Viewers extract the mean from images of the same person: A route to face learning

Robin S. S. Kramer School of Psychology, University of Aberdeen, Aberdeen, UK; Department of Psychology, University of York, York, UK

Kay L. Ritchie School of Psychology, University of Aberdeen, Aberdeen, UK; Department of Psychology, University of York, York, UK

A. Mike Burton School of Psychology, University of Aberdeen, Aberdeen, UK; Department of Psychology, University of York, York, UK

Research on ensemble encoding has found that viewers extract summary information from sets of similar items. When shown a set of four faces of different people, viewers merge identity information from the exemplars into a representation of the set average. Here, we presented sets containing unconstrained images of the same identity. In response to a subsequent probe, viewers recognized the exemplars accurately. However, they also reported having seen a merged average of these images. Importantly, viewers reported seeing the matching average of the set (the average of the four presented images) more often than a nonmatching average (an average of four other images of the same identity). These results were consistent for both simultaneous and sequential presentation of the sets. Our findings support previous research suggesting that viewers form representations of both the exemplars and the set average. Given the unconstrained nature of the photographs, we also provide further evidence that the average representation is invariant to several high-level characteristics.

Introduction

When viewers are shown sets of perceptually similar items, there is growing evidence to suggest that summary statistics, such as the mean, may be represented. This "ensemble encoding" is thought to provide an efficient way of summarizing both low-level and more complex scene information. For example, when participants were shown sets containing circles of different sizes, they tended incorrectly to identify a test circle as having been present when it had a similar size to the mean of the set (Ariely, 2001). In addition, participants were near chance when asked to identify which circles had actually been present. This now common pattern of findings is interpreted as viewers forming an accurate representation of the average of a set while retaining less (if any) information regarding individual exemplars. As well as basic size averaging, similar results have been found, for example, with judgments of orientation (Robitaille & Harris, 2011), speed (Atchley & Andersen, 1995), and dynamic displays (Albrecht & Scholl, 2010).

More recently, researchers have begun to consider whether viewers also encode the average for a more complex set of stimuli: human faces. Evidence suggests that participants form an accurate representation of the average emotional expression (Haberman, Harp, & Whitney, 2009; Haberman & Whitney, 2007, 2009), gender (Haberman & Whitney, 2007), and gaze direction (Sweeny & Whitney, 2014) from a set of faces. Further, this process is not mediated by low-level features, luminance cues, or other nonconfigural cues (Haberman & Whitney, 2009). Of note, given the rapid extraction of summary information (e.g., the average expression from 16 faces presented for 500 ms or less), this ensemble encoding is likely distinct from the prototype effect (building an abstract prototypical representation based on repeated occurrences), which typically operates over the order of minutes (e.g., Fiser & Aslin, 2001).

Citation: Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person: A route to face learning. *Journal of Vision*, *15*(4):1, 1–9, http://www.journalofvision.org/content/15/4/1, doi:10.1167/15.4.1.

doi: 10.1167/15.4.1 Received October 29, 2014; published April 17, 2015 ISSN 1534-7362

Several studies have now demonstrated that identity information is also represented by summary statistics. When shown four images of different people, participants extract the mean identity, no matter whether these faces are unfamiliar (de Fockert & Wolfenstein, 2009) or familiar (Neumann, Schweinberger, & Burton, 2013). Importantly, subsequent studies have ruled out the possibility that viewers are simply extracting the mean retinal image by presenting sets of faces that incorporated different viewpoints (Leib et al., 2014). As such, the evidence suggests that ensemble encoding can operate on viewpoint-invariant representations, which broadens its applicability and usefulness for real-world scenes.

