



PEARL

Multiple-image arrays in face matching tasks with and without memory

Ritchie, Kay L.; Kramer, Robin S.S.; Mileva, Mila; Sandford, Adam; Burton, A. Mike

Published in:
Cognition

DOI:
[10.1016/j.cognition.2021.104632](https://doi.org/10.1016/j.cognition.2021.104632)

Publication date:
2021

Link:
[Link to publication in PEARL](#)

Citation for published version (APA):

Ritchie, K. L., Kramer, R. S. S., Mileva, M., Sandford, A., & Burton, A. M. (2021). Multiple-image arrays in face matching tasks with and without memory. *Cognition*, 211(0). <https://doi.org/10.1016/j.cognition.2021.104632>

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Wherever possible please cite the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

1 Multiple-image arrays in face matching tasks with and without memory

2

3 Kay L. Ritchie¹, Robin S. S. Kramer¹, Mila Mileva², Adam Sandford³, and A. Mike Burton²

4

5 ¹ School of Psychology, University of Lincoln, UK

6 ² Department of Psychology, University of York, UK

7 ³ Department of Psychology, University of Guelph-Humber, Canada

8

9 **Corresponding author**

10 KL Ritchie

11 kritchie@lincoln.ac.uk

12 School of Psychology,

13 University of Lincoln

14 Lincoln

15 LN6 7TS

16

17 **Acknowledgements**

18 The authors would like to thank Andrew Dowsett for contributing to the work presented here,

19 Amy S. Hought for data collection for Experiment 2, Ellen Wheeler for data collection for

20 Experiment 3, and Lily Bridgewater for data collection for Experiments 4.

21

22 **Keywords**

23 Face matching; face learning; variability

24

1 Abstract

2 Previous research has shown that exposure to within-person variability facilitates face
3 learning. A different body of work has examined potential benefits of providing multiple
4 images in face matching tasks. Viewers are asked to judge whether a target face matches a
5 single face image (as when checking photo-ID) or multiple face images of the same person.
6 The evidence here is less clear, with some studies finding a small multiple-image benefit, and
7 others finding no advantage. In four experiments, we address this discrepancy in the benefits
8 of multiple images from learning and matching studies. We show that multiple-image arrays
9 only facilitate face matching when arrays precede targets. Unlike simultaneous face matching
10 tasks, sequential matching and learning tasks involve memory and require abstraction of a
11 stable representation of the face from the array, for subsequent comparison with a target. Our
12 results show that benefits from multiple-image arrays occur only when this abstraction is
13 required, and not when array and target images are available at once. These studies reconcile
14 apparent differences between face learning and face matching and provide a theoretical
15 framework for the study of within-person variability in face perception.

1 **Introduction**

2 We rely on faces to verify identity in a variety of situations ranging from buying alcohol to
3 crossing borders. It is, therefore, important to understand how accurate we are at determining
4 whether a photo-ID shows the person using it, and to identify potential ways to improve our
5 performance in such tasks.

6
7 A large body of literature suggests that recognising familiar and unfamiliar faces entail some
8 qualitatively different processes (Johnston & Edmonds, 2009; Megreya & Burton, 2006) and
9 this could have serious practical implications. On the one hand, we are very good at
10 recognising images of familiar identities even when these images are heavily distorted or
11 degraded (e.g. Bruce, 1982, 1986; Burton, Wilson, Cowan, & Bruce, 1999). On the other
12 hand, recognition of unfamiliar identities is much poorer even with images taken on the same
13 day or in the same session (e.g. Bruce et al, 1999; Burton, White, & McNeill, 2010). This
14 stark contrast between familiar and unfamiliar faces has been demonstrated using many
15 different tasks and paradigms including face memory, search and sorting tasks (Jenkins,
16 White, Van Montfort, & Burton, 2011; Klatzky & Forrest, 1984; Kramer, Hardy, & Ritchie,
17 2020) as well as face matching tasks where typically two images are presented side-by-side
18 on a computer screen, and participants are asked to judge whether the photos show the same
19 person or different people (Bruce, Henderson, Newman, & Burton, 2001; Bruce et al., 1999;
20 Clutterbuck & Johnston, 2002, 2004; Megreya & Burton, 2008; Ritchie et al., 2015). While
21 matching tasks have been generally used to approximate the process of checking photo-ID,
22 the effect of familiarity has also been documented outside the lab with findings of poor
23 performance when matching a live unfamiliar person to a photograph (Davis & Valentine,
24 2009; Kemp, Towell, & Pike, 1997; Megreya & Burton, 2008; Ritchie, Mireku, & Kramer,
25 2020). Moreover, all of this is true for many people who are employed to check photo-ID
26 such as checkout assistants (Kemp et al., 1997), passport officers (White, Kemp, Jenkins,
27 Matheson, & Burton, 2014) and police officers (Burton et al., 1999).

28
29 The difference between recognising familiar and unfamiliar faces has been attributed to the
30 types of processing involved. We have seen the faces of familiar identities in a variety of
31 contexts, situations and conditions, providing us with rich information about the many ways a
32 single person might look. This way, we are able to isolate everything that is diagnostic of the
33 person and discard any superficial image differences, leading to a more abstracted and image-
34 independent processing for familiar faces. In Bruce and Young's influential model (1986),

1 familiar recognition is conceptualised through the use of Face Recognition Units (FRUs)
2 which code structural information about known faces. FRUs must therefore store an
3 abstracted, stable representation of a familiar person that is not influenced by simple image
4 properties such as changes in head angle or expression.

5
6 Bruce (1994) first introduced the notion of stability from variation as a key familiarisation
7 mechanism. Since then, a number of behavioural and computer modelling studies have shown
8 that we can create and store stable representations of faces through exposure to within-person
9 variability – that is multiple exposures to the same person showing naturally-occurring
10 changes in their appearance. However, the same natural within-person variability that aids the
11 recognition of familiar faces, can be detrimental to unfamiliar recognition which relies to a
12 much greater extent on superficial image properties. This means that irrelevant differences in
13 the physical properties of images or simple changes in clothing or accessories can be
14 mistakenly regarded as evidence for differences in identity (Bindemann & Sandford, 2011;
15 Graham & Ritchie, 2019; Kramer & Ritchie, 2016). In fact, recent research has suggested that
16 the difference between familiar and unfamiliar face recognition may be due to our ability to
17 use or tolerate within-person variability for familiar people (Burton, 2013; Burton, Jenkins &
18 Schweinberger, 2011; Burton, Kramer, Ritchie, & Jenkins, 2016; Jenkins, White, Van
19 Montfort, & Burton, 2011). It is therefore possible that exposure to this variability can help
20 unfamiliar viewers to switch from image-based to a more abstracted processing by
21 aggregating the variability information into a single identity representation.

22
23 A growing body of research has shown that exposure to within-person variability helps when
24 learning a new identity and this has been supported by work using both behavioural and
25 computer modelling data (Dowsett, Sandford, & Burton, 2016; Jones, Dwyer, & Lewis, 2017;
26 Kramer, Young, & Burton, 2018; Longmore, Liu & Young, 2008; Longmore et al., 2017;
27 Matthews, Davis, & Mondloch, 2018; Murphy, Ipser, Gaigg, & Cook, 2015; Ritchie &
28 Burton, 2017; Robins, Susilo, Ritchie, & Devue, 2018). The benefits from access to multiple
29 images of the same identity have been shown in adults' as well as in children's face learning
30 (Matthews et al., 2018), with some evidence that children aged 6-13 need more variability
31 than adults to learn a new person from video footage (Baker, Laurence, & Mondloch, 2017).

32
33 The amount of within-person variability is also an important factor in face learning. Ritchie
34 and Burton (2017), for example, showed participants photos that were either high in

1 variability (displaying changes in head angle, lighting, camera, age, hair style, etc) from a
2 Google Images search, or photos that were low in variability, taken from a video of a single
3 event (changes only in head angle and expression). After learning the identities from these
4 images, participants' performance was tested with a name-verification and a face matching
5 task using novel images of the same identities. In both cases, participants who had learned
6 from the high variability image set outperformed those who learned from the low variability
7 set. These results suggest that exposure to variability is key to learning or abstracting a stable
8 representation of a person.

9

10 Research on the benefits of within-person variability for face matching has been less
11 consistent and conclusive. Unlike face learning, this is a purely perceptual task with no
12 demands on memory. Some studies suggest that multiple images may help to improve
13 performance on face matching. White et al. (2014) presented participants with arrays of two,
14 three, or four images of the same person and asked them to match another image to the array.
15 The multiple-image arrays gave rise to better performance than matching to a single image.
16 In a different paradigm, participants were presented with a physical photograph of a target
17 and asked to search through a pile of photos to find another image of the same person. On
18 successive trials, participants were given an additional image of the same identity and their
19 accuracy improved as the number of target images increased (Dowsett, Sandford, & Burton,
20 2016). Other recent studies, however, have failed to replicate these results with no benefits
21 reported from exposure to arrays comprising a frontal and a profile view image (Kramer &
22 Reynolds, 2018) or when matching a live person to a four-image array compared to a single
23 image (Ritchie et al., 2020).

24

25 Therefore, when it comes to the key role of within-person variability for successful
26 recognition, face *learning* and face *matching* tasks present somewhat dissimilar results.
27 Exposure to variability helps learning a new identity, whereas results with matching are
28 unclear. One possible explanation for this difference is that learning paradigms require the
29 face to be memorised whereas matching paradigms present all stimuli simultaneously,
30 without a memory component to the task. It is thus possible that exposure to variability, or
31 multiple images, is only helpful when the task requires that a representation of the face be
32 abstracted in order to be held in memory to make subsequent comparisons. This is supported
33 by evidence for the benefits of within-person variability in face matching when images are
34 presented one after the other, rather than simultaneously (Menon, White, & Kemp, 2015a).

1
2 Here, we compare face recognition accuracy in a purely perceptual simultaneous matching
3 task and a memory-dependent sequential matching task. In a series of four studies, we
4 manipulate the amount of within-person variability available, and the presentation order of
5 multiple image arrays and comparison images, allowing us to determine why variability
6 seems to be consistently aiding face learning but not face matching performance. It is
7 possible that differences in results between previous studies are due to a difference in the
8 amount of within-person variability shown in the arrays, with studies that have found a
9 multiple-image benefit (e.g. White et al., 2014) perhaps displaying more variability in the
10 arrays than those that have not found that effect (e.g. Ritchie et al., 2020). However, if the
11 difference in the utility of variability between face learning and matching is due to the
12 memory component of learning tasks, then we would expect variability to facilitate
13 performance in only sequential matching tasks. Like learning tasks, sequential matching tasks
14 may require variability to be incorporated into a stable identity representation.

