Article:
Noyes, E., Parde, C.J., Colón, Y.I. et al. (2021) Seeing through disguise: Getting to know you with a deep convolutional neural network. Cognition, 211. 104611. ISSN: 0010-0277

https://doi.org/10.1016/j.cognition.2021.104611

Seeing through disguise: Getting to know you with a deep convolutional neural network

Eilidh Noyes
University of Huddersfield

Connor J. Parde, Y. Ivette Colón, Matthew Q. Hill
The University of Texas at Dallas

Carlos D. Castillo
University of Maryland

Rob Jenkins
University of York
&
Alice J. O'Toole
The University of Texas at Dallas

Keywords: face recognition, machine learning, disguise

Corresponding author:
Eilidh Noyes, Ph.D.
Dept. of Psychology
University of Huddersfield
Huddersfield, UK
e.noyes@hud.ac.uk
Tel 01484472770

Abstract

People use disguise to look unlike themselves (evasion) or to look like someone else (impersonation). Evasion disguise challenges human ability to see an identity across variable images; Impersonation challenges human ability to tell people apart. Personal familiarity with an individual face helps humans to see through disguise. Here we propose a model of familiarity based on high-level visual learning mechanisms that we tested using a deep convolutional neural network (DCNN) trained for face identification. DCNNs generate a face space in which identities and images co-exist in a unified computational framework, that is categorically structured around identity, rather than retinotopy. This allows for simultaneous manipulation of mechanisms that contrast identities and cluster images. In Experiment 1, we measured the DCNN's baseline accuracy (unfamiliar condition) for identification of faces in no disguise and disguise conditions. Disguise affected DCNN performance in much the same way it affects human performance for unfamiliar faces in disguise (cf. Noyes & Jenkins, 2019). In Experiment 2, we simulated familiarity for individual identities by averaging the DCNN-generated representations from multiple images of each identity. Averaging improved DCNN recognition of faces in evasion disguise, but *reduced* the ability of the DCNN to differentiate identities of similar appearance. In Experiment 3, we implemented a contrast learning technique to simultaneously teach the DCNN appearance variation and identity *contrasts* between different individuals. This facilitated identification with both evasion and impersonation disguise. Familiar face recognition requires an ability to group images of the same identity together and separate different identities. The deep network provides a high-level visual representation for face recognition that supports both of these mechanisms of face learning simultaneously.

People recognise the faces of friends, family, and colleagues with ease (Burton, White, & McNeill, 2010; Jenkins, White, Van Montfort, & Burton, 2011; Kemp, Towell, & Pike, 1997). These *familiar faces* can be identified accurately even in challenging image conditions (Burton, Wilson, Cowan, & Bruce, 1999; Lander, Bruce, & Hill, 2001; Noyes & Jenkins, 2017). Recognition is more difficult for *unfamiliar faces*—those we have encountered infrequently or from brief exposures. Despite the difficulty of the task, accurate identification of unfamiliar faces can be important. For example, at passport control, an officer must decide if you are, or are not, the person photographed in the passport that you present. People make errors on these types of tasks, even with high quality images, taken on the same day, with comparable illumination, and with people displaying the same facial expression (Burton et al., 2010; Kemp et al., 1997).

Techniques that increase familiarity, by increasing viewing time and/or the number and variability of exposures to a face, improve identification accuracy (Bruce, 1994; Bruce, Doyle, Dench, & Burton, 1991; Burton, Kramer, Ritchie, & Jenkins, 2016; Clutterbuck & Johnston, 2005; Dowsett, Sandford, & Burton, 2016; Jenkins et al., 2011; Menon, Kemp, & White, 2018; Murphy, Ipser, Gaigg, & Cook, 2015; Ritchie & Burton, 2017). Although it is clear that familiarity with a face improves identification accuracy, less is known about the underlying mechanisms that facilitate robust and generalisable face learning.

Burton et al. (2005) proposed *face averaging* as a mechanism for creating robust face representations. They proposed that an "average" image representation is constructed mentally from the images of a face that we encounter. Noise from variability (e.g., illumination) is minimised when multiple images are averaged. By this account, the visual system learns faces by creating an identity representation based on the average of all of the images a person experiences across multiple encounters with a face. This "person-identity prototype" [1] model creates a central tendency representation of one's actual experience from multiple images of an individual's face.

Burton et al. (2005) tested the face-image averaging mechanism using a computational model based on principal components analysis (PCA). They first morphed the images to a common shape using bi-linear interpolation of manually set key points. Next, they used these shape-normalised

[1] We introduce the term "person-identity prototype" here to distinguish this concept from the more classical notion of a face population prototype.

images to train the model with 1, 3, 6, or 9 exemplar images of each identity, or with an average image comprising these same images. A nearest neighbour match was computed at test in Mahalanobis distance for a new image of each training identity. Training with average images yielded higher accuracy than training with exemplar images. Furthermore, the greater the number of exemplar images in the average, the higher the identification accuracy, as measured by the nearest neighbour metric. Therefore, this averaging mechanism proved effective in boosting identification performance using information from variable images.

Burton et al. (2005) also looked at the effect of averaging on human perception. They measured reaction times on a name-to-face matching task of averaged celebrity faces consisting of 3, 6, or 9 images. Reaction times decreased as the number of images included in the average increased. Additionally, reaction times for average faces were lower than for single images. In terms of accuracy, participants performed better on averaged faces in trials where the name matched the face, though in mismatched trials there was no difference between single image and averaged conditions.

