The Effect of Face Masks on Forensic Face Matching: An Individual Differences Study

Alejandro J. Estudillo[1, 2], Peter Hills[1] & Hoo Keat Wong[2]

[1]Bournemouth University

[2]University of Nottingham Malaysia

Manuscript accepted in Journal of Applied Research in Memory and Cognition

Abstract

In the forensic face matching task, observers are presented with two unfamiliar faces and must determine whether they depict the same identity. Due to the Covid-19 pandemic, some governmental authorities require the use of face masks in public spaces. However, face masks impair face identification by disrupting holistic processing of faces. The present study explores the effect of face masks on forensic face matching. Compared to a full-view condition, performance decreased when a face mask was superimposed on one face (Experiment 1) and both faces (Experiment 2) of a pair. Although a positive correlation between the full-view and the mask conditions was found, high proficiency in the full-view condition did not always generalize to the mask condition. Additionally, the mask generally has a more negative impact in those participants with better performance in the full-view condition. The theoretical and practical implications of these findings are discussed.

Keywords: forensic face matching; Covid-19; face mask; individual differences

In the forensic face matching task, two unfamiliar faces are presented simultaneously side-by-side and observers are tasked to determine whether both pictures depict the same or different identities. This task is analogous to face matching at identity checkpoints such as passport control (Bindemann, 2021) and, in clear contrast to its apparent ease, research has shown that it is highly error-prone, even under the most favourable visual conditions (Bindemann et al., 2012, 2013; Estudillo & Bindemann, 2014). This is such that mean error rates between 20% and 36% have been commonly reported (Alenezi & Bindemann, 2013; Bindemann et al., 2012; Burton et al., 2010; Megreya & Burton, 2008) and highly experienced passport control officers' performance in face matching is equally poor to that of untrained observers (Papesh, 2018; White, Kemp, Jenkins, Matheson, & Burton, 2014; White, Phillips, Hahn, Hill, & O'Toole, 2015).

Two non-mutually exclusive sources seem to explain the errors found in face matching (for review, see Fysh & Bindemann, 2017). On the one hand, according to the data limit account, the information within the pair of face stimuli is not sufficient to solve face matching (Alenezi & Bindemann, 2013; Estudillo & Bindemann, 2014; Fysh & Bindemann, 2017). In fact, image manipulations, such as changes in view (Estudillo & Bindemann, 2014; Kramer & Reynolds, 2018), image degradation (Bindemann et al., 2013; Bruce et al., 2001) and feature masking (Carragher & Hancock, 2020; Dhamecha et al., 2014) impair unfamiliar face matching. On the other hand, according to the resource limit account, observers' fail to use facial information to solve the task (Estudillo & Bindemann, 2014; Fysh & Bindemann, 2017). Evidence supporting this account comes from individual differences studies. Performance in forensic face matching tasks, rather than being uniform, ranges from chance levels, which is observed in participants with face recognition disorders such as developmental prosopagnosia, to ceiling levels, which can be observed in the so-called super-recognizers (Bobak et al., 2016; Bruce et al., 2018; Burton et al., 2010; Fysh & Bindemann, 2018; White et al., 2016). In fact, a subgroup of observers are able to solve face matching with remarkably high accuracy, even under challenging visual conditions, such as drastic changes in viewpoint (Estudillo & Bindemann, 2014), image blurring (Robertson et al., 2016) and inverted faces (White, Phillips, Hahn, Hill, & O'Toole, 2015).

Given the critical consequences that wrong identifications might have in applied scenarios, different research has explored these individual differences in face identification. However, although different degrees of correlations have been found between different face identification measures and conditions, intra-observers' consistency across these identification conditions is not always evident (Bruce et al., 2018; McCaffery et al., 2018; Verhallen et al., 2017). For example, recent research has shown that high face matching performers show limited generalization to different identification conditions (Fysh et al., 2020). These findings suggest that, despite some commonalities between different identification conditions, they might reflect different cognitive operations. From a more practical point of view, such a lack of consistency highlights the importance of using different conditions when evaluating face identification for personnel selection in applied scenarios.

The Covid-19 pandemic has changed several aspects related to the way that humans communicate and interact with each other. For instance, to prevent the spread of the virus, some governmental authorities require the use of face masks during outdoor activities and interaction with others. Thus, one important question that arises is whether, and if so, to what extent, face masks affect face identification. This question is not trivial as the outbreak of the Covid-19 pandemic poses a new challenge for people working in security settings: having to identify a person when only approximately 50% of the face is available. A recent study showed that although face masks impaired face recognition, this effect was smaller when the faces were presented upside down (Freud et al., 2020). This suggests that, compared to full-view faces, masked faces are processed in a qualitatively different way, with reduced holistic processing of faces (Freud et al., 2020).

Regarding face matching, a pre-Covid-19 pandemic study found that face masks, fake moustaches and beards decreased performance in face matching tasks (Dhamecha et al., 2014). More recently, Carragher and Hancock (2020) superimposed face masks on a set of familiar and unfamiliar faces and found worse matching performance when faces were masked compared to a full-view condition (Carragher & Hancock, 2020). Although these studies support the data limit account of face matching, only the overall performance was reported, so it is unclear if despite the constraints imposed by the face mask, some observers are still able to solve the task. Interestingly, a more recent study using real images of people wearing masks found that although face masks disrupted face matching performance, super-recognizers did not only performed better than control participants in the mask condition, but they were able to reach an accuracy level close to 90% in this mask condition (Noyes et al., 2021).