To date, research investigating the ensemble encoding of identity has only considered multiple identities. By averaging across people, the suggestion is that viewers form the gist of a crowd, for example. Although it may be useful to encode the average expression ("this crowd looks angry"), it is less clear why a representation of the average identity may be beneficial ("the average of all these people's faces would look like this"). In contrast, if we are exposed to multiple instances of a single person, perhaps over several encounters or movies, then encoding the average of those instances has clear advantages. The average of a set of instances can provide a stable representation of an individual by washing away aspects of the set that change from one photo to the next while preserving aspects that are consistent across the set (Burton, Jenkins, Hancock, & White, 2005; Jenkins & Burton, 2008). These representations are also robust to errors in that incorporating a few photographs of the wrong person makes little difference to the average (Jenkins, Burton, & White, 2006; see also Haberman & Whitney, 2010, for evidence that the visual system discounts outliers when encoding the average emotional expression). As such, encoding an average for a within-person set of images may underpin the process of familiarity through the buildup of exposure to different instances. In the current studies, we focus on this within-person encoding by only including images of a single identity in each trial.

Although some experiments involving faces have utilized simultaneous presentation (e.g., de Fockert & Wolfenstein, 2009; Neumann et al., 2013), others have implemented a sequential design (e.g., Haberman et al., 2009; Leib et al., 2014). Even when viewers were presented with face images one after the other, results have demonstrated that the average of these images was encoded. In the following two studies, we investigate both methods of presentation.

Previous experiments have, for the most part, used relatively homogeneous, gray scale stimuli (e.g., with little variability in pose). By specifically varying pose, Leib and colleagues (2014) demonstrated how encoded representations are viewpoint-invariant. Here, we use "ambient" color images (Jenkins, White, Van Montfort, & Burton, 2011). These are photographs sampled from the real world, and they incorporate a great deal of variability in pose but also in lighting, expression, focal length, etc. Therefore, encoding of the average of these images would need to take place at a sufficiently high level to deal with these differences.

Finally, given evidence of identity averaging for both unfamiliar and familiar faces (de Fockert & Wolfenstein, 2009; Neumann et al., 2013), we included consideration of familiarity in the current research in order to allow for a direct comparison of these two categories of faces. For within-person set averaging, it may be that encoding the average of a familiar person is in some way disrupted by our previous experiences with that identity, including exposure to a potentially large number of prior exemplars. Equally, recognition of that identity may hinder (or help) the encoding process.

Experiment 1: Simultaneous presentation

In this experiment, we presented participants with four images of the same person simultaneously. As such, we investigated whether participants represented simultaneously presented images of the same person as an average.

Methods

Participants

Twenty undergraduate students from the University of Aberdeen (12 women; age $M = 24.1$ years, $SD = 9.3$ years) volunteered to take part in the study and received money for their participation. All provided informed consent prior to participation (in accordance with the ethical standards stated in the 1964 Declaration of Helsinki).

Stimuli

Thirty images were downloaded from the Internet for each of 20 celebrities. We used celebrity photos in order to ensure that many images of each person were available. Ten of these celebrities (five women) were Hollywood actors and so were chosen to be familiar to participants. The other 10 (five women) were Australian celebrities, who were selected in order that they would be unfamiliar to UK participants.

For each identity, we entered the name into Google Images as a search term along with criteria specifying full-color, large, face images only. We then chose the

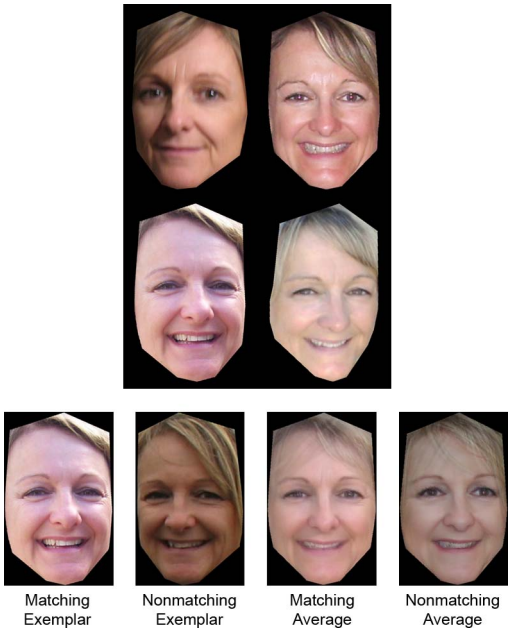

Figure 1. An example display set, followed by the four possible test faces. From left to right, a member of the presented set, an exemplar that did not appear in any display set, the morphed average of the presented set, and the morphed average of a nondisplay set. Note the variability in the display set images. (Copyright restrictions prevent publication of the original images used in these experiments. Images shown here, also used in Figure 3, feature an identity who did not appear in the experiments. She has given permission for her images to be reproduced here.)