Exemplar averaging also improved identification of a 2005 industry-standard face recognition algorithm from 54% to 100% correct (Jenkins & Burton, 2008). This was important, because algorithms at that time struggled to identify images that varied in illumination, pose, and expression (O'Toole et al., 2007; Phillips, Moon, Rizvi, & Rauss, 2000). The benefit of using averages is that they retain aspects of a face that are diagnostic of identity, while removing noise from the variability of the images encountered (Bruce, 1994; Jenkins et al., 2011).

The person-identity prototype model of familiarisation was tested for psychological relevance in Kramer, Ritchie, and Burton (2015). Participants viewed four simultaneous images of a previously unknown identity, followed by a single test image. Participants decided whether the test image was present in the four-image display. The test image was either: a.) a learned image of the identity; b.) a previously unseen image of the identity; c.) an average image of the identity created from the four learned images; or d.) an average image created from four previously unseen images of the identity. Human accuracy was higher for an average image composed of four exemplar images that were learned by a participant, than for an average image composed of four novel images of the same identity. Kramer et al. (2015) concluded that human face representations are based on mental averages of the exemplar images we encounter.

The person identity prototype theory makes progress in understanding the mechanisms that may be responsible for learning within-person image and appearance variation. However, successful face recognition requires skill on two critical tasks: learning within-person image variation and distinguishing among faces of similar appearance. Face-image averaging addresses the first task—referred to in the literature as "telling faces together" (Andrews, Jenkins, Cursiter, & Burton, 2015). A different type of mechanism is needed for distinguishing among faces of similar appearance ("telling faces apart"). This latter skill is at the core of the long-embraced definition of "face expertise" (Diamond & Carey, 1986; Gauthier & Nelson, 2001). Expertise has been considered, historically, in the following context. Faces consist of the same features (e.g., eyes, nose, and mouth) arranged into a similar configuration. The information useful for discriminating among different identities is subtle and difficult to articulate/quantify. Because people are able to discriminate among highly similar faces, humans have sometimes been considered "face experts" (though see Rossion, 2018; Young & Burton, 2018). This historical definition of face expertise was presented agnostic to the role of face familiarity.

The concept of expertise fits well with classic face prototype theory, which has been a motivating principle of many studies in face processing, over decades. This theory assumes that face representations encode the information in a face that makes it different from a *population* face average (cf. Valentine, 1991). By this account, face codes capture the uniqueness of a face relative to a population. Classic prototype theory is supported by data on the effects of face typicality (Light, Kayra-Stuart, & Hollander, 1979), as well as the benefits of facial caricaturing for creating a good likeness of a person (Benson & Perrett, 1991).

The prototype theory of face perception (Valentine, 1991) and the person-identity prototype model of Burton et al. (2005) are not mutually exclusive. Instead, these two models address different aspects of the face recognition problem—discriminating among similar face identities and dealing with appearance/image variability within an identity. To be clear, in classic prototype theory (Valentine, 1991, 2001), "average" refers to the central tendency of the population of face identities known to the individual. In the person-identity prototype model, "average" refers to the central tendency of the *images* of a single face seen by an individual. It is worth noting that classic prototype theory considers the physical structure of the face itself, independent of image and appearance variation, and assumes that there is only one representation of a face (cf. O'Toole, Castillo, Parde, Hill, & Chellappa, 2018). The person-identity prototype theory also posits a single

representation of each face, but uses a representation that reflects actual images encountered in the real world.

Notably, in the majority of studies in which classic prototype theory has been examined, especially in the early face processing literature, participants were tested with unfamiliar faces (though see Hellawell, & Hay, 1987) and viewed only a single image of each identity. Thus, the role of familiarity in discriminating among similar faces has been studied less than its role in generalising recognition across images (Young & Burton, 2017).

In addition to classic prototype theory, which considers differences from a central average, and person-identity prototype theory, which considers image/appearance differences around an identity-specific average, there are also differences *between* individual exemplar faces. How do we learn contrast between individual faces—especially those of similar appearance, in the context of image variability? This question differs from the one at the core of prototype theory, where contrast is always relative to a central prototype or population average. Along these lines, Mundy, Honey, and Dwyer (2007) demonstrated that people are better able to distinguish two similar faces (from a single image of each), if they have the opportunity to compare and contrast the two face images during face learning. Accuracy was higher when people alternately viewed each image, than when they saw the same images repeated multiple times in sequence. The role of contrast was supported further, because discrimination accuracy improved with simultaneous learning (viewing the two identities side by side) over successive exposures (viewing the two images in sequence). The authors attribute the contrast benefit to identity adaptation (Leopold, Rhodes, Müller, & Jeffery, 2005; Mundy et al., 2007; Rhodes & Jeffery, 2006).

Cavazos, Noyes and O'Toole (2019) replicated the advantage of distributed over sequential viewing order in face learning, using a face recognition task with own and other-race faces. They also tested a larger number of identities (N = 36) than Mundy et al. (2007). One caveat to the benefit of distributed learning is that the images must be recognisable to the participants as the same identity. Distributed presentation did not aid recognition when images were too variable to be readily grouped together by identity. Together, these studies indicate that viewing orders that facilitate perceiving contrasts between individual identities result in better face learning.

Another way to assess the role of between-person contrast in face learning is to manipulate the similarity of the learning faces. Dwyer and Vladeanu (2009) tested participants' ability to learn target faces from among faces of similar appearance, dissimilar appearance, or in isolation. At test,

participants were instructed to identify the target face from a line-up array. Matching accuracy was highest for identities learned from among faces of highly similar appearance. The similarity of learning items may have forced participants to focus on the distinguishing features for an identity. The authors suggested that this process might involve adaptation.