In a world where face masks are becoming highly common in our daily lives, exploring individual differences in matching full-view and masked faces is highly relevant for theoretical and practical reasons. From a theoretical point of view, an individual differences approach would reveal the role of resource limits in face matching. In other words, if some observers can consistently identify masked faces, this would indicate that face matching is, in principle, solvable even when only the top part of the face is visible. Given the low validity of self-reported measures of face identification (Bobak et al., 2019; Estudillo, 2021; Estudillo & Wong, 2021; Palermo et al., 2017), such an objective individual differences approach would help to screen those observers with superior face matching performance (Bruce et al., 2018; Ramon et al., 2019).

However, as previously mentioned, face identification performance does not always generalize to different identification conditions (Fysh et al., 2020). Thus, although some degree of association between a mask and a full-view condition can be expected, it is possible that the effect of the mask is not homogeneous across all the observers. We explored this in two different ways. First, we examined whether the effect of the mask is modulated by general face matching expertise. To test this aim, we calculated the full-view advantage as the difference in matching performance between a full-view condition and a mask condition. This measure quantifies the impairment that face masks produce to face matching. Values close to zero indicate that observers did not benefit from the full-view condition or, in other words, that the mask has no effect in face matching. Positive values indicate a full-view advantage (i.e., observers performed better in the full-view compared to the mask condition). Negative values indicate that observers performed better in the mask condition compared to the full-view condition. If the mask has a homogeneous effect across observers, it is expected that the full-view advantage would not be correlated with general matching performance. Secondly, we also analysed the performance profiles by the top 5% performers in each condition, as recent research has shown dissociations between different identification conditions at this level of performance (Fysh et al., 2020; Stacchi et al., 2020).

Experiment 1

In Experiment 1, observers performed a face matching task with pairs of faces presented either in full view (full-view condition), or with a mask superimposed on one of the faces of the pair (mask condition). To explore the general effects of face masks on matching performance, we compared sensitivity (i.e., *d-prime*) and response bias (i.e., C) across the full-view and mask conditions. In addition, we also investigated the effect of the mask at an individual level. The aim of this analysis was twofold. First, we wanted to determine the degree of correlation between the full-view and the mask conditions. For this, we explored whether performance in the mask condition is associated with performance in the full-view condition. Second, to investigate whether the effect of the mask is homogeneous across observers, we explored whether the full-view advantage (performance in full-view faces – performance in masked faces) was correlated with performance in the full-view condition. If the mask has a homogeneous effect across observers, it is expected that the full-view advantage would not correlate with general matching performance. We also examined the performance profiles by the top 5% performers in each condition, as some recent research has shown dissociations between different identification conditions at this level of performance (Fysh et al., 2020; Stacchi et al., 2020).

Method

Participants

This study was approved by the research ethics committee of Bournemouth University. We tested participants over the web. Our initial sample consisted of 220 participants recruited using the platform Testable Minds. However, 16 participants were removed due to abnormally fast response times (< 400 ms) and/or performance below chance level. Thus, our final sample consisted of 204 participants (74 females, 127 males, 2 other) with a mean age of 30 years (SD = 11.17). Due to technical issues, ethnic group was not recorded in this experiment. Participants gave their consent to participate in this study and received 3 USD as a compensation for their participation.

Stimuli and Procedure

The stimuli set comprised a total of 120 pairs of female and male Caucasian faces from the Glasgow Unfamiliar Face Database (Burton et al., 2010). The faces and identities of each pair were unique and not repeated across trials. One face photograph in each pair was taken with a high-quality digital camera, while the other was a still frame from high-quality video. All faces were shown in greyscale on a white background. Half of these pairs comprised identity matches, in which two different photographs of the same person were shown. The other half depicted identity mismatches, which showed two faces of two different identities. Sixty pairs depicted two full-view faces (30 match trials, 30 mismatch trials), while the other half depicted one full-face and one face with a face mask (30 match trials, 30 mismatch trials), with the position of the mask (left face vs. right face) counterbalanced across trials. Although the face masks were fitted to the face stimuli using Photoshop, this process was done individually for each face with the aim of covering the same features that a real mask would. The allocation of the face pairs to the mask condition was randomized across participants. Stimuli examples are presented in Figure 1.