first 30 images delivered that met the following criteria: (a) no part of the face should be obscured (for example by clothing, glasses, or a hand); (b) pose should be very broadly full-face in order to allow the placement of landmarks; and (c) pose should be standing or sitting, but not lying down, in order to limit the angle of the head to relatively upright. Note that as a result of obtaining images from the Internet, image variation (lighting, pose, expression, age, etc.) for each identity was large (for an example, see Figure 1). All images were cropped and rotated so that both pupils were aligned to the same transverse plane. Images were also resized so that they appeared as 11.8 cm high with a varying width of approximately 7.5 cm on-screen.

The first 28 images of each identity were divided into seven sets of four images, arbitrarily based on the order in which they were downloaded. For each set, the average was created by morphing across the four images using custom MATLAB software. The first four sets were chosen as the display sets, i.e., those that participants would view during the experiment. The other three formed nondisplay sets and provided additional averages for use as test faces (see below).

Procedure

The procedure closely followed that of Neumann and colleagues (2013). Participants were shown four trials for each identity (80 trials in total). In each trial, a central fixation cross appeared for 1 s. This was followed by four images presented simultaneously (the display set) for 1500 ms with each image randomly assigned to one of four specified positions on-screen. Immediately following the display set (interstimulus interval [ISI] = 0), a test face was presented for 500 ms, smaller in size than the display set images (7.9 cm × approximately 5 cm). Participants used both index fingers to indicate via button press whether the test face had or had not been present in the previous display set. Test faces were (a) a matching exemplar (a randomly selected image from the preceding display set), (b) a nonmatching exemplar (a randomly selected image that was not seen individually or within an average for that identity in other trials), (c) a matching average (the average of the four display set images), or (d) a nonmatching average (the average of four different images, randomly selected from the nondisplay sets). Figure 1 provides an example display set and possible test faces. A blank screen lasting 2200 ms followed the test face, allowing for a total response window of 2700 ms.

For each of these four conditions, 20 trials were presented—one for each identity. These 80 trials were presented in a random order for each participant.

It is important to note that for each of the four trials for a given identity, all images in the display sets and all test face exemplars and averages contained only images of that identity. As such, our focus was solely on within-person representations.

Prior to the experiment proper, participants were given 16 practice trials and provided with trial-by-trial on-screen feedback on their accuracy. (No feedback was given during the actual experiment.) Note that the correct answer to average test faces is always "absent." In order to prevent participants from learning this association, averages were not presented in the practice block. None of the four practice identities appeared in the experimental block.

After completing the practice and experimental blocks, the familiarity of the 20 experiment identities was checked by giving participants a new image (which was not part of the stimulus set) of each celebrity on a printed sheet and asking if they were familiar with that person. It was made clear that familiarity referred to prior experience rather than what participants had seen during the experiment. As expected, familiarity with Hollywood celebrities was high (number of identities recognized $M = 8.5$, $SD = 1.7$), and familiarity with Australian celebrities was low ($M = 0.7$, $SD = 0.8$).

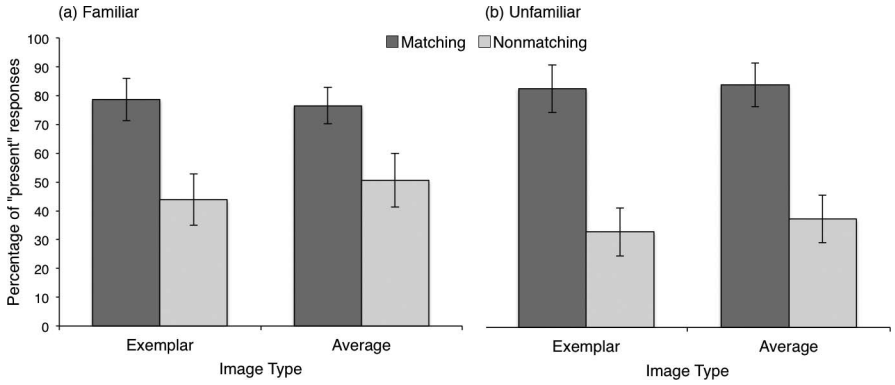

Figure 2. Mean percentage of "present" responses for (a) familiar and (b) unfamiliar test faces. Error bars represent 95% confidence intervals.