In an updated version of the person prototype model, Kramer, Young, Day, and Burton, (2017) implemented a linear discriminant analysis (LDA) to optimise separation by identity in their computational model. The model uses labelled data to learn identity classification, which can then be tested on new images of the face. They began by shape-normalizing faces in a preprocessing stage. Next, they created a face space with PCA and applied LDA to enhance the distances between individuals. The PCA+LDA model resulted in more accurate face identification than the LDA model alone. The authors conclude that face familiarity is achieved by learning idiosyncratic variation for individual faces.

Another study modeled human familiarity with individual faces by "fine-tuning" a DCNN (Blauch, Behrmann & Plaut, 2020) to individual faces. The technique of fine-tuning, applied to datasets of many identities sharing some demographic feature (e.g. race) was suggested by O'Toole et al. (2018) for adapting to the statistics of local populations. This might, for example, produce better performance for faces for which the network was tuned (see Cavazos et al., 2020). This has a downside for "other-race faces" because, it limits the ability of the network to represent critical features not characteristic of the learned race. The contrast implemented in the network of Blauch et al. (2020) applies fine-tuning to a small random set of "familiar faces". This has the effect of weighting high-level visual face features (i.e., in lower layers of the network) to better code for this set of "familiar faces". Just as fine-tuning with large demographically homogeneous sets of faces has a downside for race, it might also have a downside as a model of familiarity. Specifically, it will likely limit the quality of representations for newly encountered faces and it will also make it computationally impractical to add a new identity to the set of known identities. This impractically stems from the fact that the fine-tuning is based on a relatively small number of faces would have outsized influence on the capacity of the network to represent any face. As we detail shortly, the approach we take to familiarisation implements contrast at the level of individual known faces in a way that does not re-weight basic face features learned in training the deep network for faces in general.

Combined, the majority of studies that have demonstrated a benefit of contrast learning on recognition have not considered within-person image variability (though see Cavazos et al., 2019; Kramer, Young, & Burton, 2018; Kramer et al., 2017). Here, we approached the problem of becoming familiar with individual face identities by considering the problem of telling similar faces apart, and determining when variable images show the same person. We use images of disguised faces to investigate this problem, because different types of disguise selectively challenge the two skills needed to identify people. Impersonation disguise—a deliberate change in appearance to look like someone else—poses challenges for face discrimination, and evasion disguise—a deliberate change in appearance to look unlike one's self—challenges face recognition over variable images.

We employed recent computational models based on DCNNs to provide a viable test-bed for exploring face recognition under disguise from both perspectives. In the last five years, DCNNs have made important strides in achieving face recognition from images that show substantial within-person variability, including over variable images and personal appearance. In that sense, they are beginning to emulate human face recognition as it applies for familiar faces (O'Toole et al., 2018). Previous-generation models, though successful at recognising faces from within a single viewpoint and with reasonably well controlled illumination, are far less successful with even small changes in imaging parameters and appearance (O'Toole, An, Dunlop, Natu, & Phillips, 2012; Phillips, Hill, Swindle, & O'Toole, 2015; Turk & Pentland, 1991; Zhao, Krishnaswamy, Chellappa, Swets, & Weng, 1998). Older models, therefore, demonstrate skills similar to those humans show for unfamiliar faces (Burton et al., 2010, 1999; Kemp et al., 1997; Phillips, & O'Toole, 2014; White, Kemp, Jenkins, Matheson, & Burton, 2014).

Notably, the computations used in DCNN models are inspired by the primate visual system and consist of cascaded layers of local convolution and pooling operations. DCNNs have *general* experience with faces, because they are trained to classify identity from variable images, using 10's of thousands of training identities, from variable ('in-the-wild') images of the faces. Once trained, the output "identity" units for the training faces are removed. In this "decapitated" state, the system can generate a representation of any arbitrary face image at its top-layer. These representations come from the network's *general face knowledge history* (O'Toole et al., 2018), and support some degree of identification generalisation across image/appearance variation, even

for faces not used in training (i.e., unfamiliar faces). This generic system, however, retains no knowledge of individual faces.

To learn specific identities from one or more images, a DCNN is typically trained by adding a new layer of identity units at the output layer. A second small network learns to classify the DCNN-generated face image representations of these new identities based on a small to moderate numbers of images of each (cf. O'Toole, et al., 2018). For present purposes, DCNNs generate a face space in which identities and images co-exist in a unified computational framework (Hill et al., 2019; O'Toole et al., 2018). It is possible, therefore, to investigate the effects of learning identity contrast in the context of image/appearance variability, using the high-level visual representation produced by the DCNN.

Here we used disguise to test how familiarity with a person enables more robust face recognition with variation in appearance. In a recent test of human face matching performance for disguised faces (Noyes & Jenkins, 2019), familiar participants had a strong advantage over unfamiliar participants in seeing through disguise (Noyes & Jenkins, 2019).