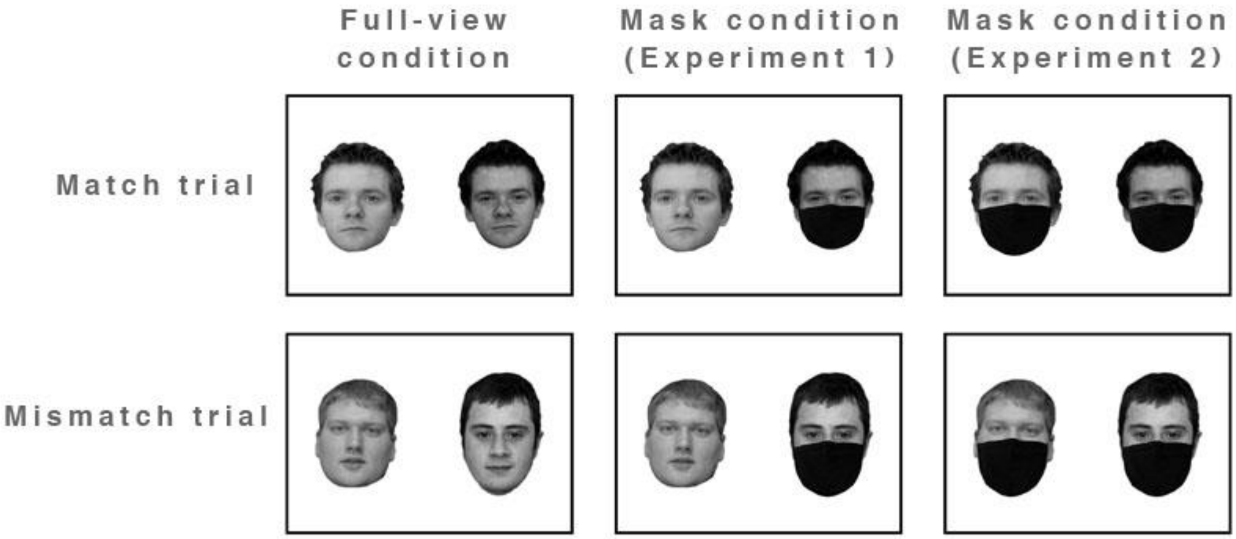

Figure 1. Example stimuli used in Experiment 1 and Experiment 2

On each trial, observers were firstly shown a central fixation cross for 1000 ms, which was replaced by two side-by-side face images. Observers were asked to determine whether the face pairs were identity matches (i.e., the same person) or mismatches (i.e., two different people) by pressing one of two buttons. This display remained onscreen until a response was made. There was no time limit for this task. Trials were randomized across participants. A demo of this task can be found at https://www.testable.org/experiment/4045/694357/start

Results[1]

Group analysis

In the first part of the analysis, to get a complete understanding of the overall effect of face masks in forensic face matching, hits (i.e., correctly identified match pairs) and false alarms (i.e.,

[1] Data for experiments 1 and 2 are available upon request

incorrectly identified mismatch pairs) were used to calculate d-prime, a measure of sensitivity, and C, a measure of response bias (Stanislaw & Todorov, 1999). Higher d-prime values indicate better matching accuracy. As some participants performed with perfect accuracy, d-prime, was corrected using Hautus' recommendations for extreme values (Hautus, 1995). Positive C values indicate a conservative bias (i.e., observers would make more mismatch responses) while negative C values show a more liberal bias (i.e., observers would make more match responses). A C value of zero indicates no response bias.

D-prime was higher in the full-view condition (mean = 2.71; SD = .77) compared to the mask condition (mean = 2.13; SD = .66). A paired-samples t-test confirmed that these differences were statistically significant, $t(203) = 13.77$, $p < .001$, d = .96, 95% CI = .79 – 1.12. We found that C was higher in the mask condition (mean = .05; SD = .39) compared to the full-view condition (mean = -.03; SD = .37), $t(203) = 3.60$, $p < .001$, d = .25, 95% CI = .11 – .39. We also explored whether these conditions produced a response bias (i.e., $C \neq 0$). One sample t-tests confirmed that while observers did not present a response bias in the full-view condition, $t(203) = 1.48$, $p = .14$, observers showed a conservative bias in the mask condition, $t(203) = 1.99$, $p < .05$, d = .14, 95% CI = .02 – .25.

Individual differences analysis

In the second part of our analysis, we explored the degree of correlation between the full-view and the mask conditions. For d-prime, the relationship between these two conditions is presented in Figure 2A. As shown in this figure, observers' performance in the full-view was positively correlated with their performance in the mask condition, r = .66, $p < .001$, 95% CI = .57 - .73.

Relationship between the magnitude of the full-view advantage and face matching performance

We also explored whether the effect of the mask was equivalent across all the range of scores in the full-view condition. The relationship between the full-view advantage (i.e., differences in performance between the full-view and mask conditions) and the performance in the full-view condition is presented in Figure 2B. As shown in this figure, observers' scores in the full-view condition were positively associated with the full-view advantage, r = .55, $p < .001$, 95% CI = .45 - .64. This correlation indicates that as participants performed better with full-view faces, masks tended to have a greater impact on matching performance.

Analysis of top performing observers

As shown in Figure 2A, a few observers completed the task with remarkably high-performance in either the full-view condition, the mask condition or both conditions. In an attempt to fully characterize the performance pattern of these observers across conditions, in line with other authors (Fysh et al., 2020; Stacchi et al., 2020), we isolated observers with the top 5% scores in each viewing condition (10 observers per condition). These observers were selected firstly based on the highest d-prime scores in each condition. In cases where more than 10 observers achieved the highest d-prime scores, they were then selected based on the shortest median correct response times in each condition, as when complemented with accuracy, correct reaction times is a measure of the efficiency to solve a task (Jensen, 2006)[2]. Two observers met these criteria in both conditions (crosses in Figures 2A and B). For this reason, we could not formally compare both groups, however, visual inspections to Figure 2A and Table 1 suggest that the top 5%

[2] The same results were obtained when all the observers who reached the higher d-prime score were included in the analysis (see supplementary results 1).

observers full-view group (squares in Figure 2A and B) performed better in the full-view condition compared to the top 5% observers mask group (triangles in Figure 2A and B). In contrast, the top 5% observers mask group performed better in the mask condition compared to the top 5% observers full-view group.

Table 1. Mean d-prime, C and magnitude of the mask effect (standard deviations in brackets) for the top 5% observers in each viewing condition. Columns represent the top 5% observers in each condition, and each row performance of these observers on each condition.