15
16 Experiment 1 investigates the effect of array variability on face matching performance in a
17 simultaneous task. Experiment 2 compares performance in simultaneous versus sequential
18 matching tasks. Finally, Experiments 3 and 4 compare performance on two different
19 sequential tasks – one that allows for variability to be integrated into a single mental
20 representation and one that does not.

21

22 **Experiment 1 – array variability**

23 The evidence to date is mixed as to whether multiple images improve matching performance
24 (Menon et al., 2015; Ritchie et al., 2020; Sandford & Ritchie, under review; White et al.,
25 2014), and so it could be that these experiments used arrays of differing degrees of
26 variability, resulting in different effects. In this first experiment, we investigated the effect of
27 array variability on face matching performance. We constructed high and low variability
28 arrays from an existing image set (Ritchie & Burton, 2017). Participants compared a target
29 image to either a high or low variability array, and we also tested accuracy in a one-to-one
30 condition. It is possible that multiple-image arrays only facilitate face matching when the
31 arrays are high in variability.

32

33 **Method**

34 *Participants*

1 Thirty-one participants took part in this experiment (7 male, mean age: 22 years, range: 17-40
2 years). All participants were students or other members of the University of York. All
3 participants gave informed consent, and the study was granted ethical approval by the
4 University of York Psychology Ethics Committee.

5

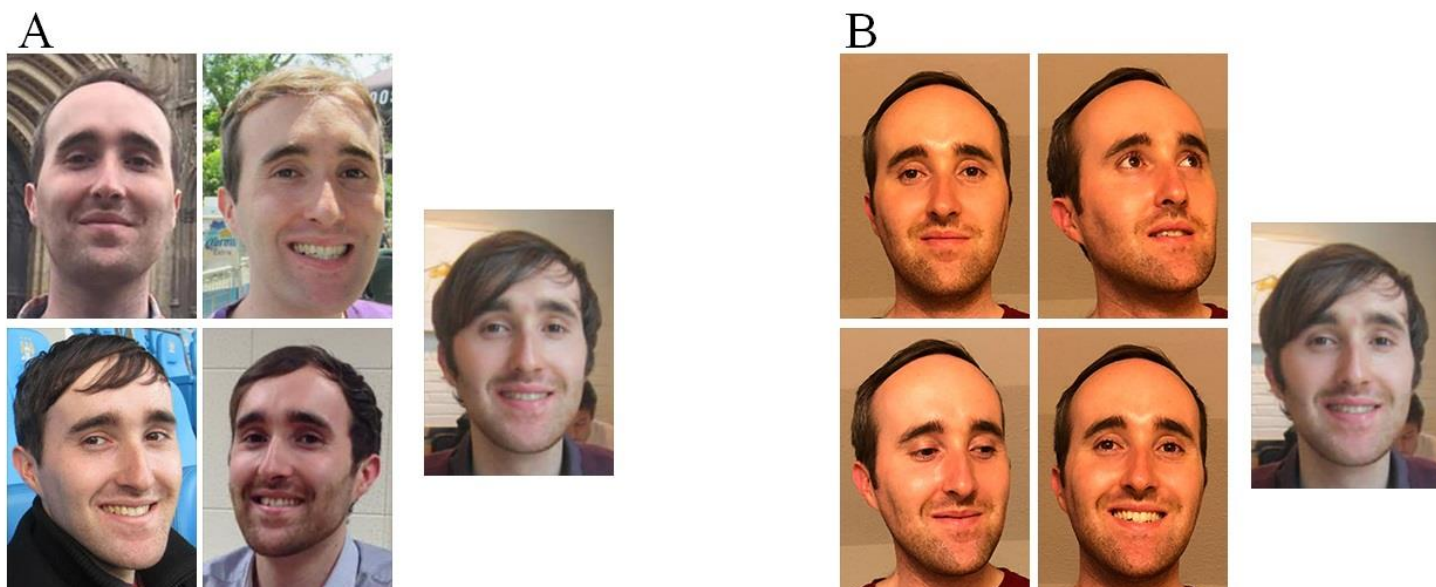
6 *Stimuli and Procedure*

7 The stimuli were images from a previous set of high and low variability ambient images used
8 for face learning research (Ritchie & Burton, 2017). The set comprised five high and four low
9 variability images of each of ten Australian celebrities (five female), specifically chosen to be
10 unfamiliar in the UK. The high variability images were downloaded from a Google Images
11 search for each identity and varied in head angle, expression, lighting, age, etc. The low
12 variability images were screenshots from single interview videos, allowing for variation in
13 head angle and expression, but now taken seconds apart under the same lighting and with the
14 same camera (see Figure 1).

15

16 For the matching task, we constructed four-image arrays from the high variability (Google
17 Images) images, and the low variability (video screenshots) images. All four images in each
18 array always showed the same person. In half of the face matching trials, participants were
19 presented with two images side by side on the screen (one-to-one condition). For half of the
20 one-to-one trials, the image on the left of the screen was from the high variability set, and
21 from the low variability set in the other half of the trials. The image on the right was either a
22 match (an image from the high variability set showing the same identity) or a mismatch (a
23 foil image showing a different identity that matched the verbal description of the target
24 identity, e.g., young man, dark hair). In the other half of the trials, participants were presented
25 with a four-image array paired with either a match or a foil image (four-to-one condition).
26 The multiple-image array was from the high variability set in one half of the trials and from
27 the low variability set in the other. It was always presented on the left of the screen (see
28 Figure 1), and participants were informed in the four-to-one condition that these four images
29 showed the same person. The comparison (match or mismatch) image was presented on the
30 right and participants were prompted with on-screen instructions to respond via keypresses to
31 indicate whether the comparison image showed the same person as displayed on the left of
32 the screen. Each participant completed a total of 40 trials – 20 in the one-to-one condition
33 (half with a high variability image, half with a match image) and 20 in the four-to-one

1 condition (half with a high variability image array, half with a match image). Each identity
 2 was seen once in each condition (high/low variability, one/four images, match/mismatch).



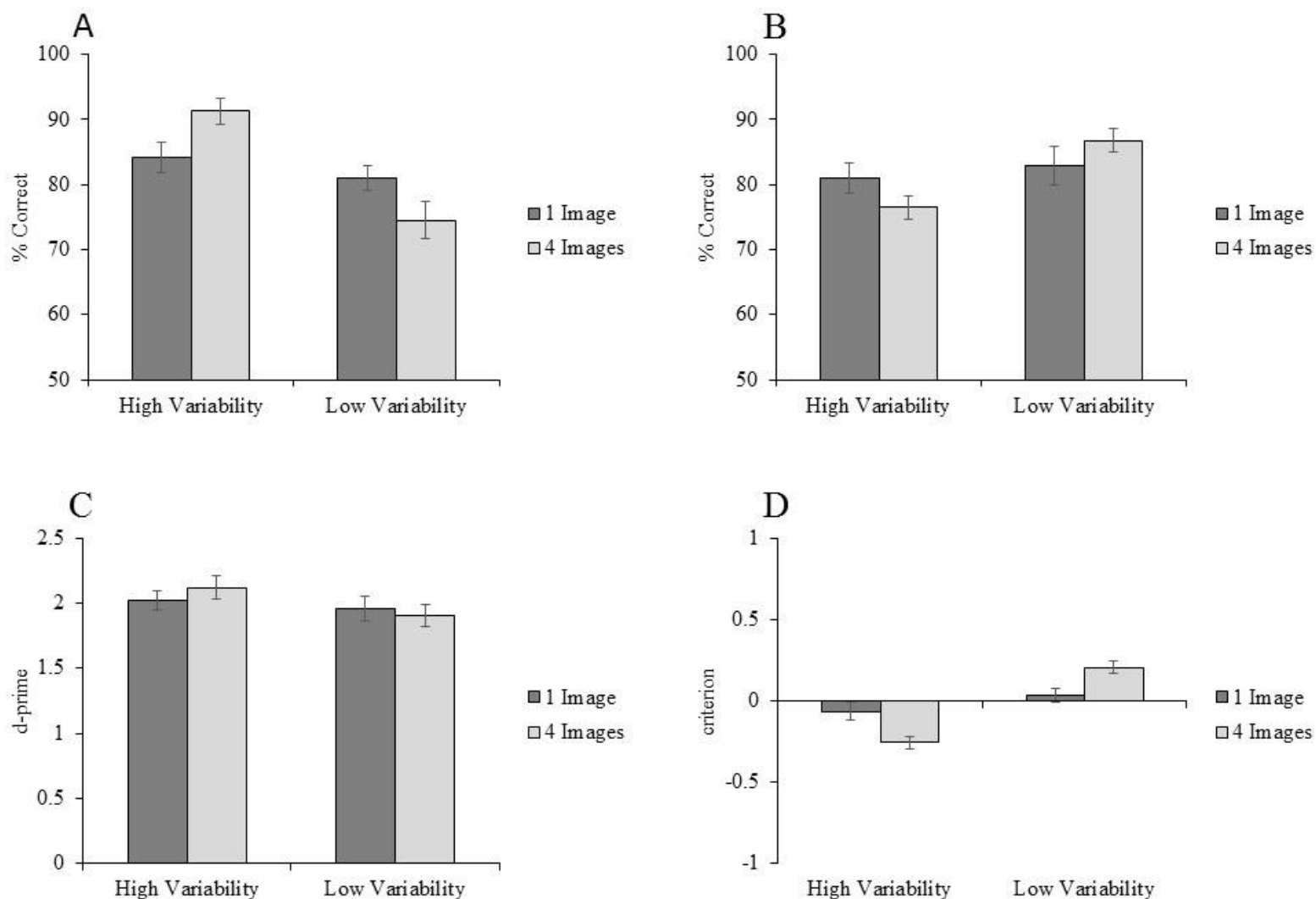
4 **Figure 1.** Example stimuli used in Experiment 1. A) High variability array match trial. B)
 5 Low variability array match trial. (Copyright restrictions prevent publication of the images
 6 used in the experiment. Images in Figure 1, also in Figures 3 and 6, are illustrative of the
 7 experimental stimuli and depict someone who did not appear in the experiments but has given
 8 permission for the images to be reproduced here).

9

10 **Results**

11 Previous research has found that performance on match and mismatch trials is not correlated
 12 (Megreya & Burton, 2007), and many studies have found that experimental manipulations
 13 affect performance on match *or* mismatch trials, but rarely both (e.g. Megreya & Burton,
 14 2006, 2007; Menon et al., 2015; Ritchie & Burton, 2017; White et al., 2014). Therefore, for
 15 all experiments reported here, match and mismatch trial accuracy are analysed separately. In
 16 addition to traditional frequentist hypothesis testing, we included Bayes factors using JASP
 17 (JASP Team, 2020), which allowed us to quantify the extent to which the data support the
 18 alternative hypothesis (BF_{10}). Bayes factors for the simple main effects analyses do not take
 19 into account the full ANOVA, and so indicate simple strengths of differences between
 20 conditions. Mean accuracy for Experiment 1 is shown in Figure 2. No participants in any of
 21 the experiments indicated familiarity with any of the stimulus identities.