Results and discussion

Response data for Experiment 1 are shown in Figure 2. Data were entered into a 2 (Familiarity: Familiar, Unfamiliar) × 2 (Image Type: Exemplar, Average) × 2 (Test Face: Matching, Nonmatching) ANOVA. All factors were within-subjects. We found a significant main effect of Test Face, $F(1, 19) = 213.87$, $p < 0.001$, $\eta^2_p = 0.92$, with participants responding "present" more often for test faces that matched the preceding set ($M = 80.4\%$) than for those that did not ($M = 41.3\%$). We also found a significant Familiarity × Test Face interaction, $F(1, 19) = 24.41$, $p < 0.001$, $\eta^2_p = 0.56$. Simple main effects showed higher "present" responses for matching test faces in both the Familiar condition, $F(1, 19) = 131.93$, $p < 0.001$, $\eta^2_p = 0.87$, and the Unfamiliar condition, $F(1, 19) = 168.16$, $p < 0.001$, $\eta^2_p = 0.90$, with the interaction being driven by a larger effect for unfamiliar faces. No other effects or interactions were significant.

This is an interesting result. It is clear from the "exemplar" conditions that participants were sensitive to the faces they had actually seen, giving significantly more "present" responses to images they had seen over those they had not. However, what is particularly interesting is that this effect was exactly replicated for the "average face" test stimuli. So participants were just as likely to claim that they had seen the average of the display set. This effect is not a simple preference for averages: Participants claimed to have seen the average of the particular photos from the display set, making fewer "present" responses to another average of this person derived from different photos. Given the fact that averages of a particular face will eventually converge, given larger sample sizes, this is rather a striking result. It suggests that the average formation is quite tightly image-bound while, at the same time, not being due to low-level perceptual averaging (Leib et al., 2014).

We propose that this averaging process may underlie face learning. If people are able to extract an average from different photos of the same person, this is a fast route to forming a robust representation that can be used for subsequent recognition of that person. However, in natural settings, we never see the same face represented in different ways simultaneously. In order for this mechanism to be a plausible one for the process of face learning, it should also be evident following sequential presentation of images. This is explored in Experiment 2.

Experiment 2: Sequential presentation

In this experiment, we presented participants with four images of the same person sequentially. In other respects, the design was the same as in Experiment 1. As above, our aim is to establish whether participants acquire an average representation of multiple images of the same person. In particular, do they believe they have seen an average of the set when, in fact, they have not?

Methods

Participants

A further 20 undergraduate students from the University of Aberdeen (16 women; age $M = 22.1$ years, $SD = 4.7$ years) volunteered to take part in the study and received money for their participation. None had taken part in the previous experiment. All provided informed consent prior to participation (in accordance with the ethical standards stated in the 1964 Declaration of Helsinki).