		Top 5% observers full-view	**Top 5% observers mask**
Full-view condition	d-prime	4.28 (.00)	3.41 (.62)
	C	.00 (.00)	-.18 (.30)
Mask condition	d-prime	2.96 (.50)	3.60 (.18)
	C	.02 (.35)	.11 (.34)

Within-group comparisons across conditions revealed that the top 5% observers full-view group performed better in the full-view condition than in the mask condition, $t(9) = 8.21$, $p < .001$, d = 5.81, 95% CI = 3.46 – 8.02, with similar response bias across both conditions $t(9) = .19$, $p = .81$[3]. Additionally, the top 5% observers mask group show similar performance and response bias in the full-view condition and the mask condition, both $ts(9) \leq 1.94$, $p \geq .08$.

[3] Performance of these observers was perfect in the full-view condition (d-prime score = 4.28; C = 0), so the variance was equal to 0. For this reason, two one sample t-tests were conducted comparing these observers' scores in the mask condition with the values 4.28 for d-prime and 0 for C.

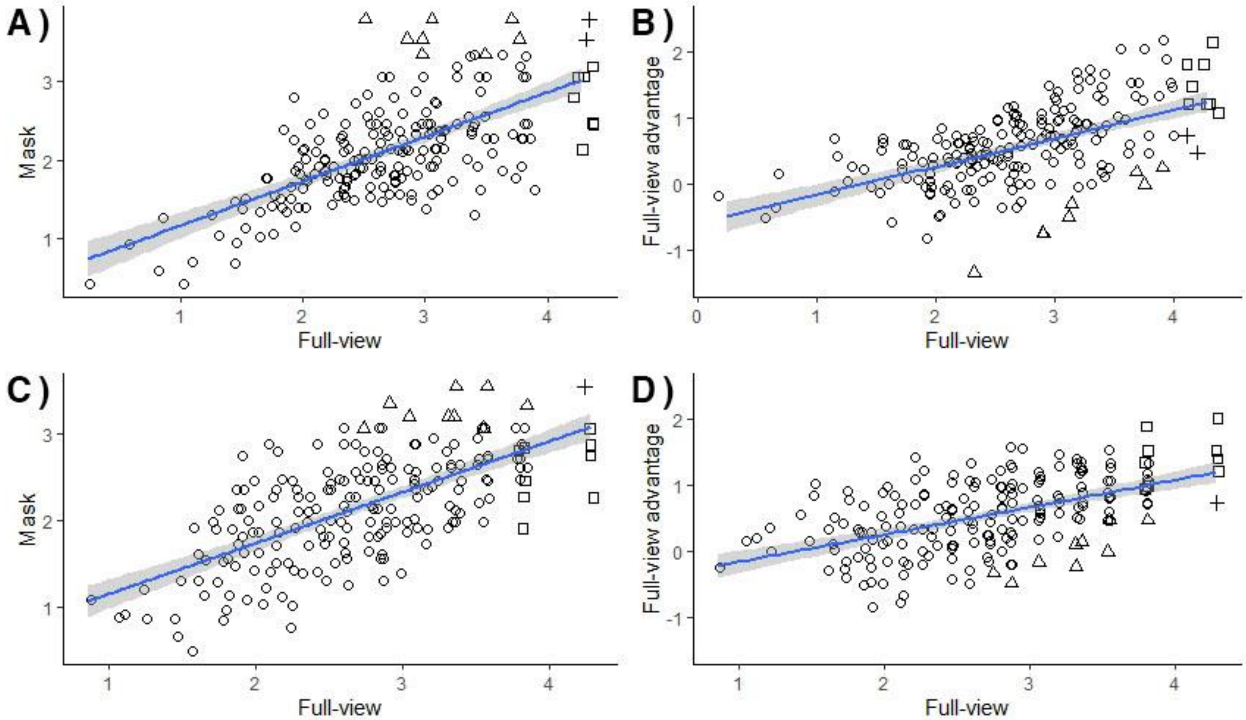

Figure 2. Top panel represents Experiment 1 and bottom panel Experiment 2. Relationships between the full-view and the mask condition for d-prime (A, C). (B, D) relationship between the full-view and the full-view advantage (full-view – mask performance). Squares denote observers with the top 5% scores in the full-View condition. Triangles denote observers with top 5% scores in the mask condition. Crosses denote observers with to 5% scores in both conditions. Note: to avoid overlapping between datapoints, jittering was used for illustration purposes.

Discussion

In Experiment 1, we found that observers were more accurate in the full-view condition compared to the mask condition, replicating recent results (Carragher & Hancock, 2020; Noyes et al., 2021). In addition, the analysis of response bias revealed that, compared to the full-view condition, participants are more likely to make more “mismatch” responses in the mask condition. Despite this overall pattern, the individual differences analysis revealed three interesting results. First, it is noteworthy that some observers performed with remarkably high sensitivity in the mask condition. Interestingly, when compared to the full-view condition, some of these observers did not show any impairment in matching as consequence of the mask. This high-performance level is striking, as the face masks covers approximately 50% of the face. This high performance also reveals an important finding: the effect of the face mask on face matching is, in fact, solvable most of the time.

Second, although the performance in the mask condition was predicted by performance in the full-view condition, we found that the mask had a more negative impact in those participants with better performance in the full-view condition. We will discuss the theoretical implications of this finding in the general discussion, but from an applied point of view this finding highlights an apparent contradiction for personnel selection in security settings, as those observers with better performance with full-view faces would be more vulnerable to mask effects. However, our top performers analysis revealed that although the top 5% observers in the full-view condition performed better in the full-view condition than in the mask condition, such a dissociation was not found in the top 5% observers in the mask condition. This highlights the importance of using different identification conditions when evaluating face identification abilities in applied scenarios (Fysh et al., 2020; Stacchi et al., 2020).