22



2 **Figure 2.** Data for Experiment 1. A) Match trials. B) Mismatch trials. C) d-prime. D)
 3 criterion. Error bars show the within-subjects standard error (Cousineau, 2005).

4
 5 First, for match trials, a 2 (variability: high, low) x 2 (number of images: 1, 4) within subjects
 6 ANOVA showed a significant main effect of variability $F(1,30) = 32.63, p < .001, \eta_p^2 = .52,$
 7 $BF_{10} = 9.46,$ a non-significant main effect of number of images $F(1,30) = 0.02, p = .888,$
 8 $\eta_p^2 < .01, BF_{10} = 0.74,$ and a significant interaction $F(1,30) = 9.22, p = .005, \eta_p^2 = .24$
 9 $BF_{10} = 0.28.$ Simple main effects showed a significant improvement in performance with high
 10 compared to low variability images for four-image arrays $F(1,60) = 35.00, p < .001, \eta_p^2 = .37$
 11 $BF_{10} = 17,014.43,$ but not for one-to-one match trials $F(1,60) = 1.29, p = .261, \eta_p^2 = .02,$
 12 $BF_{10} = 0.46,$ meaning there was no difference in matching performance when the single
 13 comparison image came from the high or the low variability set of images. Simple main

1 effects also showed an effect of number of images for both high variability images
 2 $F(1,60) = 5.00, p = .029, \eta_p^2 = .08, BF_{10} = 2.08$, and low variability images $F(1,60) = 4.13,$
 3 $p = .047, \eta_p^2 = .06, BF_{10} = 1.16$, such that four-image arrays helped when they were high in
 4 variability (1 image $M = 84.19\%$, 4 images $M = 91.29\%$), but hindered when they were low
 5 in variability (1 image $M = 80.97\%$, 4 images $M = 74.52\%$).

6
 7 For mismatch trials, there was a significant main effect of variability $F(1,30) = 7.54,$
 8 $p = .010, \eta_p^2 = .20, BF_{10} = 9.47$, a non-significant main effect of number of images
 9 $F(1,30) = 0.02, p = .888, \eta_p^2 < .01, BF_{10} = 0.19$, and a significant interaction $F(1,30) = 6.95,$
 10 $p = .013, \eta_p^2 = .19, BF_{10} = 1.72$. Simple main effects showed a significant effect of variability
 11 for four-image arrays $F(1,60) = 14.18, p < .001, \eta_p^2 = .19, BF_{10} = 30.32$, where accuracy was
 12 poorer with high variability image arrays ($M = 76.45\%$) than with low variability arrays
 13 ($M = 86.77\%$). There was a non-significant effect of variability for one-to-one match trials
 14 $F(1,60) = .05, p = .824, \eta_p^2 = .03, BF_{10} = 0.33$ meaning there was no difference in
 15 performance on mismatch trials when the single comparison image came from the high or the
 16 low variability set of images. The simple main effects for number of images for the high and
 17 low variability images were both non-significant (both $ps > .05$, both $BF_{10} < 1$; see Figure 2).

18
 19 We can also analyse the data using signal detection measures. Here, hits correspond to correct
 20 match trials, and false alarms to incorrect mismatch trials (see Figure 2, lower panels). For d-
 21 prime (d') values, there was a non-significant main effect of variability $F(1,30) = 1.79,$
 22 $p = .192, \eta_p^2 = .06, BF_{10} = 0.46$, a non-significant main effect of number of images
 23 $F(1,30) = 0.05, p = .831, \eta_p^2 < .01, BF_{10} = 0.19$, and a non-significant interaction
 24 $F(1,30) = 0.57, p = .457, \eta_p^2 = .02, BF_{10} = 0.29$. For criterion values (a measure of bias), there
 25 was a significant main effect of variability $F(1,30) = 43.90, p < .001, \eta_p^2 = .59,$
 26 $BF_{10} = 22,278.74$, a non-significant main effect of number of images $F(1,30) = 0.05,$
 27 $p = .825, \eta_p^2 < .01, BF_{10} = 0.19$, and a significant interaction $F(1,30) = 14.47, p < .001,$
 28 $\eta_p^2 = .33, BF_{10} = 63.71$. Simple main effects showed a significant effect of number of images
 29 for high variability images $F(1,60) = 6.67, p = .012, \eta_p^2 = .10, BF_{10} = 2.39$, whereby
 30 participants were more biased toward responding “match” with four high variability images
 31 ($M = -.26$) than one image ($M = -.06$). Simple main effects also showed a significant effect of
 32 number of images for low variability images $F(1,60) = 4.98, p = .029, \eta_p^2 = .08, BF_{10} = 4.42,$
 33 whereby participants were more biased toward responding “mismatch” with four low
 34 variability images ($M = .20$) than one image ($M = .04$).

1
2 In this experiment, using four-image arrays, we have shown that for match trials, high
3 variability arrays improve performance, and low variability arrays impair performance, as
4 compared to one-to-one matching trials. For mismatch trials, however, high variability arrays
5 impaired performance compared to low variability arrays, and there was no benefit for high
6 variability four-image arrays over single images. Taken together, these results suggest that
7 there is no evidence for an increase in overall accuracy (match and mismatch trials taken
8 together as in d') with multiple-image arrays. These results are aligned with two recent
9 studies which showed no overall benefit of variability when the array and the target are
10 presented simultaneously (Ritchie et al., 2020; Sandford & Ritchie, under review).

11 12 **Experiment 2 – simultaneous vs sequential matching**

13 This experiment investigated the effect of four-image arrays in simultaneous and sequential
14 matching. The simultaneous and sequential tasks have different task demands, being purely
15 perceptual- and memory-based respectively. This allows us to investigate the effect of
16 variability on these two different processes. If memory is important for the multiple-image
17 advantage, then we should see that four-image arrays produce higher matching accuracy only
18 in a sequential and not a simultaneous matching task.

19 20 **Method**

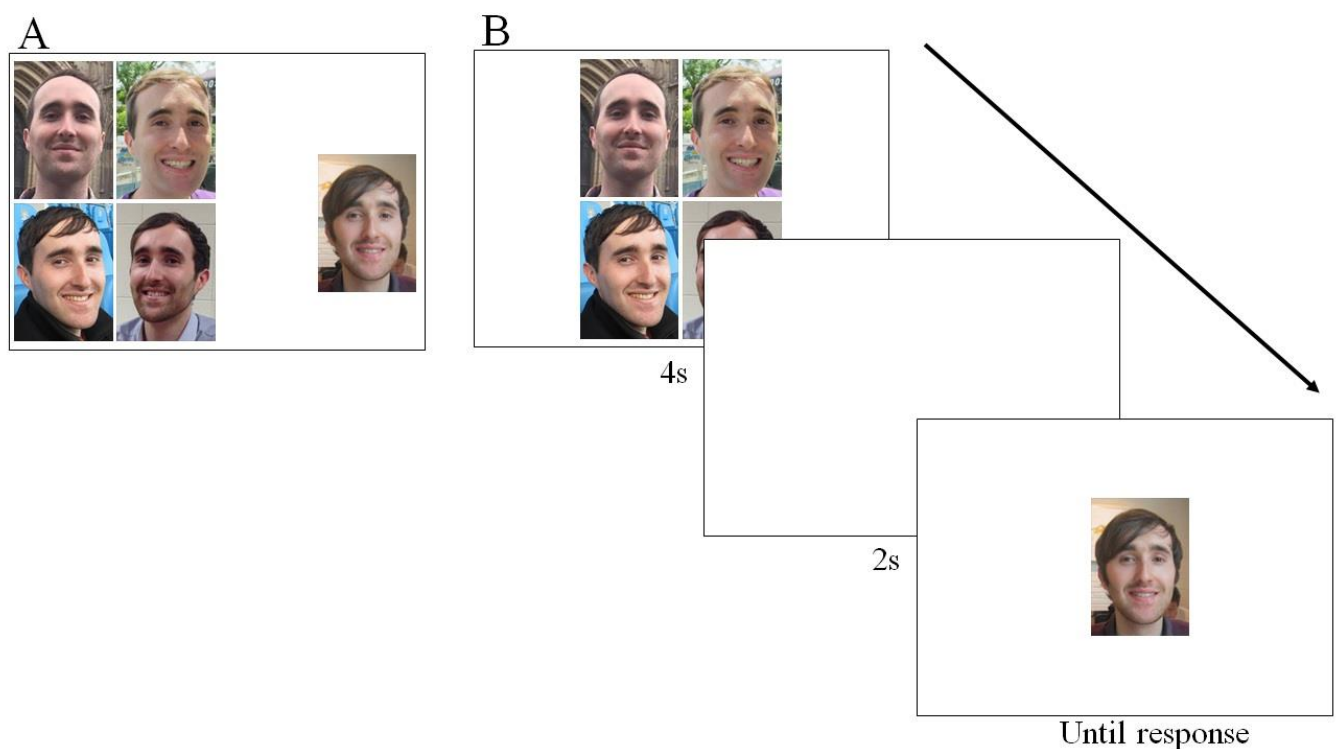
21 *Participants*

22 Forty participants took part (6 male, mean age: 20 years, range: 18-27 years). All participants
23 were students or other members of the University of Lincoln. All participants gave informed
24 consent, and the study was granted ethical approval by the University of Lincoln School of
25 Psychology Research Ethics Committee.