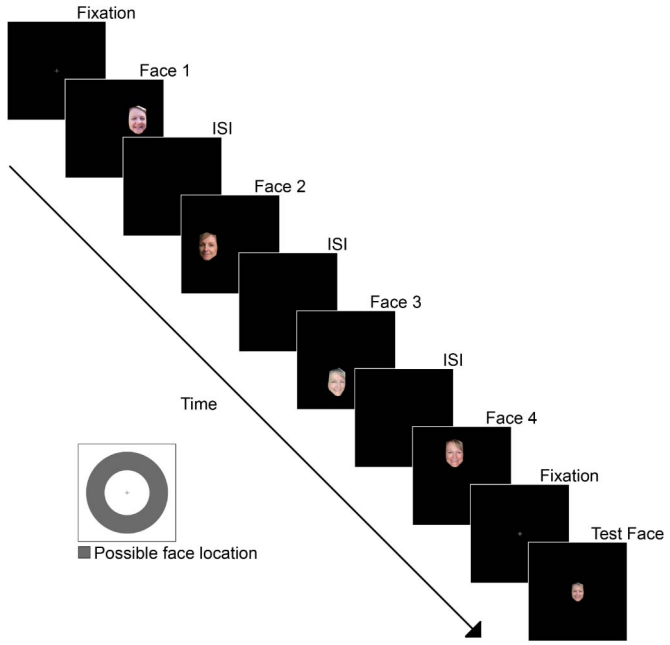

Figure 3. Example of sequential presentation. Each face in the sequence appeared onscreen for 375 ms, and the ISI was also 375 ms.

Procedure

The stimuli were those used in Experiment 1. The procedure was also identical to the first experiment with one important difference: The four images for each display set were presented sequentially. In each trial, a central fixation cross appeared for 1 s. This was followed by the four images, presented one at a time in a random order. Each image appeared on-screen for 375 ms (a quarter of the presentation time for all four images in Experiment 1; Neumann et al., 2013). In order to avoid the possibility of low-level perceptual averaging simply due to the overlap in locations on-screen of the four images (e.g., if all images were presented centrally), each image appeared with its center at a random position along the circumference of a circle of radius 4.2 cm (see Figure 3). Within each trial, no two images appeared at an angle of less than 30° to each other around this circle. A blank screen appeared after each image for 375 ms. Prior to the presentation of the test face, a central fixation cross appeared on-screen for 1 s in order to highlight for participants that the sequence had finished and the next image would be the test face. All other details remained unchanged from the first experiment.

After completion of the practice and experimental blocks, the familiarity of the identities was checked as in Experiment 1. As expected, familiarity with Hollywood celebrities was high ($M = 8.1$, $SD = 2.1$), and familiarity with Australian celebrities was low ($M = 0.7$, $SD = 0.9$).

Results and discussion

Response data for Experiment 2 are shown in Figure 4. Data were entered into a 2 (Familiarity: Familiar, Unfamiliar) × 2 (Image Type: Exemplar, Average) × 2 (Test Face: Matching, Nonmatching) ANOVA. All factors were within-subjects. We found a significant main effect of Test Face, $F(1, 19) = 113.25$, $p < 0.001$, $\eta^2_p = 0.86$, with participants responding "present" more often for test faces that matched the preceding set ($M = 82.8\%$) than for those that did not ($M = 43.5\%$). We also found a significant Familiarity × Test Face interaction, $F(1, 19) = 7.56$, $p = 0.013$, $\eta^2_p = 0.29$. Simple main effects showed a larger "present" response for matching test faces in both the Familiar condition, $F(1, 19) = 54.16$, $p < 0.001$, $\eta^2_p = 0.74$, and the Unfamiliar condition, $F(1, 19) = 119.00$, $p < 0.001$, $\eta^2_p = 0.86$, with the interaction being driven by a larger effect for unfamiliar faces.

We also found a significant Image Type × Test Face interaction, $F(1, 19) = 4.44$, $p = 0.049$, $\eta^2_p = 0.19$. Simple main effects showed a larger "present" response for matching test faces in both the Exemplar condition, $F(1, 19) = 101.06$, $p < 0.001$, $\eta^2_p = 0.84$, and the Average condition, $F(1, 19) = 62.40$, $p < 0.001$, $\eta^2_p = 0.77$, with the interaction being driven by a larger effect for exemplars. No other effects or interactions were significant.

Once again, we find the same effect as in Experiment 1. Participants are equally willing to claim that they have seen the average of a set as they are to recognize a real exemplar. As in the previous experiment, this effect is tied specifically to the average of the images they have seen with averages of novel photos being rejected at a similar rate as novel instances. This seems to be good evidence for the proposal that viewers automatically extract the average of a set of faces of the same person—a mechanism that could plausibly underlie face learning.