The goal of the current study was to test learning mechanisms that can support human-like face recognition under evasion and impersonation disguise. Here, human-like performance includes both familiar and unfamiliar face recognition. We used a DCNN as a model of face learning that begins as a face recognition system that is "unfamiliar" with the individual test faces. In Experiment 1, we show that the DCNN, in this state, performs comparably to humans who are unfamiliar with disguised (evasion, impersonation) faces. In Experiment 2, we test a face-averaging mechanism to familiarise the DCNN with the test identities. The averaging mechanism is applied to the *high-level visual face representations* produced at the top layer of DCNN units. This approach differs from that of Burton et al. (2005) who averaged information directly available in face *images*. We demonstrate how our work differs from previous averaging and contrast learning methods in Table 1. Face representations at the top level of a DCNN have been shown to represent face identity, in addition to image characteristics (cf. O'Toole et al., 2018), including viewpoint and illumination (Hill et al., 2019; Parde et al., 2017, Parde 2020). This is a strong departure from modeling familiarity with mechanisms aimed at altering the representation of image-based information in a face space. We show that averaging the high-level visual representations that result from the DCNN improves performance on trials that benefit from

learning within-person variation. The cost of this improvement, however, is decreased performance for images of different identities that are similar in appearance.

Paper	Pre-processing of image	Similarity computation (Unfamiliar)	Calculation	Post Processing (Familiar)	Result	Limitations
Burton et al. (2005)	Shape normalisation	PCA	Nearest neighbour matches. 'Hit' if nearest neighbour the same ID.	Averages – (Study 3) created by morphing together shape free images. Arithmetic means at pixel level. Number of images contributing to average varied (3,6,9). Studies 4 and 5 - created as described for Study 3 with the additional step of morphing the shape free average back to the average 2D shape outline of the individual in the images.	Image average based systems outperform instance based versions	Shape information extracted through pre-processing Viewpoint generalization limitations
Kramer et al. (2017,2018)	Shape normalisation	PCA	Nearest centroid.	PCA+LDA – trained on identity	" "	" "
This paper	None	DCNN	Similarity score compared against criterion.	Averaging – of DCNN-generated high level visual representations for images used to familiarise the identity. Number of images contributing to the average varied. SVM – trained to discriminate each identity from all other identities.	Improved performance on same-identity trials Improved performance on same-identity trials, high performance on different-identity trials.	

Table 1. The current work uses a DCNN and averaging of DCNN-generated face representations to model face familiarity. The table outlines how this approach differs to previous models.

In Experiment 3, we implemented a version of contrast learning to capture learning of both within- and between-person variation in the high-level visual representations produced by DCNNs and measure the contribution of each to face learning. We show that general face learning and averaging across image variability in these high-level visual representations cannot solve the problem of recognising disguised faces at a level that compares with familiar humans. Instead, it is necessary to reshape the top-layer DCNN face space by providing the algorithm with information about both within- and between-person variation.

Experiments

Experiment 1 – Unfamiliar Identification of Disguised Faces by a DCNN

In Experiment 1, we tested DCNN accuracy on the FAÇADE face-matching test; a face-matching task that includes photos of faces taken in evasion disguise, impersonation disguise and no disguise

(cf. Noyes & Jenkins, 2019). We compared DCNN performance against human accuracy, measured in Noyes and Jenkins (2019).

First, we asked whether the DCNN could perform the face recognition task with disguised images, and if so, whether the level of performance was more similar to unfamiliar or familiar human participants. Second, we asked whether the pattern of accuracy for the DCNN across evasion and impersonation disguise mirrored the human pattern. Based on the human data from Noyes and Jenkins (2019), we expected disguise to impair DCNN identification accuracy. We also expected DCNN identification accuracy to be lower for evasion than impersonation trials. Because the DCNN representation is based on general knowledge of faces, rather than on specific identities, we expected machine performance to mirror that of humans who are unfamiliar with the faces.

Method

Materials

The images used in this study were the 156 matching task image pairs from the Noyes and Jenkins (2019) FAÇADE image set. The database consists of images of models photographed in no disguise, evasion disguise, and impersonation disguise conditions (see Figure 1). In the creation of this unique dataset, each person provided his or her work identification photograph to act as a "reference image". This reference image showed the model with no disguise. All other images of the model (in disguise and no disguise) were compared against this reference image. A second "no disguise" image of each identity was taken during a photograph session. This second no disguise image was paired with the reference image to provide a 'same-identity- no-disguise' image pair in the matching task. The 'different-identity no-disguise' image pairs consisted of the reference image of one identity and the no disguise image of another identity.

Next, the models created both an evasion and impersonation disguise. For the evasion disguise, they were asked to change their appearance to look "as different as possible" from their reference image. Models often went to great lengths to do this, by wearing wigs, growing/shaving beards, changing their hairstyles, and by applying or removing make-up worn in the reference image (See Figure 1). The same-identity disguise image pairs consisted of the reference image and the image of the model in evasion disguise.

Figure 1. Example of same-identity (top row) and different-identity image pairs (middle and bottom row) in no disguise and disguise conditions. The far-right image in the top row provides and an example of evasion disguise, the far-right image in the middle row provides an example of impersonation similar disguise, and the far-right image in the bottom row is an example of impersonation random disguise.

For the impersonation photographs, the models were asked to change their appearance to look like the reference photograph of two other individuals in the data set. One of these individuals was selected due to high similarity in appearance to the model, whereas the other one was chosen at random from the available reference images (within gender). These images were used to create different-identity disguise image pairs, which consisted of a reference image and an impersonation image.

The resulting image pairs for the DCNN matching task consisted of *same identity trials*, and two types of *different identity trials* (see Figure 1). The different person trials were categorised as *different similar* (the two people photographed were paired intentionally, because they have a similar appearance naturally) and *different random* (the pair was created by selecting two identities of the same gender from the dataset at random).