26 27 *Stimuli and procedure*

28 The stimuli here were of 80 identities (40 female), including the Australian celebrities used in
29 Experiment 1. All images were high in variability. As in Experiment 1, we used a matching
30 task with either a single image or a four-image array which were paired with an image of the
31 same identity in match trials and with an image of a foil identity in mismatch trials. Each
32 identity was presented once, with a random assignment of identities to conditions across
33 participants (keeping the number of males and females in each condition equal). Each
34 participant completed two separate face matching blocks - one simultaneous and one

1 sequential. For the simultaneous block, the array or single target image was presented on the
 2 left with the comparison image (match or foil) on the right (as in Experiment 1). In the
 3 sequential block, the target image/array was displayed first for 4 s, followed by a blank
 4 screen for 2 s, followed by the comparison (match or foil) image which remained on screen
 5 until the participant responded (see Figure 3). All images and arrays were centred on the
 6 screen for the sequential procedure. The order of blocks was counterbalanced across
 7 participants. Each block contained 40 trials: 10 single image match, 10 single image
 8 mismatch, 10 four-image array match, 10 four-image array mismatch. For the simultaneous
 9 block, the array or single target image was presented on the left with the comparison image
 10 (match or foil) on the right (as in Experiment 1). In the sequential block, the target
 11 image/array was displayed first for 4 s, followed by a blank screen for 2 s, followed by the
 12 comparison (match or foil) image which remained on screen until the participant responded
 13 (see Figure 3). All images and arrays were centred on the screen for the sequential procedure.
 14

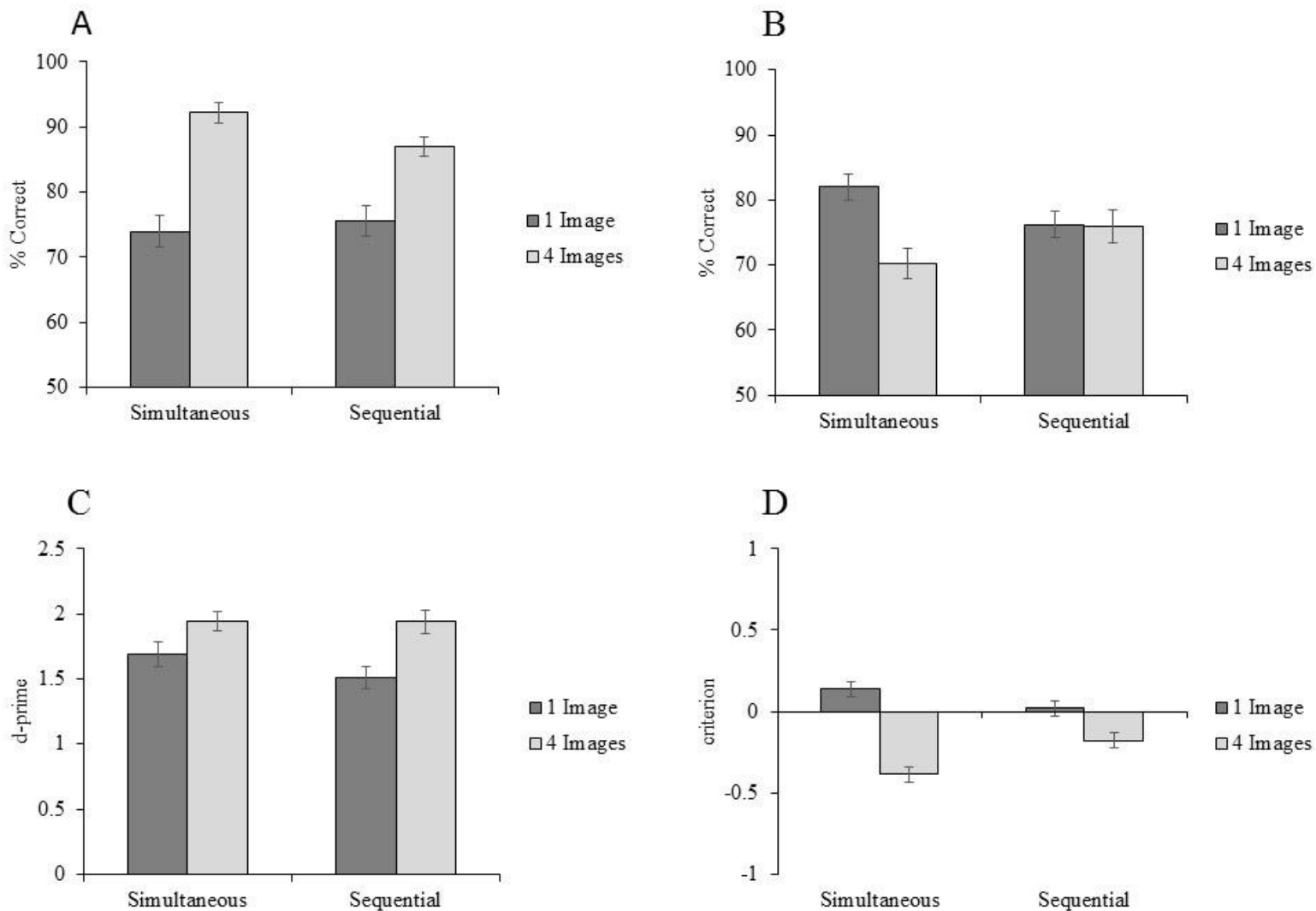


15
 16 **Figure 3.** Procedure for Experiment 2. A) Simultaneous match trial. B) Sequential match trial.

17

18 **Results**

1 Here, we analysed the data using a 2 (number of images: 1,4) x 2 (presentation type:
 2 simultaneous, sequential) ANOVA separately for match and mismatch trials. Figure 4 shows
 3 the results of Experiment 2.



5 **Figure 4.** Results of Experiment 2 using single images and four-image arrays in both
 6 simultaneous and sequential face matching tasks. A) Match trials. B) Mismatch trials. C) d-
 7 prime. D) criterion. Error bars show the within-subjects standard error (Cousineau, 2005).

8

9 For match trials, there was a significant main effect of number of images $F(1,39) = 68.70$,
 10 $p < .001$, $\eta_p^2 = .64$, $BF_{10} = 3.75 \times 10^9$, with significantly higher accuracy for four images
 11 ($M = 89.63\%$) than one image ($M = 74.75\%$). There was a non-significant main effect of
 12 presentation type $F(1,39) = 0.92$, $p = .343$, $\eta_p^2 = .02$, $BF_{10} = 0.23$, and a non-significant
 13 interaction $F(1,39) = 2.30$, $p = .137$, $\eta_p^2 = .06$, $BF_{10} = 0.61$.

1
 2 For mismatch trials, there was a significant main effect of number of images $F(1,39) = 5.55$,
 3 $p = .024$, $\eta_p^2 = .12$, $BF_{10} = 3.39$, a non-significant main effect of presentation type
 4 $F(1,39) < 0.001$, $p = .975$, $\eta_p^2 < .01$, $BF_{10} = 0.17$, and a significant interaction $F(1,39) = 4.85$,
 5 $p = .034$, $\eta_p^2 = .11$, $BF_{10} = 3.59$. Simple main effects showed a significant effect of number of
 6 images for simultaneous $F(1,78) = 10.39$, $p = .002$, $\eta_p^2 = .12$, $BF_{10} = 60.53$ but not sequential
 7 trials $F(1,78) < 0.001$, $p = .951$, $\eta_p^2 < .001$, $BF_{10} = 0.23$. For mismatch trials, performance
 8 was higher with one image ($M = 82.00\%$) compared to four images ($M = 70.25\%$). As in
 9 Experiment 1, we see that four images help for match trials but harm for mismatch trials
 10 when presented simultaneously with the target image. In contrast, when the array is shown
 11 prior to the comparison image, we see an advantage for four images in match trials without
 12 the accompanying decrease in performance in mismatch trials.

13
 14 Signal detection analyses showed a similar pattern of results. A 2 (number of images: 1, 4) x
 15 2 (presentation type: simultaneous, sequential) ANOVA on d' values showed a significant
 16 main effect of number of images $F(1,39) = 11.24$, $p = .002$, $\eta_p^2 = .22$, $BF_{10} = 48.61$, a non-
 17 significant main effect of presentation type $F(1,39) = 0.95$, $p = .336$, $\eta_p^2 = .02$, $BF_{10} = 0.26$
 18 and a non-significant interaction $F(1,39) = 0.83$, $p = .368$, $\eta_p^2 = .02$, $BF_{10} = 0.57$. d' values
 19 were higher with four images ($M = 1.94$) than one image ($M = 1.60$).

20
 21 For criterion values, there was a main effect of number of images $F(1,39) = 43.82$, $p < .001$,
 22 $\eta_p^2 = .53$, $BF_{10} = 2.69 \times 10^7$, a non-significant main effect of presentation type $F(1,39) = 1.05$,
 23 $p = .312$, $\eta_p^2 = .03$, $BF_{10} = 0.21$ and a significant interaction $F(1,39) = 6.96$, $p = .012$,
 24 $\eta_p^2 = .15$, $BF_{10} = 16.47$. Simple main effects showed an effect of number of images at both
 25 simultaneous $F(1,78) = 40.28$, $p < .001$, $\eta_p^2 = .34$, $BF_{10} = 4.02 \times 10^7$ and sequential trial types
 26 $F(1,78) = 5.64$, $p = .020$, $\eta_p^2 = .07$, $BF_{10} = 4.14$ whereby participants were more likely to
 27 respond “match” to four compared with one image. Simple main effects also showed a
 28 significant effect of presentation style only for four images $F(1,78) = 7.57$, $p = .007$, $\eta_p^2 = .09$
 29 whereby participants were more likely to respond “match” for the simultaneous presentation
 30 ($M = -.12$) than the sequential presentation ($M = -.08$).

31
 32 Paired samples t-tests were run to analyse performance in the single image condition. There
 33 was a non-significant difference between one-to-one matching performance for both match
 34 ($t(39) = 0.41$, $p = 1$, $BF_{10} = 0.18$) and mismatch trials ($t(39) = 2.06$, $p = .092$, although the

1 Bayes factor provided some evidence for a difference $BF_{10} = 1.14$). Therefore, in this
2 experiment, there was no detrimental effect of presenting the one-to-one matching task
3 sequentially as opposed to simultaneously.

4
5 This experiment has demonstrated that four-image arrays presented simultaneously with the
6 comparison image produce a benefit on match trials but a cost on mismatch trials, whereas
7 four-image arrays presented sequentially with (before) the target image produce the benefit at
8 match trials without the corresponding deficit at mismatch trials. This suggests that the
9 multiple-image benefit can be found when there is a substantial memory component to the
10 task (sequential) but not when the task is purely perceptual (simultaneous). Of course, even
11 the simultaneous matching task requires some memory as participants look from the array
12 images to the target image, engaging visual short-term memory (e.g. Henderson, Pollatsek, &
13 Rayner, 1987). Here we mean that our sequential matching task has a more substantial
14 memory component lasting seconds, as opposed to the milliseconds it takes to execute an eye
15 movement. Therefore, we argue that our sequential matching task engages memory and
16 forces the abstraction of a representation of the face as in a learning task.

17
18 Our results are in accord with research on face learning (e.g. Murphy et al., 2015; Ritchie &
19 Burton, 2017; Robins et al., 2018) which has shown an advantage of seeing multiple variable
20 images when learning a new identity. In the learning paradigms, there is a delay between
21 learning and test, meaning that participants must extract a representation of the identity and
22 store that, in order for it to be compared to subsequent images during the test phase. If the
23 variability advantage is due to the memory component of the task, this explains why we do
24 not find a variability advantage both in Experiment 1 and in the simultaneous condition of
25 this experiment, and elsewhere (Ritchie et al., 2020), as the simultaneous matching task does
26 not have a memory component. If memory is key to the variability advantage, then we should
27 only see the benefit of the multiple-image array when the array is presented first in the
28 sequence, before the target image, and not after. We address this in our final two
29 experiments.

