comparison of list-method and item-method directed forgetting. *Journal of Memory and Language, 61*(1), 134–152.

Sahakyan, L., Delaney, P. F., Abushanab, B., & Foster, N. L. (2013). List-Method Directed Forgetting in Cognitive and Clinical Research: A Theoretical and Methodological Review. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 59). Elsevier.

Scully, I. D., & Hupbach, A. (2020). Directed forgetting affects how we remember and judge other people. *Journal of Applied Research in Memory and Cognition, 9*(3), 336–344.

Sell, A. J. (2016). Applying the intentional forgetting process to forgiveness. *Journal of Applied Research in Memory and Cognition, 5*(1), 10–20.

Shriver, E. R., & Hugenberg, K. (2010). Power, individuation, and the cross-race recognition deficit. *Journal of Experimental Social Psychology, 46*(5), 767–774.
Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7*(1), 36–97.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*, 137–149.
Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*(1), B1–B9.
Tay, P. K. C., & Yang, H. (2017). Angry faces are more resistant to forgetting than are happy faces: Directed forgetting effects on the identity of emotional faces. *Journal of Cognitive Psychology, 29*(7), 855–865.
Tullis, J. G., Benjamin, A. S., & Liu, X. (2014). Self-pacing study of faces of different races: Metacognitive control over study does not eliminate the cross-race recognition effect. *Memory & Cognition, 42*(6), 863–875.
Waldum, E. R., & Sahakyan, L. (2012). Putting congeniality effects into context: Investigating the role of context in attitude memory using multiple paradigms. *Journal of Memory and Language, 66*(4), 717–730.
Wang, T. H., Placek, K., & Lewis-Peacock, J. A. (2019). More Is Less: Increased Processing of Unwanted Memories Facilitates Forgetting. *The Journal of Neuroscience, 38*(17), 3551–3560.
Xie, H., Hu, X., Mo, L., & Zhang, D. (2021). Forgetting positive social feedback is difficult: ERP evidence in a directed forgetting paradigm. *Psychophysiology, 58*(5), 14.
Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin, 133*(5), 800–832.
Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and Motivation in Face Recognition: A Critical Review of Theories of the Cross-Race Effect. *Personality and Social Psychology Review, 16*(2), 116–142.

Open practices statement The data for all experiments are available on the Open Science Framework (https://osf.io/uxe32/), and none of the experiments were preregistered.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.