All image trial types (same identity, different similar, different random) were included in the *no disguise* and *disguise* conditions. Every image pair included a reference image and another image that varied according to disguise condition and trial type. For the *no disguise* condition, this image was always a no-disguise image, but was either the same identity as the reference image or a different identity, depending on the trial type. For the disguise conditions, this image was an *evasion disguise* for same identity trials, and an *impersonation disguise* for different identity trials (see Figure 1).

Deep Convolutional Neural Network: Structure and Training

The DCNN we tested (Sankaranarayanan, Alavi, Castillo, & Chellappa, 2016) consists of 7 convolutional layers, 3 pooling layers, and 3 fully connected layers. These layers work to extract features (the computer's numerical representation of an image) from input images. The network uses Parametric Rectified Linear Units (PReLUs) between each layer. This method was chosen, because it allows negative output values and improves the convergence rate (Sankaranarayanan et al., 2016). When the network is used to extract features from images, features are taken from layer fc7. This layer has 512 units. The network was trained using a publicly available dataset (CASIA-Webface Dataset) (Yi, Lei, Liao, & Li, 2014). This database consists of approximately half a million images and includes over 10,000 identities.

Procedure

All images from the FAÇADE matching task were processed by the face-identification DCNN, which generated a 512-element feature vector for each image. As noted, this face representation consisted of the activation levels of the top-layer DCNN units in response to the input image (Sankaranarayanan et al., 2016). We refer to this, henceforth, as the *DCNN face representation*. For any given pair of test images, similarity between faces was calculated by measuring the cosine distance (similarity) between the face representation vectors of the two images in the pair.

These similarity scores were compared, in turn, against a criterion value to determine whether the similarity between the two DCNN representations was high enough to consider the pair a matched identity. Image pairs were assigned a "same" or "different" identification response depending on whether the similarity of the images was greater than the criterion. The criterion was selected to mimic standard use of computational models of face recognition, which typically employ a decision threshold chosen to keep false alarms (i.e., images of different people incorrectly identified as the same person) low. Specifically, we set the threshold to 0.46, the value needed to maintain a false alarm rate of 0.1% for "in-the-wild" images. This criterion corresponded to the similarity score at which only one in one-thousand non-match pairs in the IJB-A database would be falsely judged a match (Klare et al. 2015).

To determine DCNN accuracy, the DCNN response (same- or different- identity) was compared against the correct response for each image pair and the percentage of correct responses was calculated for each image type and disguise condition.

Results and Discussion

The DCNN achieved perfect matching accuracy for no disguise image pairs for both same and different identity trials, and for impersonation trials. Identification accuracy for evasion trials was equal to chance. There are striking similarities between the patterns of performance of the DCNN and the data from unfamiliar human observers reported by Noyes and Jenkins (2019) (Figure 2).

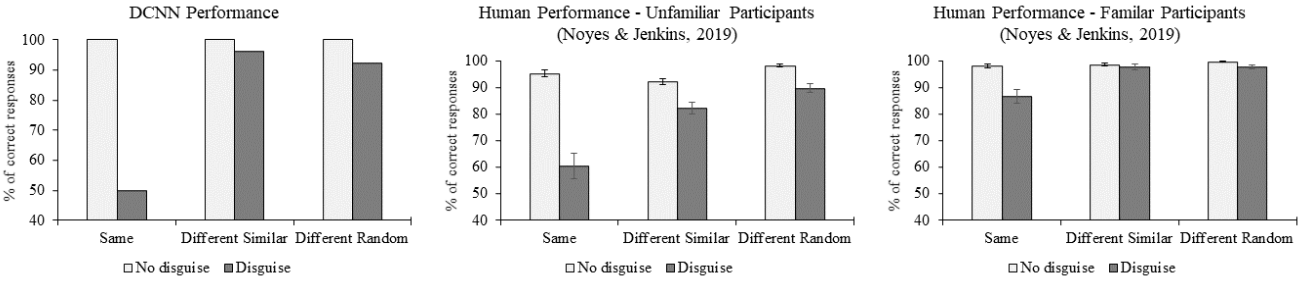

Figure 2. Patterns of performance accuracy for the DCNN mirrors that of unfamiliar human participants from Noyes and Jenkins (2019).

The low accuracy of the DCNN for evasion disguise exposes a weakness in the DCNN that humans share. Because DCNNs are currently the state-of-the-art for computer-based face recognition, this might pose a security threat in applied scenarios. From a research utility perspective, however, this finding provides an opportunity to explore mechanisms of human face learning. To test this, we incorporated within person variability in face learning and contrast-learning methods of familiarity. The methods and results of these manipulations are provided in Experiments 2 and 3.

As noted, humans familiar with the people/faces tested in Noyes and Jenkins (2019) matched disguised faces with higher accuracy than the DCNN (Figure 2). Next, we return to theories of face learning in an attempt to improve DCNN performance.

Experiment 2. Familiarity through 'averaging' DCNN Representations

In this experiment, we modelled human familiarity by trying to increase the accepted range of within-person variability through face representation averaging. Whereas previous studies attempted to create a more stable face representation through image averaging, here we averaged the DCNN face representations. Averaging images places emphasis on low-level (retinotopic) visual representations. In contrast, averaging DCNN representations operates on higher-level

visual codes that index face identity categorically across image and appearance variation. These might be considered analogous to codes in the fusiform face area (Grill-Spector & Weiner, 2014).