30
31 **Experiment 3 – sequential presentation varying the order of array and comparison**
32 **image**

33

1 We hypothesise that the variability advantage found above in a sequential matching task
2 relies on the task having a memory component. This is also the case for the variability
3 advantage found elsewhere in the face learning literature, as learning tasks require memory.
4 If this is the case, then we should see this advantage only when we present the array first in a
5 sequential matching task as this will require participants to abstract a unified identity
6 representation from the variability that is inherent in different images of the same person.
7 Therefore, in this experiment, we vary the order of the target image/array and the comparison
8 (match/foil) image.

9

10 **Method**

11 *Participants*

12 Fifty new participants took part (12 male, mean age: 23 years, range: 18-61 years). All
13 participants were students or other members of the University of Lincoln. All participants
14 gave informed consent, and the study was granted ethical approval by the University of
15 Lincoln School of Psychology Research Ethics Committee.

16

17 *Stimuli and procedure*

18 The stimuli were images of a new set of 80 identities. These were celebrities from different
19 countries, specifically chosen to be unfamiliar to our participants in the UK. For each
20 identity, we downloaded five images and one image of a foil identity from Google Images.
21 For the four-image array conditions, the four images were randomly picked from the five
22 images of the identity, with the remaining image used as the match comparison image. The
23 80 identities were randomly assigned to conditions, and each identity was seen only once by
24 each participant.

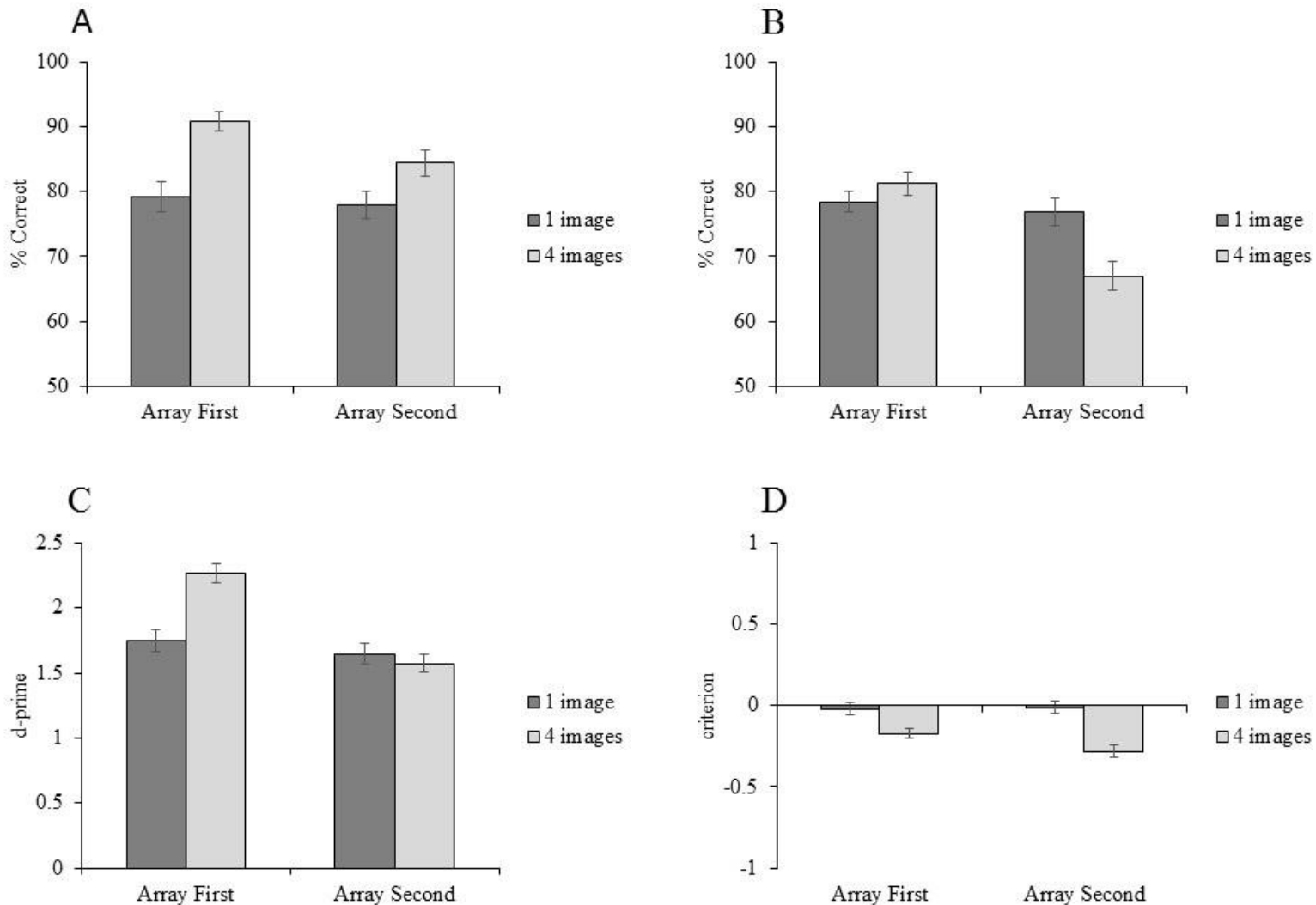
25

26 Participants completed two blocks of trials, 'array first' and 'array second', counterbalanced
27 across participants. Each block contained 40 trials: 10 single image match, 10 single image
28 mismatch, 10 four-image array match, 10 four-image array mismatch. The 'array first' block
29 was identical to the sequential matching procedure described in Experiment 2. The target
30 image/array was presented first for 4 s, followed by a blank screen for 2 s, followed by the
31 comparison (match or foil) image which stayed on screen until the participant responded. In
32 the 'array second' block, the order of the target image/array and the comparison image was
33 swapped so that the comparison image was displayed for 4 s, followed by a blank screen for
34 2 s, followed by the target image/array until response.

1

2 **Results**

3 Figure 5 shows mean performance across conditions in Experiment 3. We analysed match
 4 and mismatch trials separately, using a 2 (array order: array first, array second) x 2 (number
 5 of images: 1, 4) within subjects ANOVA.



7 **Figure 5.** Results of Experiment 3 presenting the array first or second in a sequential
 8 matching task. A) Match trials. B) Mismatch trials. C) d-prime. D) criterion. Error bars show
 9 the within-subjects standard error (Cousineau, 2005).

10

11 The ANOVA for match trials showed a non-significant main effect of array order
 12 $F(1,49) = 3.12, p = .084, \eta_p^2 = .06, BF_{10} = 0.93$, a significant main effect of number of images
 13 $F(1,49) = 30.53, p < .001, \eta_p^2 = .38, BF_{10} = 2.34$ whereby participants were more accurate

1 with four ($M = 87.60\%$) compared to one image ($M = 78.60\%$), and a non-significant
 2 interaction $F(1,49) = 2.23, p = .142, \eta_p^2 = .04, BF_{10} = 0.56$.

3

4 The ANOVA for mismatch trials showed a significant main effect of array order
 5 $F(1,49) = 14.15, p < .001, \eta_p^2 = .22, BF_{10} = 517.71$, a non-significant main effect of number
 6 of images $F(1,49) = 4.00, p = .051, \eta_p^2 = .08, BF_{10} = 0.70$, and a significant interaction
 7 $F(1,49) = 20.95, p < .001, \eta_p^2 = .30, BF_{10} = 57.23$. Simple main effects showed a non-
 8 significant effect of number of images when the array was presented first $F(1,98) = 1.58$,
 9 $p = .212, \eta_p^2 = .02, BF_{10} = 0.58$, but a significant effect of number of images when the array
 10 was presented second $F(1,98) = 19.36, p < .001, \eta_p^2 = .16, BF_{10} = 65.20$ with poorer
 11 performance with a four-image array ($M = 67.00\%$) than a single image ($M = 76.80\%$).
 12 Simple main effects also showed a non-significant effect of array order with one image
 13 $F(1,98) = 0.41, p = .523, \eta_p^2 < .001, BF_{10} = 0.27$, but a significant effect of array order with
 14 four-image arrays $F(1,98) = 31.99, p < .001, \eta_p^2 = .25, BF_{10} = 16,165.33$ with poorer
 15 performance when the array was presented second ($M = 67.00\%$) compared to first
 16 ($M = 81.20\%$).

17

18 Signal detection analysis showed a similar pattern of results. An ANOVA on d' values
 19 showed a significant main effect of array order $F(1,49) = 13.13, p < .001, \eta_p^2 = .21$,
 20 $BF_{10} = 831.073$, a significant main effect of number of images $F(1,49) = 10.28, p = .002$,
 21 $\eta_p^2 = .17, BF_{10} = 1.92$, and a significant interaction $F(1,49) = 14.89, p < .001, \eta_p^2 = .23$,
 22 $BF_{10} = 35.02$. Simple main effects showed an effect of number of images only when the array
 23 was presented first $F(1,98) = 25.14, p < .001, \eta_p^2 = .20, BF_{10} = 1,995.71$ with higher
 24 sensitivity for four images ($M = 2.27$) compared to one ($M = 1.75$). The simple main effect
 25 was non-significant when the array was shown second $F(1,98) = 0.53, p = .468, \eta_p^2 = .01$,
 26 $BF_{10} = 0.26$. Simple main effects also showed an effect of array order with four images in the
 27 array $F(1,98) = 26.82, p < .001, \eta_p^2 = .21, BF_{10} = 94,114.22$, with poorer performance when
 28 the array was presented second ($M = 1.57$) compared to first ($M = 2.27$). The simple main
 29 effect of 'array order' when the array consisted of just one image was non-significant
 30 $F(1,98) = 0.59, p = .444, \eta_p^2 = .01, BF_{10} = 0.28$.

31

32 For criterion values, there was a non-significant main effect of array order $F(1,49) = 1.51$,
 33 $p = .225, \eta_p^2 = .03, BF_{10} = 0.27$ and a significant main effect of number of images
 34 $F(1,49) = 19.43, p < .001, \eta_p^2 = .28, BF_{10} = 6,941.18$ whereby participants were more likely

1 to respond ‘match’ with four-image arrays ($M = -.23$) than one image ($M = -.02$). The
2 interaction was non-significant $F(1,49) = 2.35, p = .132, \eta_p^2 = .05 \text{ BF}_{10} = 0.51$.