Combined analysis of Experiments 1 and 2

Given the similar patterns of results for the two experiments, we carried out a mixed ANOVA to determine whether there was a significant effect of the type of presentation. Data from the two experiments were entered into a 2 (Presentation Type: Simultaneous, Sequential) × 2 (Familiarity: Familiar, Unfamiliar) × 2 (Image Type: Exemplar, Average) × 2 (Test Face: Matching, Nonmatching) ANOVA. Presentation Type was between-subjects, and the remaining factors were within-subjects. We found no main effect of Presentation Type, $F(1, 38) = 0.75$, $p = 0.392$, $\eta^2_p = 0.02$,

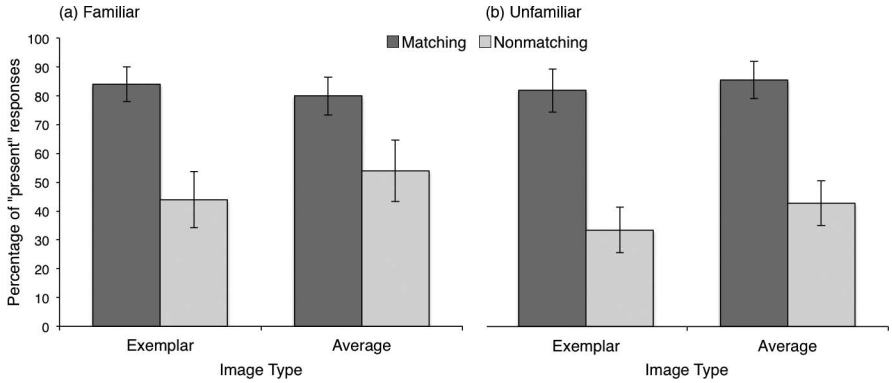

Figure 4. Mean percentage of "present" responses for (a) familiar and (b) unfamiliar test faces. Error bars represent 95% confidence intervals.

and no significant interactions involving this factor (all ps > 0.387).

We found a significant main effect of Test Face, $F(1, 38) = 295.73$, $p < 0.001$, $\eta^2_p = 0.89$, with participants responding "present" more often for test faces that matched the preceding set ($M = 81.6\%$) than for those that did not ($M = 42.4\%$). In addition, we found a significant main effect of Familiarity, $F(1, 38) = 4.69$, $p = 0.037$, $\eta^2_p = 0.11$, with a larger "present" response for familiar test faces ($M = 63.9\%$) than for unfamiliar ones ($M = 60.1\%$). There was also an almost significant main effect of Image Type, $F(1, 38) = 4.06$, $p = 0.051$, $\eta^2_p = 0.10$, with participants responding "present" more often for average test faces ($M = 63.8\%$) than for exemplars ($M = 60.2\%$).

These main effects were qualified by two interactions. The first was a significant Familiarity × Test Face interaction, $F(1, 38) = 27.13$, $p < 0.001$, $\eta^2_p = 0.42$. Simple main effects showed a larger "present" response for matching test faces in both the Familiar condition, $F(1, 38) = 148.02$, $p < 0.001$, $\eta^2_p = 0.80$, and the Unfamiliar condition, $F(1, 38) = 280.97$, $p < 0.001$, $\eta^2_p = 0.88$, with the interaction being driven by a larger effect for unfamiliar faces. The second was a significant Image Type × Test Face interaction, $F(1, 38) = 4.82$, $p = 0.034$, $\eta^2_p = 0.11$. Simple main effects showed a larger "present" response for average test faces in the Nonmatching condition, $F(1, 38) = 7.29$, $p = 0.010$, $\eta^2_p = 0.16$, but no difference between averages and exemplars in the Matching condition, $F(1, 38) = 0.03$, $p = 0.874$, $\eta^2_p < 0.01$.

To sum, these results mirrored those found when each experiment was analyzed separately as we would expect because there was no effect of presentation type. In addition, we found that participants responded "present" significantly more often when a nonmatching average was presented in comparison with a nonmatching exemplar while no difference was found for matching test faces. This result was suggested by the findings of Experiment 2 and has been confirmed here.

General discussion

We investigated set averaging for both simultaneous and sequential presentation designs. In contrast with previous work, we used color images that incorporated a large amount of variability, and each display set contained images of only one identity. For both experiments, we found a consistent pattern of results.