Nevertheless, there is an important shortcoming in our stimuli. Previous research has demonstrated the importance of congruency between stimuli on face identification. For example, when observers encode a face with a ski-mask or a Niqab, participants perform better in a line-up identification task when that face is also wearing a ski-mask or Niqab, compared to a full-view face (Manley et al., 2019; Toseeb et al., 2014). More relevant to the present study, in a face matching task, participants performed better when both faces of the pair were presented with glasses or without glasses, compared to when only one of the faces of the pair had glasses (Kramer & Ritchie, 2016; Toseeb et al., 2014). In our first experiment, in the mask condition, only one of the faces wore a mask, making the pair of faces incongruent. However, in the control condition, both faces are presented in full-view, making the pair of faces congruent. Thus, it is possible that in our first experiment, the effect of the mask is confounded with the effect of incongruency. To rule out this possibility, in the mask condition of Experiment 2, both faces of the pair were partially concealed by a face mask.

Experiment 2

This experiment aims to further investigate the effect of face masks on face identification. As in Experiment 1, observers were presented with pair of faces and were asked to determine whether the two pictures depict the same or two different identities. While in half of the trials both faces depict two full-view frontal faces, in the other half, both faces were covered by a face mask. This design avoids confounding effects of congruency between both faces of the pair and masking.

Method

Participants

This study was approved by the research ethics committee of Bournemouth University. We tested participants over the web, none of whom took part in Experiment 1. Our initial sample consisted of 222 participants recruited using the platform Testable Minds. However, 18 participants were removed due to abnormally fast response times (< 400 ms) and/or performance below chance level. Thus, our final sample consisted of 204 participants (89 females, 113 males, 2 others) with a mean age of 31 years (SD = 10.23). 117 participants reported to be Caucasian, 66 Asian, 12 Black, 8 Hispanic and 11 defined themselves as others. Participants gave their consent to participate in this study and received a 3 USD for their participation.

Stimuli and Procedure

Stimuli and procedure were identical to that of Experiment 1 with the difference that in the mask condition, both faces were covered by a face mask. A demo of this task can be found at https://www.testable.org/experiment/4045/145900/start

Results

Group analysis

As in Experiment 1, d-prime was first analysed. D-prime was higher in the full-view condition (mean = 2.71; SD = .71) compared to the mask condition (mean = 2.15; SD = .62). A paired-samples t-test confirmed that this difference was statistically significant, $t(203) = 14.27$, $p < .001$, d = 1.00, 95% CI = .83 – 1.16. In contrast, C was similar for the full-view condition (mean = .00; SD = .35) and the mask condition (mean = .03; SD = .39), $t(203) = 1.33$, $p = .18$. One sample t-tests revealed no response bias either in the full-view or mask conditions, both $ts(203) \leq 1.29$, $ps \geq .19$.

Individual differences analysis

Figure 2C represents the correlation between the full-view and the mask condition in Experiment 2. As can be seen in this figure, observers' performance in the full-view condition was positively correlated with their performance in the mask condition, r = .66, $p < .001$, 95% CI = .58 - .73.

Relationship between the magnitude of the full-view advantage and face matching performance

Finally, we explored whether the mask had an equivalent effect across all the range of scores in the full-view condition. The correlation between the full-view advantage and the performance in the full-view condition is presented in Figure 2D. As shown in this figure, observers' d-prime scores in the full-view condition were positively associated with the full-view advantage, r = .53, *p* < *.001,* 95% CI = .42 - .62. This correlation shows that as participants performed better with full-view faces, masks tended to have a stronger impact on matching performance.

Analysis of top performing observers

Figure 2C also reveals that a few observers completed the task with remarkably high-performance in either the full-view condition, the mask condition or both conditions. Thus, as in Experiment 1 (see also Fysh et al., 2020; Stacchi et al., 2020), in an attempt to fully characterize the performance pattern of these observers across conditions, we isolated observers with top 5% scores in each viewing condition (Table 2). One participant met the criteria in both conditions (crosses in Figures 2C, and D). Due to this overlapping, we could not formally compare both groups. Visual inspections to Figure 4 and Table 2 suggest that the top 5% observers full-view group (squares in Figures 2C, and D) performed better in the full-view condition compared to the top 5% observers mask group (triangles in Figures 2C, and D). In contrast, the top 5% observers mask group performed better in the mask condition compared to the top 5% observers full-view group. Additionally, the top 5% observers full-view group performed better in the full-view condition than in the mask condition, $t(9) = 10.65$, $p < .001$, d = 3.36, 95% CI = 1.70 – 5.00. However, the response bias for this group was similar across the full-view and the mask conditions, $t(9) = .09$, $p = .92$. The top 5% observers mask group performed similarly in the full-view condition and the mask condition, $t(9) = .65$, $p = .53$. This group also showed a higher response bias in the mask condition compared to the full-view condition, t(9) = 2.26, p = .05, d = .71, 95% CI = .00 – 1.40.

Table 2. Mean d-prime and C (standard deviations in brackets) for the top 5% observers in each viewing condition. Columns represent the top 5% observers in each condition, and each row performance of these observers on each condition.