3

4 These results demonstrate that match trial performance improved with four images when the
5 array was presented first, without a deficit in mismatch performance. When the array was
6 presented second, although match performance increased, mismatch performance was poorer
7 than with one image. This was confirmed by signal detection analyses which showed no
8 overall benefit in sensitivity when the array was presented second, but a clear benefit when
9 the array came first. This experiment suggests that multiple-image arrays only provide a
10 benefit to performance when the array is shown before the target image, requiring memory,
11 and not when it is displayed after the target image. This suggests that the variability
12 advantage found in Experiment 3 above, and in the face learning literature (e.g. Murphy et
13 al., 2015; Ritchie and Burton, 2017; Robins et al., 2018) is due to the memory component of
14 the task, forcing participants to abstract a representation of the person from the variable
15 images in order to compare a subsequent image to that representation. When the array was
16 presented second in the current experiment, a representation had not been abstracted from
17 variability, but simply relied on a single image, and the variability shown in the array was not
18 helpful. This is similar to the effects reported in Experiments 1 and 2, and elsewhere (Ritchie
19 et al., 2020), whereby an array presented simultaneously with the target image does not result
20 in an overall benefit to performance.

21

22 In order to strengthen our conclusions about the importance of memory for the variability
23 advantage, we sought to replicate our results using a different paradigm and a different set of
24 images.

25

26 **Experiment 4 – applying the array order manipulation to a new task**

27 This experiment further examined the effect of the presentation order of the array and the
28 comparison image. Here we used an adaptation of the sequential matching paradigm used in
29 Dowsett, Sandford, and Burton (2016). This allowed us to investigate whether the variability
30 advantage is still found in a different face matching paradigm which includes a memory
31 component.

32

33 **Method**

34 *Participants*

1 Forty-five participants took part (3 male, mean age: 20 years, range: 18-25 years). All
2 participants were students or other members of the University of Lincoln. All participants
3 gave informed consent, and the study was granted ethical approval by the University of
4 Lincoln School of Psychology Research Ethics Committee.

5

6 *Stimuli and procedure*

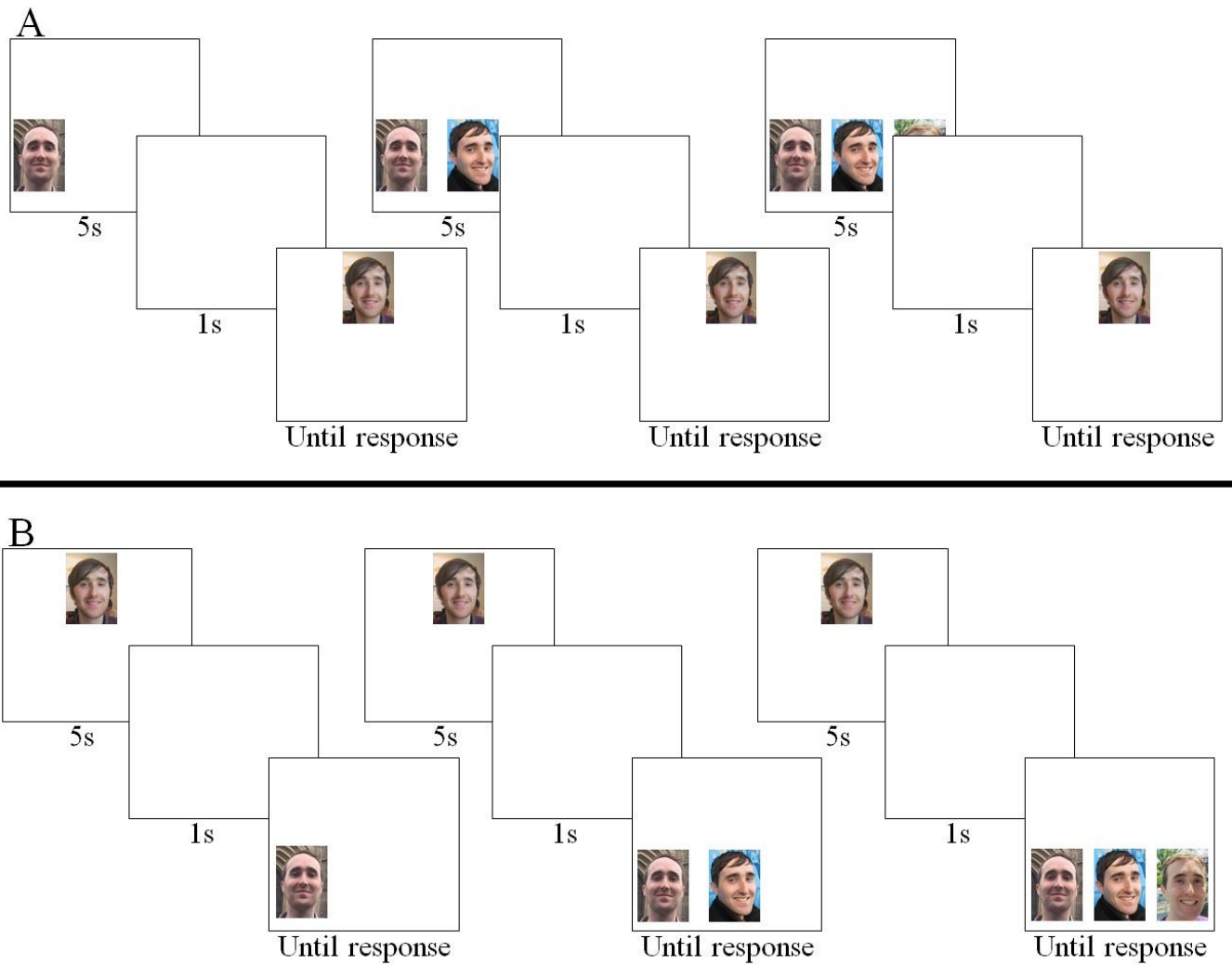
7 The stimuli used in this experiment were a subset of 60 of the 80 identities used in
8 Experiment 3. Participants completed six blocks, where we manipulated the number of
9 images in the array as well as the presentation order of the array and comparison image. The
10 task was designed to be similar to a computerised version of the task used with physical cards
11 by Dowsett, Sandford, and Burton (2016) and is also akin to the paradigm used in a recent
12 study by Sandford and Ritchie (under review).

13

14 Blocks 1-3 showed a sequential matching task with the array appearing *before* the target
15 image, with the size of the array increasing across blocks. Blocks 4-6 showed a sequential
16 matching task with the array appearing *after* the target image, again with the size of the array
17 increasing across blocks. Blocks 1 and 4 showed a one-to-one sequential matching task.
18 Blocks 2 and 5 showed a two-image array paired with a single comparison image, and blocks
19 3 and 6 showed a three-image array paired with a single comparison image.

20

21 In every block, the first image/array was shown for 5s, followed by a blank screen for 1s, and
22 then the target image/array was presented until response. The array images were shown at the
23 bottom of the screen (image 1 on the left, image 2 in the middle, and image 3 on the right),
24 and the target was shown at the top centre (see Figure 6). Each block showed half match and
25 half mismatch trials. Participants completed Blocks 1-3 (array first) then 4-6 (array second),
26 or 4-6 then 1-3 (counterbalanced across participants). Identities were randomly assigned to
27 blocks, with each identity appearing once in Blocks 1-3 and once in Blocks 4-6.

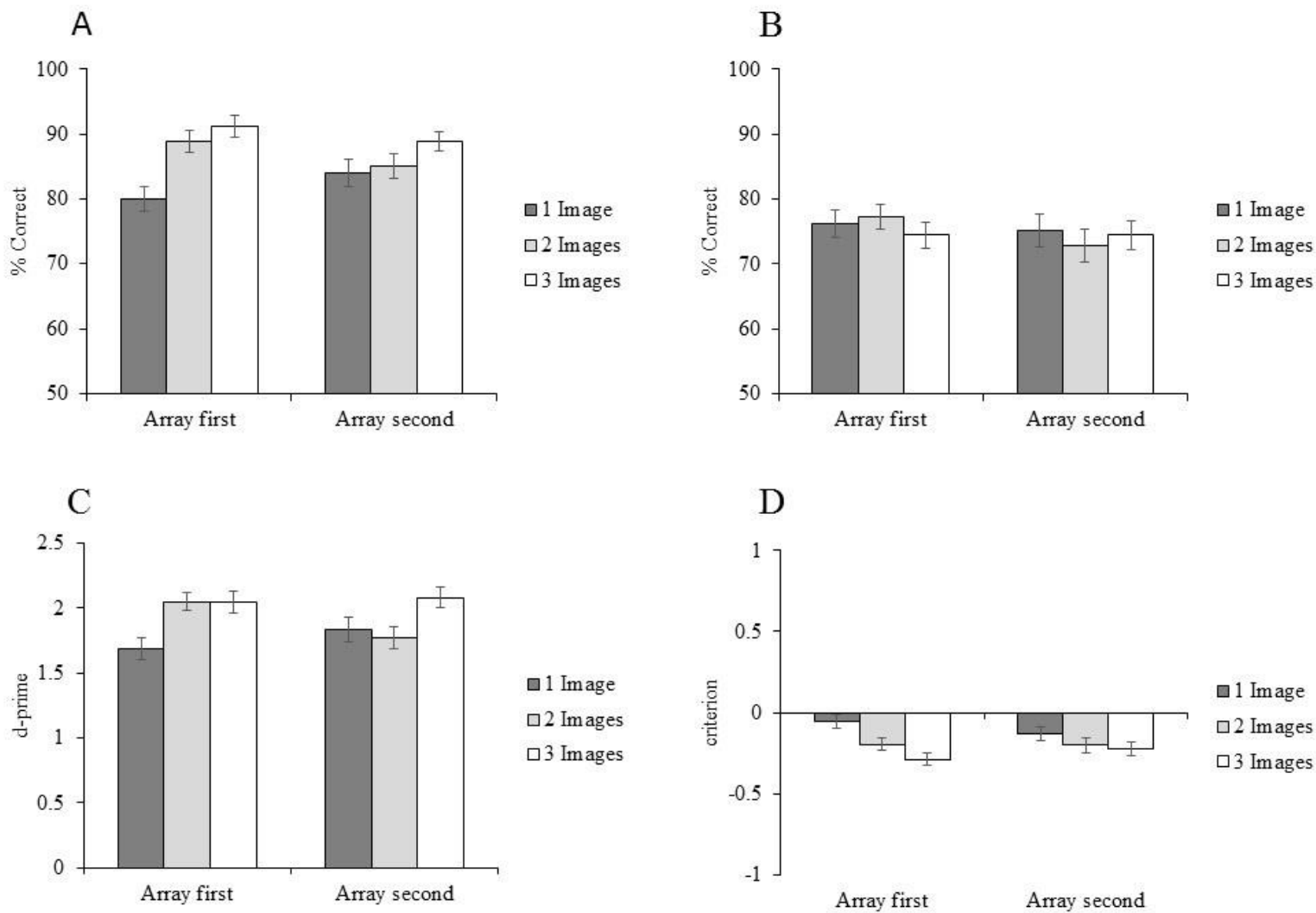


1
2 **Figure 6.** Schematic of the paradigm used in Experiment 4. A) Array first conditions (Blocks
3 1-3). B) Array second conditions (Blocks 4-6). Left) One-to-one match trial. Middle) Two-
4 image array match trial. Right) Three-image array match trial.