First, participants demonstrated good memory for exemplars for both methods of presentation. That is to say, participants were able to report accurately that a test exemplar was present (approximately 80% correct) in the display set they had previously seen. Participants found it harder to correctly report the absence of a test exemplar (approximately 40% incorrect). This inaccuracy is higher than in previous work (de Fockert & Wolfenstein, 2009; Neumann et al., 2013) and is likely to arise because in the experiments presented here all four display set images were of the same identity. Perceiving that a fifth, novel image of the same identity was not present may be more difficult than comparing this image with four previous images that all depicted different identities as was the case in previous studies.

Second, we find clear evidence for encoding the average. In both experiments, participants responded "present" in around 80% of trials in which the test face was the matching average. This suggests that viewers formed an average representation of the four images with the result that they believed they had previously seen that average when asked. Importantly, participants were significantly less likely to think they had seen an average of a different set of four images of the *same identity*. This is an important result, demonstrating that it is not simply any average image that causes viewers to respond "present"—even an average made up of different images of the same person. Therefore, participants must be forming an average representation that is specific to the images they have encountered.

Third, although we find these patterns of results for both familiar and unfamiliar faces, the sizes of the effects are larger for unfamiliar identities. Figures 2 and 4 suggest that this may be due to more "present" responses for nonmatching test faces in the familiar condition. This makes intuitive sense in that, for familiar identities, even when the test face is nonmatching (a new image of the person or a new average), participants are more likely to think they have already seen the image because they have prior experience with that identity and so may feel they have seen images that have not appeared in the experiment. However, given the relatively small effect size of this interaction, we would recommend further research before drawing any conclusions from this result.

Fourth, for the sequential method of presentation, we found a just significant interaction between Image Type and Test Face. This effect was also present in the combined analysis. Inspection of the figures suggests that viewers are slightly less likely to respond "present" for nonmatching exemplars in comparison with nonmatching averages. Again, this follows intuition in that a nonmatching average may appear more similar to the display set images than a nonmatching exemplar, resulting in more incorrect "present" responses. However, we note the very small effect size here.

Other than this slight difference for nonmatching exemplars and averages, we find no effects due to Image Type. We find no evidence in the current experiments to suggest viewers form a strong representation of the matching average while failing to represent the individual exemplars. This result is in line with some research (Neumann et al., 2013) but contrasts with other work (Haberman & Whitney, 2007, 2009). Although representing both exemplars and their average appears inefficient as a solution, it may be that hierarchical representations in working memory are formed at multiple levels of abstraction (Brady & Alvarez, 2011). Items in working memory may benefit from a combination of representations, in which information about the average can increase accuracy when exemplar memory is unreliable or inaccurate.

However, there is an alternative interpretation. Although viewers appear to remember individual exemplars, it may be that a strong representation of the average is sufficient for producing this apparent accuracy. Matching exemplars will always be more similar to the matching average than nonmatching exemplars. Therefore, simply by referencing an encoded average representation, viewers may be able to discriminate, at least to some extent, between matching and nonmatching exemplars. Indeed, previous research provides direct evidence against the idea that individual exemplars are represented (Ariely, 2001; Corbett & Oriet, 2011). The designs of the experiments presented here do not allow for a test of this interpretation, and so further research is required in order to address this specifically.

In the current work, "present" responses should only have been given in 25% of trials: those in which the test face was a matching exemplar. It is possible that participants had inflated expectations regarding the required ratio of "present" responses because of experience with the practice block (50% correct "present" responses) or psychology experiments more generally. However, previous research that controlled for this by informing participants of the correct frequency of "present" responses suggests that this possibility is unlikely to account for the results presented here (experiments 2 and 3, Neumann et al., 2013). In addition, even if participants were motivated to increase the number of "present" responses given, this should not favor any particular condition. As such, this account fails to explain why matching averages received more "present" responses than nonmatching averages.