		Top 5% observers full-view	**Top 5% observers mask**
Full-view condition	d-prime	4.04 (.25)	3.38 (.45)
	C	-.02 (.17)	.15 (.32)
Mask condition	d-prime	2.67 (.46)	3.30 (.18)
	C	-.03 (.45)	.36 (.31)

Discussion

In Experiment 2, observers were more accurate in the full-view condition compared to the mask condition. In contrast to Experiment 1, at group level, we found no differences in response bias between the full-view and the mask condition. This finding, in conjunction with those of Experiment 1, are consistent with previous research showing that incongruency between faces in a matching task lead to a stronger bias to make mismatch responses (Kramer & Ritchie, 2016).

Our results also showed that some observers were able to complete the task in the mask condition with a remarkably high, albeit not perfect, performance, demonstrating that face matching is solvable in most situations, even under challenging visual constraints. We also found a correlation between performance in the full-view and mask conditions. In addition, the magnitude of the full-view advantage was positively correlated with performance in the full-view condition, suggesting that the mask had a stronger effect in better matching performers. However, while the top 5% observers in the full-view condition performed better in the full-view condition than in the mask condition, the top 5% observers in the mask condition performed similarly across both conditions. Yet, the latter group also showed a more conservative response bias in the mask condition compared to the full-view condition.

General discussion

The aim of this study was to investigate the effect of face masks on forensic face matching. In Experiment 1, in the mask condition, only one of the faces of the pair was covered by a face mask. We found that observers' performance was worse in the mask condition compared to the full-view condition, replicating recent findings (Carragher & Hancock, 2020; Noyes et al., 2021). We also found that performance in the mask condition was predicted by performance in the full-view condition. Further analysis revealed that some observers were able to match faces with remarkably high sensitivity, even in the mask condition. Interestingly, as the performance in the full-view increased so did the full-view advantage. To avoid potential congruency effects (Kramer & Ritchie, 2016), in Experiment 2, the face mask were imposed to both faces. Despite this change, the results of this experiment replicated those of Experiment 1.

It has been previously suggested that an important part of the errors in face matching can be explained by insufficient facial information in the face pair (Alenezi & Bindemann, 2013; Fysh & Bindemann, 2017). However, it is noteworthy to highlight that despite covering approximately 50% of the face, some observers completed the task in the mask condition with remarkably high, although not perfect, performance. In fact, when compared to the full-view condition, some of these observers did not show any impairment in matching as consequence of the mask. This high performance suggests not only that the pair of faces contain enough information to achieve face matching, but also that the top part of the face could suffice for this task in most cases (Hills & Pake, 2013; Mckelvie, 1976; Megreya & Bindemann, 2018; Peterson & Eckstein, 2012; Sadr et al., 2003). Our results are in line with previous studies showing that some observers could match faces with consistent high accuracy, even under highly challenging visual conditions, such as profile view of faces (Estudillo & Bindemann, 2014), image blurring (Robertson et al., 2016) and inverted faces (White et al., 2015). Altogether these results suggest that, although data limits undoubtably affect matching accuracy, errors in face matching are better explained by resource limits, that is, observers fail to efficiently use facial information to match faces (Estudillo & Bindemann, 2014; Fysh & Bindemann, 2017).

We found that performance in the mask condition was predicted by performance in the full-view condition. Despite this relationship between the full-view and the mask conditions, analysis of the top 5% of performers in each condition revealed different patterns of performance across these two groups. Specifically, while observers in the top 5% full-view group performed better in the full-view condition compared to the mask condition, no such differences were found in the top 5% mask group. This pattern of individual differences between conditions, which has been previously reported (Bobak et al., 2016; Fysh et al., 2020; Stacchi et al., 2020), shows that some observers in the top end of the performance distribution present low generalization between different identification conditions (Fysh et al., 2020; Stacchi et al., 2020). Thus, during personnel selection procedures in forensic and security settings, it is important to combine different face identification conditions tapping different potential real-world identification scenarios (Fysh et al., 2020).

To quantify the impairment caused by the mask to face matching, we calculated the full-view advantage as the difference in performance between the full-view condition and the mask condition. A recent study has shown that the effect of the mask is reduced with inverted faces, which reflects, according to the authors, that face masks impair holistic processing (Freud et al., 2020). Thus, if this is assumed, the full-view advantage would reflect the contribution of holistic processing to face matching, with higher values indicating a stronger reliance in holistic processing. It has been argued that, in contrast to face recognition (Rossion, 2013; Wong et al., 2021), face matching might rely on a more featural processing mode (Megreya, 2018; Megreya & Bindemann, 2018; Megreya & Burton, 2006; Towler et al., 2017). Remarkably, we found that higher performance in face matching was associated with stronger mask effects, suggesting not only that holistic processing is important for face matching, but also that stronger holistic processing is associated with better performance in face matching. However, as in our study we did not include an alternative measure of holistic processing (i.e., inverted faces) as a control, we are cautious about this interpretation.

Importantly, even if it is assumed that face matching relies on holistic processing the fact that the top 5% performers in the mask condition generally matched faces with identical accuracy in both the full-view and the mask conditions reflects that featural processing can be important to achieve high performance in this task. This conclusion is supported by previous research showing that highly experienced face matching performers show a less prominent face inversion effect (White et al., 2015). These findings further reflect the importance of using different identification conditions tapping different cognitive process when evaluating face identification for personnel selection in applied scenarios (Fysh et al., 2020; Stacchi et al., 2020).