In Experiment 1, we compared the similarity of feature vectors for the two images in each pair. In Experiment 2, we replaced the single feature vector obtained for the reference image with an average feature vector for the *identity*. To achieve this, we entered new data into our algorithm in the form of "ambient" images of each of our models. The DCNN calculated a feature vector for each of these new images. These face representation vectors were then averaged for each identity. We hypothesised that an averaged representation should be a more robust identity code than a single comparison image. We further hypothesised that this familiarisation method should be especially effective for improving the model's ability to tolerate challenging within-person variation from evasion disguise.

Specifically, in Experiment 2 we expected that DCNN matching accuracy would increase for evasion disguise items. However, it is less clear how face representation averaging will affect different-identity trials. This scenario has not been tested by previous experiments. Average representations may reduce accuracy on different person trials, because expanding tolerance for within-person variation may "blur" the boundaries of identity for different faces, especially when those faces are of similar appearance.

Methods

Stimuli

The FAÇADE models provided a set of no-disguise ambient images (M = 20 per identity) to be used as "*familiarisation images"* for the DCNN. These images included naturalistic appearance changes across time, such as changes in hairstyle, pose, expression and illumination. Notably, the images did not include any deliberate attempt to disguise appearance (see Figure 3). The test stimuli were again the 156 FAÇADE matching task pairs.

Figure 3. Example of the types of ambient images provided by the models for use in this experiment. The images in Figure 3 are illustrative of the experimental stimuli and depicts author EN who did not appear in the experiments.

Procedure

An average feature vector was obtained for each identity as follows. First, all familiarisation images were processed by the DCNN. This produced a feature vector for each familiarisation image. These feature vectors were then averaged for each identity with increasing group sizes (*N* = 3, 5, 10, 15 and 20 images) to simulate increasing familiarity (cf. Clutterbuck & Johnston, 2002, 2005). For each group size condition (with the exception of group size *N* = 20), a random sample of *N* images was drawn, and the experiment was repeated 5 times. The average performance of these 5 iterations served as the accuracy measure. The procedure to obtain similarity scores largely mirrored that used in Experiment 1, with the exception that the newly calculated average identity feature vector replaced the previous single feature vector from the reference image.

Results and Discussion

Experiment 1 revealed that the DCNN achieved chance levels of matching performance for evasion trials. In Experiment 2, accuracy for the familiarised DCNN for evasion image pairs increased with the number of images averaged (See Table 2). Feature vectors consisting of 20 images increased DCNN matching accuracy for evasion pairs to 65% over its original level of chance performance.

		Number of Gallery Images				
		0	3	5	10	20
Same ID	No Disguise	100	100	100	100	100
	Disguise (Evasion)	50.0	46.9	52.3	64.6	65.4
Different ID Similar	No Disguise	100.0	99.2	98.5	96.2	96.2
	Disguise (Impersonation Similar)	96.2	95.4	87.7	85.4	87.7
Different ID Random	No Disguise	100.0	100.0	100.0	100.0	100.0
	Disguise (Impersonation Random)	92.3	96.2	95.4	95.4	96.2

Table 2. An improvement in accuracy for same-identity trials came at a cost of more errors on different-identity trials. Reported numbers are the percentage of correct responses.

Although performance increased for evasion trials, accuracy *declined* for different identity trials when the reference and model images were of similar appearance. On different identity (similar appearance) trials, accuracy decreased with the number of images included in the average. Thus, although averaged feature vectors expanded the acceptable range of within-identity appearance variation, it decreased the ability of the DCNN to discriminate among different people of a similar appearance.

It is evident from the results of Experiment 2 that increasing the accepted range of variability of a face does not solve all aspects of the face recognition problem. Human familiarity relies not only on creating a stable representation of an identity, but also on learning between-identity contrasts and stable representations of these other known identities. In Experiment 2, averages were created only for the reference identity. Notably, the familiar humans in Noyes and Jenkins (2019) were familiar with all identities from the FAÇADE dataset. We incorporate these factors into Experiment 3.

Experiment 3. Contrast Method of Familiarity

In this experiment, we incorporated a learning mechanism to enhance between-person contrast, and model tolerance of within-person variability. We used Support Vector Machine (SVM) classifiers to simulate the between-person contrast familiarity mechanism with all identities in the FAÇADE dataset.

Methods

Stimuli

As in Experiment 2, the familiarisation images consisted of the ambient image sets provided by the FAÇADE models. The test stimuli remained constant across all experiments and were the 156 FAÇADE matching task pairs.

Procedure

SVM learning was used to familiarise the DCNN with each identity in the FAÇADE database. An SVM is a kernel-based binary classification technique. The goal of an SVM is to optimally separate data by finding a vector or hyperplane in a data space that allows for the greatest margin of separation between the closest points of the two distributions (Vapnik, Golowich, & Smola, 1997).

This hyperplane or vector that best separates the target distribution from the rest of the data is called the "support vector" for that distribution.

Separate support vector classifiers were trained for each identity from the FAÇADE database. Images of the identity to be learned were labelled as positive examples, and all other images were labelled as negative examples. Each identity took a turn as the positive example to create a support vector for each identity. Thus, each identity's support vector provided a representation of a single identity, learnt in contrast to all other identities in the dataset. We believe that contrast is a mechanism that is available in the real world to help us learn differences between faces (Cavazos et al. 2019, Roark et al. 2007). To incorporate within-person variability, the number of images that contributed to the SVM was varied from 3 to 20 images, as in Experiment 2.