Previous research on the ensemble encoding of faces has mainly focused on simultaneous presentation (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009; Neumann et al., 2013). However, evidence also supports the idea that facial expressions (Haberman et al., 2009) and identities (Leib et al., 2014; experiment 4, Neumann et al., 2013) are averaged across sequential presentations. Indeed, this mechanism appears to be viewpoint-invariant in that the average identity can be formed by averaging images that vary in viewing angle (Arnold & Siéroff, 2012; Leib et al., 2014). Here, by using ambient images, we provide additional evidence that ensemble encoding can operate on viewpoint-invariant representations. However, our stimuli also varied in expression, lighting, gaze direction, and numerous other real-world factors. As such, our findings provide a strong argument that encoding is invariant with regard to multiple high-level features.

Although previous research has shown that viewers represent the averages for both familiar (Neumann et al., 2013) and unfamiliar (de Fockert & Wolfenstein, 2009) sets of images containing different identities, the current work demonstrates that this is also true for sets containing different images of the same identity. Mechanisms that result in the averaging of faces across identities provide no obvious advantages (for discussion, see Neumann et al., 2013). However, it is easy to imagine the benefits of averaging together different images of the same identity. We know from previous research that the average of a set of instances can provide a more stable and robust representation of an individual (Burton et al., 2005). The formation of a within-person average may therefore provide a useful tool that could explain why viewers perform much better with familiar face recognition (Bruce, 1986) and matching (Bruce et al., 1999).

Although there may be clear advantages to creating the average representation for a single identity, previous evidence that we also average across identities may suggest that the current findings are the result of a more general ensemble encoding mechanism rather than anything specific to the creation of stable person representations. In line with this idea, developmental prosopagnosics can extract ensemble characteristics from sets of faces equivalently to controls (Leib et al., 2012), perhaps suggesting a general process that has been co-opted for faces. However, that ensemble encoding of faces is not mediated by low-level features, luminance cues, or other nonconfigural cues (Haberman & Whitney, 2009), and that it is able to operate on high-level, view-invariant information (Leib et al., 2014) may suggest at least some specialization regarding faces.

In conclusion, we have shown that viewers extract an average from different images of the same identity while apparently continuing to represent the individual exemplars. This process appears unaffected by whether the images are presented simultaneously or sequentially. Our findings may provide one account through which stable representations of identities are formed. However, representing the average alone remains a very limited statistical summary, and we recommend further investigation in order to determine whether other important information, such as the distribution of a set of faces (Burton, Kramer, Ritchie, & Jenkins, in press), is also encoded.

Keywords: set representation, ensemble encoding, face, identity, averaging

Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.323262, and from the Economic and Social Research Council, UK (ES/J022950/1).

Commercial relationships: none.
Corresponding author: Robin S. S. Kramer.
Email: remarknibor@gmail.com.
Address: Department of Psychology, University of York, York, UK.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science, 21,* 560–567.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162.

Arnold, G., & Siéroff, E. (2012). Temporal integration of face view sequences and recognition of novel views. *Visual Cognition, 20,* 793–814.

Atchley, P., & Andersen, G. (1995). Discrimination of speed distributions: Sensitivity to statistical properties. *Vision Research, 35,* 3131–3144.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22,* 384–392.

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception, 15,* 387–397.

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,* 339–360.

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51,* 256–284.

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (in press). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138,* 289–301.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology, 62,* 1716–1722.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12,* 499–504.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*(11):1, 1–13, http://www.journalofvision.org/content/9/11/1, doi:10.1167/9.11.1. [PubMed] [Article]

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17,* R751–R753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of*

Experimental Psychology: Human Perception and Performance, 35, 718–734.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72,* 1825–1838.

Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science, 319,* 435.

Jenkins, R., Burton, A. M., & White, D. (2006). Face recognition from unconstrained images: Progress with prototypes. In *Proceedings of the 7th IEEE International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 10-12 April (pp. 25–30)*. Los Alamitos, CA: IEEE Computer Society.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121,* 313–323.

Leib, A. Y., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. *Journal of Vision, 14*(8):26, 1–13, http://www.journalofvision.org/content/14/8/26, doi:10.1167/14.8.26. [PubMed] [Article]

Leib, A. Y., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia, 50,* 1698–1707.

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition, 128,* 56–63.

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11*(12):18, 1–8, http://www.journalofvision.org/content/11/12/18, doi:10.1167/11.12.18. [PubMed] [Article]

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science, 25,* 1903–1913.