It must be noted that three different properties of our stimuli might have potentially contributed to our results. Firstly, it is possible that artificially superimposing the mask on our stimuli could have affected our results. In their recent study, Carragher and Hancock (2020) also superimposed the masks on the faces and, remarkably, our effect sizes are similar. In comparison, Noyes and colleagues (Noyes et al., 2021), who used real pictures of people wearing face masks, reported a smaller reduction in matching accuracy by face masks. Although, undoubtedly, face masks conceal shape information of the bottom part of the face, as in real life face masks are tightly fit to the face, it is possible that part of this shape information is retained in natural pictures of people wearing masks. Thus, superimposing the masks on face stimuli could indeed overstate the effect of face masks on face identification.

Secondly, it is possible that the high performance shown by some observers in the mask conditions could be consequence of the face database used in this study. In fact, the Glasgow Unfamiliar Face Database comprises high-quality frontal images that were taken under optimal lighting conditions. In addition, pictures from each model were taken during the same day, and present little variation in the external features of the face, such as hairstyle. These external features of the face seem to be particularly important for unfamiliar face identification, (Bruce et al., 1999; Clutterbuck & Johnston, 2002; Estudillo, 2012; Kemp et al., 2016), and although they are also available in the full-view condition, it is possible that the mask encourages observers to rely even more on these external features. With the current data, we cannot rule out this possibility, but as previously mentioned, recent research using ambient face images that present substantial variations in these external features, have shown that some observers still present high matching performance with masked faces (Noyes et al., 2021).

Thirdly, the face stimuli used in our two studies were Caucasian faces. As people tend to be worse recognizing other-ethnicity faces, the so-called other ethnicity-effect (Chiroro & Valentine, 1995; Estudillo et al., 2020; Malpass & Kravitz, 1969; Valentine et al., 2016; Wong et al., 2020), the inclusion of non-Caucasian observers in our sample (87 out of 204 in Experiment 2) could have affected our results. However, supplementary analysis showed that both Caucasian and non-Caucasian observers showed a clear full-view advantage (i.e., better performance in the full-view compared to the mask condition) and the magnitude of this effect was positively associated with the performance in the full-view condition (for details, see Supplementary Results).

In conclusion, while face masks impaired overall face matching performance, individual differences analysis revealed that some observers performed with a remarkably high accuracy even in the mask condition. This is such that, when compared to the full-view condition, some of these observers in the top end of the distribution did not show any impairment as consequence of the mask.

References

Alenezi, H. M., & Bindemann, M. (2013). The Effect of Feedback on Face-Matching Accuracy. *Applied Cognitive Psychology*, *27*, 735–753. https://doi.org/10.1002/acp.2968

Bindemann, M. (2021). *Forensic Face Matching: Research and Practice*. Oxford University Press.

Bindemann, M., Attard, J., Leach, A. M. Y., & Johnston, R. A. (2013). *The Effect of Image Pixelation on Unfamiliar-Face Matching*. https://doi.org/10.1002/acp.2970

Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology. Applied*, *18*(3), 277–291. https://doi.org/10.1037/a0029635

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016). *Super-recognisers in Action : Evidence from Face-matching and Face Memory Tasks*. *91*(October 2015), 81–91.

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Quarterly Journal of Experimental Psychology*, *72*, 872–881. https://doi.org/10.1177/1747021818776145

Bruce, V., Bindemann, M., & Lander, K. (2018). Individual differences in face perception and person recognition. *Cognitive Research: Principles and Implications*, *3*(1), 18. https://doi.org/10.1186/s41235-018-0109-4

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental*

Psychology: Applied, *5*(4), 339–360. https://doi.org/10.1037/1076-898X.5.4.339

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*(3), 207–218. https://doi.org/10.1037/1076-898X.7.3.207

Burton, M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

Carragher, D. J., & Hancock, P. J. B. (2020). Surgical face masks impair human face matching performance for familiar and unfamiliar faces. *Cognitive Research: Principles and Implications*, *5*(1), 59. https://doi.org/10.1186/s41235-020-00258-x

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. https://doi.org/10.1080/14640749508401421

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*(8), 985–994. https://doi.org/10.1068/p3335

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology*, *96*(1), 57–70. https://doi.org/10.1016/j.jecp.2006.07.001

Dhamecha, T. I., Singh, R., Vatsa, M., & Kumar, A. (2014). Recognizing disguised faces: Human and machine evaluation. *PLoS ONE*, *9*(7), 99212. https://doi.org/10.1371/journal.pone.0099212

Estudillo, A. J. (2012). Facial Memory: The Role of the Pre-Existing Knowledge in Face

Processing and Recognition. *Europe's Journal of Psychology*, *8*(2), 231–244. https://doi.org/10.5964/ejop.v8i2.455

Estudillo, A. J. (2021). Self-reported face recognition abilities for own and other-race faces. *Journal of Criminal Psychology*. https://doi.org/10.31234/osf.io/ru973

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *I-Perception*, *5*(7), 589–601. https://doi.org/10.1068/i0669

Estudillo, A. J., Lee, J. K. W., Mennie, N., & Burns, E. (2020). No evidence of other-race effect for Chinese faces in Malaysian non-Chinese population. *Applied Cognitive Psychology*, *34*, 270–276. https://doi.org/10.1002/acp.3609

Estudillo, A. J., & Wong, H. K. (2021). Associations between self-reported and objective face recognition abilities are only evident in above- and below-average recognisers. *PeerJ*, *9*, e10629. https://doi.org/10.7717/peerj.10629

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., & Ganel, T. (2020). The COVID-19 pandemic masks the way people perceive faces. *Scientific Reports*, *10*(1), 1–8. https://doi.org/10.1038/s41598-020-78986-9

Fysh, M., & Bindemann, M. (2017). Forensic face matching: a review. In A. M. Megreya & M. Bindemann (Eds.), *Face Processing: Systems, Dissorders and Cultural Differences* (pp. 1–20). Nova Science Publishing.