In Experiments 1 and 2, the similarity of images was calculated as the cosine similarity of the DCNN's top-level feature vectors for the images in an image pair. In Experiment 3, this top-level feature output was used as the input to each SVM. This produced an "identity vector" for each image—a vector in which each element represents the likelihood that the image is the person that the SVM was trained to classify. Similarity scores for each image pair were calculated by computing the cosine similarity of these new SVM-based identity vectors to determine whether the two images in the image pair should be classified as the same identity, and if so, whether the identity classification would be correct. More formally, the similarity score was again compared against a criterion, to determine image classification—same or different identity.

The SVM-based method operates in a qualitatively different similarity space than the method implemented in our previous experiments, and therefore required a new criterion score. To select this score, support vector classifiers were created from a sample of the IJB-A dataset (Klare et al. 2015) to match the number of models in the FAÇADE dataset. The cosine similarity between the identity support vectors and the corresponding true identity vectors (for test images) was calculated, and the similarity value at the 0.1% false alarm rate was taken as the criterion cut off.

Results and Discussion

Accuracy on same-identity disguise trials increased when contrast familiarisation was implemented with support vectors. Performance increased with the number of ambient images used in the support vector training, from 60.77% for 3 images to 71.54% for 20 images. This occurred whilst maintaining generally high performance across the other image types (see Table

3), similar to that observed by Noyes & Jenkins 2019 for familiar viewers (see Figure 2). Notably, familiar participants in Noyes & Jenkins 2019 had high levels of personal familiarity with the models across time whereas the SVM results are based on max N=20 images of each model.

		Number of Gallery Images			
		3	5	10	20
Same ID	No Disguise	100	100	100	100
	Disguise (Evasion)	60.8	67.7	71.5	71.5
Different ID Similar	No Disguise	97.7	99.2	100.0	100.0
	Disguise (Impersonation Similar)	95.4	97.7	98.5	98.5
Different ID Random	No Disguise	99.2	100.0	100.0	100.0
	Disguise (Impersonation Random)	94.6	96.2	98.5	98.5

Table 3. Learning within person variability and between person contrast improved performance for all image types. The more within and between variability learnt, the higher the performance. Reported numbers are the percentage of correct responses.

This familiarity method expanded the range of accepted appearances for an identity, but critically, also refined the representation. In addition to an increase in accuracy on same identity trials, different identity errors were reduced.

General Discussion

Successful recognition means that we do not confuse people who resemble one another, and that we can perceive identity-constancy across variable images of the same person. Although it is theoretically clear that successful familiar face recognition requires an ability to group images of the same identity together and different identities apart, few studies have considered both of these mechanisms of face learning simultaneously. Where both have been tested (Kramer et al. 2017), the averaging mechanism was implemented at a pre-processing (image-level) stage, prior to the application of the neural network. This implements the averaging mechanism (morphing), at a stage analogous to low-level visual processing (i.e., at the image level)—where it seems very much at odds with the strong retinotopy of early visual areas. The DCNN model, by contrast, allows for

both processes to be applied to high-level visual representations. The advantage of this model over previous approaches stems from the fact that high-level visual representations, processed through early visual processing mechanisms are categorical in form. In these representations, categorical variation between identities is enhanced relative to variation of image data around identities (Hill et al. 2019, Parde et al. 2020). Critically, this is more consistent with what is known about how retinotopic visual representations in early vision are converted into categorical visual representations in inferotemporal cortex, than is the implementation of morphing in early visual areas.

Deep networks provide a compelling framework for exploring theories of face learning and the mechanisms that support human performance for familiar face recognition. This perspective can be tested in a particularly useful way when we consider recognition of disguised faces, which challenge both the skill of telling people apart (evasion) and of telling people together (impersonation). The network can be used to examine how the representation of an individual face evolves as the "system" is exposed to increasing numbers of image exemplars. DCNNs can support separate learning mechanisms that encourage increased distinctions between individuals and/or decreased distinctions among images of the same person. In the context of recent work on how humans recognise familiar and unfamiliar people under evasion and impersonation disguise (Noyes & Jenkins, 2019), deep learning can offer insight into face-learning mechanisms that go beyond modeling familiarity simply as "seeing more images of an individual". In our work, we asked how increased image exposure could be exploited to best create a representation that balances image generalisation with identity separation.

To begin, unfamiliar face identification for both disguised and non-disguised faces was modeled here using the *general face knowledge history* of the system (O'Toole, et al. 2018) (Exp. 1). Operating in this mode, the DCNN only has general expertise for faces, which it acquires through training with a vast dataset of labeled face images. In this baseline configuration, the network performs like a human participant with a lifetime of experience with faces, but with no knowledge of the individual faces in the test set. Both the DCNN, and humans unfamiliar with the faces we tested (Noyes & Jenkins, 2019), showed extremely poor performance for evasion disguise, and moderately impaired performance for impersonation disguise. Notably, even without specific knowledge of individual faces, the DCNN and humans accurately recognised undisguised images of the same individuals. The ability of these networks to recognize "unfamiliarized" (i.e.,

previously unknown and undisguised) faces is well known from the computational literature at large (e.g., Sankaranarayanan, Alavi, Castillo & Chellappa, 2016; Schroff, Kalenichenko, & Philbin, 2015; Taigman, Yang, Ranzato, & Wolf, 2014), and from the psychological literature (Blauch, et al. 2020; Hill et al., 2019; Parde et al. 2017). These unfamiliar representations, however, were not sufficient to overcome the challenge of disguise.