5

6 **Results**

7 We analysed match and mismatch trials separately using a 2 (array order: array first, array
8 second) x 3 (number of images: 1, 2, 3) within subjects ANOVA. Figure 7 shows the results
9 of Experiment 4.



2 **Figure 7.** Results of Experiment 4 presenting the increasing numbers of array images first or
 3 second in a sequential matching task. A) Match trials. B) Mismatch trials. C) d-prime. D)
 4 criterion. Error bars show the within-subjects standard error (Cousineau, 2005).

5

6 The ANOVA for match trials showed a non-significant main effect of array order
 7 $F(1,44) = 0.28, p = .599, \eta_p^2 = .01, BF_{10} = 0.15$, a significant main effect of number of images
 8 $F(2,88) = 15.35, p < .001, \eta_p^2 = .26, BF_{10} = 5,510.41$, and a significant interaction
 9 $F(2,88) = 3.12, p < .01, \eta_p^2 = .07, BF_{10} = 1.52$. Simple main effects showed a non-significant
 10 effect of number of images when the array was presented second $F(1,176) = 2.71, p = .102,$
 11 $\eta_p^2 = .03, BF_{10} = 0.68$, but a significant effect of number of images when the array was
 12 presented first $F(1,176) = 14.29, p < .001, \eta_p^2 = .14, BF_{10} = 22,546.40$. Tukey HSD tests
 13 showed a significant improvement in accuracy when the array presented first contained two

1 images ($M = 88.89\%$) or three images ($M = 91.11\%$) compared to a single image
 2 ($M = 80.00\%$), both $ps < .05$. No other comparisons were significant ($ps > .05$).

3

4 The ANOVA for mismatch trials showed a non-significant main effect of array order
 5 $F(1,44) = 0.84$, $p = .364$, $\eta_p^2 = .02$, $BF_{10} = 0.24$, a non-significant main effect of number of
 6 images $F(2,88) = 0.17$, $p = .844$, $\eta_p^2 < .001$, $BF_{10} = 0.05$, and a non-significant interaction
 7 $F(2,88) = 0.74$, $p = .480$, $\eta_p^2 = .02$, $BF_{10} = 0.12$.

8

9 An ANOVA on d' values showed a non-significant main effect of array order $F(1,44) = 0.16$,
 10 $p = .691$, $\eta_p^2 < .001$, $BF_{10} = 0.14$, and a significant main effect of number of images
 11 $F(2,88) = 5.55$, $p = .005$, $\eta_p^2 = .11$, $BF_{10} = 5.81$, with Tukey's HSD tests showing higher
 12 sensitivity with three images ($M = 2.06$) compared to one ($M = 1.76$) or two images
 13 ($M = 1.91$), $ps < .05$. There was a non-significant interaction $F(2,88) = 3.07$, $p = .051$,
 14 $\eta_p^2 = .07$, $BF_{10} = 0.72$.

15

16 For criterion values, there was a non-significant main effect of array order $F(1,44) = 0.02$,
 17 $p = .888$, $\eta_p^2 < .001$, $BF_{10} = 0.13$, a significant main effect of number of images
 18 $F(2,88) = 7.51$, $p < .001$, $\eta_p^2 = .15$, $BF_{10} = 15.12$, and a non-significant interaction
 19 $F(2,88) = 1.20$, $p = .306$, $\eta_p^2 = .03$, $BF_{10} = 0.19$. Tukey's HSD tests showed significantly
 20 more bias for arrays containing two ($M = -.20$) and three images ($M = -.26$) than single
 21 images ($M = -.09$), $ps < .05$. This shows a tendency to respond "match" more for multiple-
 22 image arrays. No other comparisons were significant, $ps > .05$.

23

24 In this final experiment, using a different paradigm, we find that presenting a multiple-image
 25 array helps on match trials, without harming performance on mismatch trials, only when that
 26 array is displayed before and not after the target image. Other than the lack of an overall
 27 benefit in terms of sensitivity here compared to Experiment 3, these results show the same
 28 pattern as observed in our previous experiments.

29

30 **General Discussion**

31 Across the four experiments presented here, we see a clear pattern of results whereby
 32 multiple-image arrays lead to improved face matching performance in sequential matching
 33 tasks. This effect is only present when the array is presented before and not after the target
 34 image. We do not find the multiple image advantage for simultaneous face matching tasks.

1 These results reconcile the differences between the face learning literature which shows that
2 exposure to within-person variability and multiple images help with face learning (e.g.
3 Dowsett, Sandford, & Burton, 2016; Longmore et al., 2017; Matthews et al., 2018; Murphy et
4 al., 2015; Ritchie & Burton, 2017) but do not necessarily help face matching (Kramer &
5 Reynolds, 2018; Ritchie et al., 2020; Sandford & Ritchie, under review). Learning paradigms
6 require a representation of the identity to be abstracted from multiple images and held in
7 memory until the time of testing. Likewise, our sequential matching paradigm (Experiments
8 2-4) requires a representation of the identity to be abstracted from the array, held in memory,
9 and compared to the target image. When we present the array and the target image
10 simultaneously (Experiments 1 and 2), or the target image before the array (Experiments 3
11 and 4), we do not see a multiple-image benefit, as there is either no memory component or
12 only a single image to be held in memory.

13

14 There are two different mechanisms that could potentially account for the improvement in
15 face matching with access to multiple naturally varying images of the same person. The first
16 follows from the Bruce and Young model (1986) and the concept of FRUs. Here, information
17 from the multiple-image array is aggregated together in a way that preserves what is
18 diagnostic of the identity, while ignoring superficial image differences, to form a stable
19 mental representation that can easily support recognition. An alternative explanation,
20 however, is that by increasing the number of images available to participants, we are also
21 increasing the chance of finding an image that is particularly similar to the comparison image
22 (i.e., a closest match). This will also lead to an improvement in accuracy but superficial
23 image characteristics might still be attended to and taken into consideration. The results from
24 our simultaneous and sequential matching tasks might help us differentiate between these two
25 strategies. While both mechanisms could be used in a sequential task, there is no need to
26 create a mental representation of the identity in a simultaneous task. We can see all images at
27 the same time, therefore the set up of a simultaneous task might instead encourage
28 participants to adopt a closest match strategy. Since our results show a multiple image
29 advantage in sequential tasks only, they provide support for the concept of stability from
30 variation where different images of the same person are integrated into a single identity
31 representation. This is consistent with previous work by Menon et al. (2015b) who presented
32 participants with two images and either instructed them that they belonged to the same person
33 (to encourage integration) or that they belonged to two different people (to stop integration).
34 Differences between these two conditions were only found using a sequential (but not a

1 simultaneous) task, again, suggesting that a closest match strategy is more likely to be used in
2 a simultaneous matching task.

3

4 It is important to consider what form a ‘stable mental representation’ of an identity might
5 take. When a set of similar items are presented, it has been shown that viewers extract
6 summary information, a process referred to as ‘ensemble coding’. Viewers incorrectly report
7 having seen an image which represents the mean of the set (also referred to as the average, or
8 prototype) when that image was in fact never displayed. This has been shown for circles
9 (Ariely, 2001) as well as faces (e.g. de Fockert & Wolfenstein, 2009; Neumann,
10 Schweinberger, & Burton, 2013). We have previously shown that viewers extract the mean
11 from images of faces, whether presented simultaneously or sequentially (Kramer, Ritchie, &
12 Burton, 2015). We have also argued, however, that face averages do not consistently improve
13 face matching accuracy (Ritchie et al, 2020; Ritchie, White et al., 2018) and do not give rise
14 to higher likeness ratings than specific exemplars (Ritchie, Kramer & Burton, 2018).

15 Therefore, we do not suggest here that a stable mental representation of an identity must
16 necessarily constitute a simple ‘average’ or prototype. Instead, it seems likely that robust
17 representations of a familiar faces incorporate both abstractive and instance-specific
18 information.

19

20 In addition to this finding, Experiment 1 manipulated the amount of variability in the arrays.
21 It is possible that previous experiments that found a multiple-image advantage in
22 simultaneous face matching (White at al., 2014) simply presented more variability in their
23 arrays than the experiments that did not (Ritchie et al., 2020; Sandford & Ritchie, under
24 review). Experiment 1, however, showed no overall benefit of either low or high variability
25 arrays on face matching performance. Therefore, the amount of variability likely does not
26 explain the differences between previous results. Nevertheless, assuming that a closest match
27 strategy is more likely to be used in such a situation, then the conflicting results from these
28 studies could simply be due to subtle differences in the image sets used.

29

30 Three recent studies have looked at the utility of providing multiple images when searching
31 for a face in an array or a crowd. Dunn, Kemp, and White (2018) showed participants one or
32 four images of a target identity for 3 s, and then had participants search for a new image of
33 the person in an array of faces. Searching for unfamiliar people was improved, both in terms
34 of higher accuracy and faster reaction times, when participants had seen four compared to

1 only one image of the target. Two subsequent studies had participants search for unfamiliar
2 people in videos of crowds, and presented the image(s) of the target identity simultaneously
3 with the crowd video (Kramer, Hardy, & Ritchie, 2020; Mileva & Burton, 2019). Mileva and
4 Burton (2019) found that providing participants with three ID-document images of the target
5 improved search performance over one image, with no further increase when 16 images were
6 provided. Kramer et al. (2020) also found an increase in performance with three recent
7 images of the target compared to one. Here, we find that multiple images improve face
8 matching performance in a sequential task (akin to Dunn et al., 2018). We do not find here
9 that arrays improve face matching when the array and the target are presented
10 simultaneously, but both Mileva and Burton (2019) and Kramer et al. (2020) do find that
11 arrays help with searching for faces in crowds when the array and the crowd video are
12 presented simultaneously. This difference in results may be due to the different nature of the
13 tasks, with searching being a much more difficult and complex task, perhaps inherently
14 involving an aspect of memory, where participants may try to memorise the target images
15 then view the video. In fact, Mileva and Burton (2019) lend some support for this idea in an
16 experiment which gave participants as a reference a video of the target rotating their head.
17 The authors report “Informally, we observed that searchers typically froze the target video
18 while searching the CCTV clip, suggesting that two simultaneous moving displays impose
19 too high a load to be useful” (Mileva & Burton, 2019, p. 11). Neither in the search studies nor
20 the studies presented in this paper can we rule out that participants also found high variability
21 arrays to be too high a load to be useful. In fact, our observed change in bias in simultaneous
22 matching tasks for high variability images (Experiments 1 and 2) may speak to this in that
23 participants may have been overwhelmed by the variability in the array and so simply
24 responded ‘match’ more often than ‘mismatch’.