Fysh, M., & Bindemann, M. (2018). The Kent Face Matching Test. *British Journal of Psychology*, *109*(2), 219–231. https://doi.org/10.1111/bjop.12260

Fysh, M., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: implications for theory, research and personnel selection. *Royal Society Open Science*, *7*(9), 200233. https://doi.org/10.1098/rsos.200233

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

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. https://doi.org/10.1016/j.cognition.2013.08.012

Jensen, A. R. (2006). Clocking the Mind. In *Clocking the Mind*. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-044939-5.X5000-9

Kemp, R. I., Caon, A., Howard, M., & Brooks, K. R. (2016). Improving Unfamiliar Face Matching by Masking the External Facial Features. *Applied Cognitive Psychology*, *30*(4), 622–627. https://doi.org/10.1002/acp.3239

Kramer, R. S. S., & Reynolds, M. G. (2018). Unfamiliar Face Matching With Frontal and Profile Views. *Perception*, *47*(4), 414–431. https://doi.org/10.1177/0301006618756809

Kramer, R. S. S., & Ritchie, K. L. (2016). Disguising Superman: How Glasses Affect Unfamiliar Face Matching. *Applied Cognitive Psychology*, *30*(6), 841–845. https://doi.org/10.1002/acp.3261

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334. https://doi.org/10.1037/h0028434

Manley, K. D., Chan, J. C. K., & Wells, G. L. (2019). Do masked-face lineups facilitate eyewitness identification of a masked individual? *Journal of Experimental Psychology: Applied*, *25*(3), 396–409. https://doi.org/10.1037/xap0000195

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(1), 21. https://doi.org/10.1186/s41235-018-0112-9

Mckelvie, S. J. (1976). The Role of Eyes and Mouth in the Memory of a Face. *The American Journal of Psychology*, *89*(2), 311–323.

Megreya, A. M. (2018). Feature-by-feature comparison and holistic processing in unfamiliar face matching. *PeerJ, 2018*(2), e4437. https://doi.org/10.7717/peerj.4437

Megreya, A. M., & Bindemann, M. (2018). Feature instructions improve face-matching accuracy. *PLOS ONE*, *13*(3), e0193455. https://doi.org/10.1371/journal.pone.0193455

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876. https://doi.org/10.3758/BF03193433

Megreya, A. M., & Burton, A. M. (2008). Matching Faces to Photographs: Poor Performance in Eyewitness Memory (Without the Memory). *Journal of Experimental Psychology: Applied*, *14*(4), 364–372. https://doi.org/10.1037/a0013464

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, *31*(5), 553–566. https://doi.org/10.1068/p3339

Noyes, E., Davis, J. P., Petrov, N., Gray, K. L. H., & Ritchie, K. L. (2021). The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *Royal Society Open Science*, *8*(3), rsos.201169. https://doi.org/10.1098/rsos.201169

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., Albonico, A., Malaspina, M., Daini, R., Irons, J., Al-Janabi, S., Taylor, L. C., Rivolta, D., & McKone, E. (2017). Do People Have Insight into their Face Recognition Abilities? *Quarterly Journal of Experimental Psychology*, *70*(2), 218–233. https://doi.org/10.1080/17470218.2016.1161058

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), E3314-23. https://doi.org/10.1073/pnas.1214269109

Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology*, *110*(3), 461–479. https://doi.org/10.1111/bjop.12368

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological Science*, *22*(4), 464–471. https://doi.org/10.1177/0956797611401753

Robertson, D. J., Noyes, E., Dowsett, A., Jenkins, R., Burton, A. M., & Burton, M. (2016). Face recognition by Metropolitan Police Super-recognisers. *PLoS ONE, 11*, e0150036. https://doi.org/10.1371/journal.pone.0150036

Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, *32*(3), 285–293. https://doi.org/10.1068/p5027

Stacchi, L., Huguenin-Elie, E., Caldara, R., & Ramon, M. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive Research: Principles and Implications*, *5*(1), 8. https://doi.org/10.1186/s41235-019-0205-0

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Toseeb, U., Bryant, E. J., & Keeble, D. R. T. (2014). The Muslim Headscarf and Face Perception: "They All Look the Same, Don't They?" *PLoS ONE*, *9*(2), e84754. https://doi.org/10.1371/journal.pone.0084754

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47–58. https://doi.org/10.1037/xap0000108

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-Space: A Unifying Concept in Face Recognition Research. *Quarterly Journal of Experimental Psychology*, *69*(10), 1996–2019. https://doi.org/10.1080/17470218.2014.990392

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, *141*, 217–227. https://doi.org/10.1016/J.VISRES.2016.12.014

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169–177. https://doi.org/10.1177/0956797611420575

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814). https://doi.org/10.1098/rspb.2015.1292

White, D., Rivolta, D., Burton, A. M., Al-Janabi, S., & Palermo, R. (2016). Face matching impairment in developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology*, *0218*(May), 1–11. https://doi.org/10.1080/17470218.2016.1173076

Wong, H. K., Stephen, I. D., & Keeble, D. R. T. (2020). The Own-Race Bias for Face Recognition in a Multiracial Society. *Frontiers in Psychology*, *11*, 208. https://doi.org/10.3389/fpsyg.2020.00208