25

26 Our results should be viewed within the context of the wider literature on face learning, face
27 matching, and representations of familiar and newly learned identities. It is evident from the
28 face learning literature that exposure to variability does give rise to a representation that is
29 stable enough to support recognition of new images of the newly learned people. However,
30 we argue that exposure to variability is not sufficient to produce fast “learning” in order to
31 help in a simultaneous matching task. Our results suggest that exposure to variability is only
32 helpful for face processing tasks which require an element of memory, where the learning
33 and test stimuli are presented sequentially. Future research should establish the limits of this
34 variability advantage using different short-term and long-term memory tasks.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

Supplementary Material

The data for all four experiments is available at [Cognition to add URL].

References

- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162. DOI: 10.1111/1467-9280.00327
- Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19-30. DOI: 10.1016/j.cognition.2016.12.0120010-0277
- Bindemann, M., & Sandford, A. (2011). Me, myself and I: Different recognition rates for three photo-IDs of the same person. *Perception*, *40*, 625–627. DOI: 10.1068/p700
- Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*, 105–116. DOI: 10.1111/j.2044-8295.1982.tb01795.x
- Bruce, V. (1994). Stability from variation: The case of face recognition. *Quarterly Journal of Experimental Psychology*, *47*, 5–28. DOI: 10.1080/14640749408401141
- Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception*, *15*(4), 387-397. DOI: 10.1068/p150387
- 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. DOI: 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. DOI: 10.1037//1076-898X.7.3.207

- 1 Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of*
2 *Psychology*, 77, 305–327. DOI: 10.1111/j.2044-8295.1986.tb02199.x
3
- 4 Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The
5 importance of variability. *Quarterly Journal of Experimental Psychology*, 66(8), 1467–
6 1485. DOI: 10.1080/17470218.2013.800125
7
- 8 Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar
9 faces. *British Journal of Psychology*, 102(4), 943–958. DOI: 10.1111/j.2044-
10 8295.2011.02039.x
11
- 12 Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation:
13 Representations of faces derived from multiple instances. *Cognitive Science*, 40, 202-
14 223. DOI: 10.1111/cogs.12231
15
- 16 Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior*
17 *Research Methods*, 42(1), 286-291. DOI: 10.3758/BRM.42.1.286
18
- 19 Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality
20 video: Evidence from security surveillance. *Psychological Science*, 10(3), 243-248. DOI:
21 10.1111/1467-9280.00144
22
- 23 Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an
24 indirect face-matching measure. *Perception*, 31, 985–994. DOI: 10.1068/p3335
25
- 26 Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual*
27 *Cognition*, 11(7), 857–869. DOI: 10.1080/13506280444000021
28
- 29 Cousineau, D. (2005). Confidence intervals in within-subjects designs: A simpler solution to
30 Loftus and Masson’s method. *Tutorials in Quantitative Methods for Psychology*, 1(1),
31 42-45.
32

- 1 Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the
2 defendant in the dock. *Applied Cognitive Psychology*, *23*(4), 482–505. DOI:
3 10.1002/acp.1490
4
- 5 de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of
6 faces. *Quarterly Journal of Experimental Psychology*, *62*, 1716–1722. DOI:
7 10.1080/17470210902811249
8
- 9 Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images
10 leads to fast acquisition of familiarity for specific individuals. *Quarterly Journal of*
11 *Experimental Psychology*, *69*(1), 1–10. DOI: 10.1080/17470218.2015.1017513
12
- 13 Dunn, J. D., Kemp, R. I., & White, D. (2018). Search templates that incorporate within-face
14 variation improve visual search for faces. *Cognitive Research: Principles and*
15 *Implications*, *3*, 37. DOI: 10.1186/s41235-018-0128-1
16
- 17 Graham, D. L., & Ritchie, K. L. (2019). Making a spectacle of yourself: The effect of glasses
18 and sunglasses on face perception. *Perception*. Advance online publication. DOI:
19 10.1177/0301006619844680
20
- 21 Henderson, J. M., Pollatsek, A., & Rayener, K. (1987). Effects of foveal priming and
22 extrafoveal preview on object identification. *Journal of Experimental Psychology:*
23 *Human Perception and Performance*, *13*, 449–463. DOI: 10.1037/0096-1523.13.3.449
24
- 25 JASP Team (2020). JASP (Version 0.14) [Computer software].
26
- 27 Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of
28 the same face. *Cognition*, *121*(3), 313–323. DOI: 10.1016/j.cognition.2011.08.001
29
- 30 Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A
31 review. *Memory*, *17*(5), 577–596. DOI: 10.1080/09658210902976969
32
- 33 Jones, S. P., Dwyer, D. M., & Lewis, M. B. (2017). The utility of multiple synthesized views
34 in the recognition of unfamiliar faces. *Quarterly Journal of Experimental Psychology*,

- 1 70(5), 906-918. DOI: 10.1080/17470218.2016.1158302
- 2
- 3 Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs,
4 credit cards and fraud. *Applied Cognitive Psychology, 11*(3), 211–222. DOI:
5 10.1002/(SICI)1099-0720(199706)11:3<211::AID-ACP430>3.0.CO;2-O
- 6
- 7 Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory*
8 *and Cognition, 12*(1), 60–70. DOI: 10.3758/BF03196998
- 9
- 10 Kramer, R. S. S., Hardy, S. C., & Ritchie, K. L. (2020). Searching for faces in crowd
11 chokepoint videos. *Applied Cognitive Psychology, 34*(2), 343-356. DOI:
12 10.1002/acp.3620
- 13
- 14 Kramer, R. S. S., & Reynolds, M. G. (2018). Unfamiliar face matching with frontal and
15 profile views. *Perception, 47*(4), 414–431. DOI: 10.1177/0301006618756809
- 16
- 17 Kramer, R. S. S., & Ritchie, K. L. (2016). Disguising superman: How glasses affect
18 unfamiliar face matching. *Applied Cognitive Psychology, 30*, 841–845. DOI:
19 10.1002/acp.3261
- 20
- 21 Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from
22 images of the same person: A route to face learning. *Journal of Vision, 15*(4):1, 1-9.
23 DOI: 10.1167/15.4.1
- 24
- 25 Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity.
26 *Cognition, 172*, 46–58. DOI: 10.1016/j.cognition.2017.12.005
- 27
- 28 Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs.
29 *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 77–
30 100. DOI: 10.1037/0096-1523.34.1.77
- 31
- 32 Longmore, C. A., Santos, I. M., Silva, C. F., Hall, A., Faloyin, D., & Little, E. (2017). Image
33 dependency in the recognition of newly learnt faces. *Quarterly Journal of Experimental*
34 *Psychology, 70*, 863–873. DOI: 10.1080/17470218.2016.1236825

- 1
2 Matthews, C. M., Davis, E. E., & Mondloch, C. J. (2018). Getting to know you: The
3 development of mechanisms underlying face learning. *Journal of Experimental Child*
4 *Psychology, 167*, 295-313. DOI: 10.1016/j.jecp.2017.10.012
5
- 6 Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a
7 matching task. *Memory & Cognition, 34(4)*, 865–876. DOI: 10.3758/BF03193433
8
- 9 Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A
10 familiarity-based dissociation. *Perception & Psychophysics, 69(7)*, 1175-1184. DOI:
11 10.3758/BF03193954
12
- 13 Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance
14 in eyewitness memory (without the memory). *Journal of Experimental Psychology:*
15 *Applied, 14(4)*, 364–372. DOI: 10.1037/a0013464
16
- 17 Menon, N., White, D., & Kemp, R. I. (2015a). Variation in photos of the same face drives
18 improvements in identity verification. *Perception, 44(11)*, 1332-1341. DOI:
19 10.1177/0301006615599902
20
- 21 Menon, N., White, D., & Kemp, R. I. (2015b). Identity-level representations affect unfamiliar
22 face matching performance in sequential but not simultaneous tasks. *The Quarterly*
23 *Journal of Experimental Psychology, 68(9)*, 1777-1793. DOI:
24 10.1080/17470218.2014.990468
25
- 26 Mileva, M., & Burton, A. M. (2019). Face search in CCTV surveillance. *Cognitive Research:*
27 *Principles and Implications, 4*, 37. DOI: 10.1186/s41235-019-0193-0
28
- 29 Murphy, J., Ipser, A., Gaigg, S., & Cook, R. (2015). Exemplar variance supports robust
30 learning of facial identity. *Journal of Experimental Psychology: Human Perception and*
31 *Performance, 41*, 577–581. DOI: 10.1037/xhp0000049
32

- 1 Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and
2 individual identity from sets of famous faces. *Cognition*, *128*, 56–63. DOI:
3 10.1016/j.cognition.2013.03.006
4
- 5 Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of*
6 *Experimental Psychology*, *70*(5), 897-905. DOI: 10.1080/17470218.2015.1136656
7
- 8 Ritchie, K. L., Kramer, R. S. S., & Burton, A. M. (2018). What makes a face photo a ‘good
9 likeness’? *Cognition*, *170*, 1-8. DOI: 10.1016/j.cognition.2017.09.001
10
- 11 Ritchie, K. L., Mireku, M. O., & Kramer, R. S. S. (2020). Face averages and multiple images
12 in a live matching task. *British Journal of Psychology*, *111*(1), 92-102. DOI:
13 10.1111/bjop.12388
14
- 15 Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015).
16 Viewers base estimates of face matching accuracy on their own familiarity: Explaining
17 the photo-ID paradox. *Cognition*, *141*, 161-169. DOI: 10.1016/j.cognition.2015.05.002
18
- 19 Ritchie, K. L., White, D., Kramer, R. S. S., Noyes, E. Jenkins, R., & Burton, A. M. (2018).
20 Enhancing CCTV: Averages improve face identification from poor-quality images.
21 *Applied Cognitive Psychology*, *32*, 671-680. DOI: 10.1002/acp.3449
22
- 23 Robins, E., Susilo, T., Ritchie, K. L., & Devue, C. (2018, July 23). Within-person variability
24 promotes learning of internal facial features and facilitates perceptual discrimination and
25 memory. DOI: 10.31219/osf.io/5scnm
26
- 27 Sandford, A. L. R., & Ritchie, K L. (under review). Unfamiliar face matching, within-person
28 variability, and multiple-image arrays.
29
- 30 White, D., Burton, A. M., Jenkins, R., & Kemp, R. (2014). Redesigning photo-ID to improve
31 unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*,
32 *20*(2), 166-173. DOI: 10.1037/xap0000009
33

- 1 White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport
- 2 officers' errors in face matching. *PLoS ONE*, 9, e103510. DOI:
- 3 10.1371/journal.pone.0103510