Previous studies indicate that familiarity improves human face identification performance and supports generalised recognition over image variation, including variation due to challenging levels of disguise (e.g., Noyes & Jenkins, 2019). We tested learning mechanisms that promote recognition robustness in different ways. In adding a familiarity component to the baseline DCNN, we drew on evidence that exposure to multiple diverse images contributes to recognition accuracy (Burton et al., 2016). We also drew on previous models' use of image-based averaging to stabilise identity-specific information across images, while minimising image variation noise (Burton, Jenkins, Hancock, & White, 2005; Jenkins & Burton, 2008; Jenkins et al., 2011; Kramer et al., 2017). The approach we took to familiarisation via averaging, however, differed in important ways from these previous models, which have averaged face images through a morphing process prior to processing the images through the network (Kramer et al., 2018, 2017). Instead, we processed images through the deep network to produce "identity" representations at the top layer of the network, as a first step. Only then, did we average the representations of the face that emerged at the top layer of the network. This takes advantage of the general face knowledge history of the network to produce an identity representation that benefits from the DCNNs general ability to "untangle" identity and image-based information in face images (Hill et al., 2019; O'Toole et al., 2018; Parde et al., 2020).

Using the DCNN's ability to transform an image-based, retinotopic representation into a categorical representation seems consistent with the way human vision operates in the natural world. Indeed, when we encounter an image of either a known or unknown individual, this image is processed first through low-level visual mechanisms. What emerges at higher order visual areas represents face identity more saliently than any particular face image experienced (Parde et al., 2020). A decision of whether or not an individual is known is almost certainly made at this post-processing stage. If other images of the individual have been encountered, the averaging hypothesis would posit that the representation in memory consists of a better (i.e., more noise-resistant) representation made from multiple image exposures.

The averaging approach we employed resulted in better recognition of faces in evasion disguise, but *less accurate* recognition of impersonation disguise. The reason for this is clear. Evasion disguise challenges the ability of the network to group widely variable images of a person together into a single identity. Representation averaging extends the acceptable range of variation, to subsume images of people trying to evade identification. This mechanism facilitates learning of within-person variability at the cost of accepting some impersonators as examples of the identity they are trying to look like.

Whilst averaging goes part way toward an account of familiarity, a separate contrast mechanism is needed in face learning to reinforce distinctions between similar identities. This mechanism must draw on differences between an individual identity and all other identities in the set of faces we know. The SVM provides a mechanism for both telling faces apart and clustering multiple images of each identity together. Specifically, the SVM, which we applied to learning individual "familiar" identities, via their DCNN-generated representations, improved accuracy on evasion trials, while maintaining high performance for impersonation trials. For each individual in the "familiar set", the SVM learns to separate the representation of the images of each identity from those of all other identities in the set. Thus, the network learns what makes a person different from a local population of other people with whom it is familiar. Both of these mechanisms – averaging and contrast – must come together for the facilitation of successful face learning that can see through evasion disguise and detect impersonation.

Returning to the larger issues, these findings inform theories of face learning and contribute to a revision of our understanding of face space theory. The classic face space model, in the original metaphorical formulation implemented with eigenfaces, assumes that a face can be represented by a single image of each identity, as a point in the space, and therefore does not accommodate within person variability in appearance (for a different perspective on formulation on classic face space see Lewis, 2004) . A more recent tenant of the face space model incorporates this variability in the form of an identity region within a face space that encompasses many exemplar images of a single identity in close proximity to each other (Hill et al., 2019). This occurs in a way that allows for image location in the identity-space to be structured according to the type and salience of image variation (e.g., viewpoint, illumination). When DCNNs are successful at face recognition, it follows that identity regions for different individuals remain distinct. As psychological theories of face familiarity progress, they must begin to consider mechanisms for distinguishing among

individual face identities and for seeing individual identities despite appearance and image variability. Deep networks for face recognition offer a base from which both sides of this question can be studied.

Acknowledgements

Funding provided by National Eye Institute Grant R01EY029692-01 to AOT and by the Intelligence Advanced Research Projects Activity (IARPA). This research is based in part on work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012 and 2019-022600002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorised to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology* , 68(10), 2041–2050. Retrieved from https://doi.org/10.1080/17470218.2014.1003949

Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*. Retrieved from https://doi.org/10.1080/09541449108406222

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2020). Computational insights into human perceptual expertise for familiar and unfamiliar face recognition. Cognition, 104341. Retrieved from https://doi.org/10.1016/j.cognition.2020.104341

Bruce, V. (1994). Stability from Variation: The Case of Face Recognition the M.D. Vernon Memorial Lecture. *The Quarterly Journal of Experimental Psychology Section A*. Retrieved from https://doi.org/10.1080/14640749408401141

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition*, 38(2), 109–144. Retrieved from https://doi.org/10.1016/0010-0277(91)90049-a

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: the power of averages. *Cognitive Psychology*, 51(3), 256–284. Retrieved from https://doi.org/10.1016/j.cogpsych.2005.06.003

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity From Variation: Representations of Faces Derived From Multiple Instances. *Cognitive Science*, 40(1), 202–223. Retrieved from https://doi.org/10.1111/cogs.12231

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*. Retrieved from https://doi.org/10.3758/brm.42.1.286

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face Recognition in Poor-Quality Video:

Evidence From Security Surveillance. *Psychological Science*. Retrieved from https://doi.org/10.1111/1467-9280.00144

Cavazos, J. G., Noyes, E., & O'Toole, A. J. (2019). Learning context and the other-race effect: Strategies for improving face recognition. *Vision Research*. Retrieved from https://doi.org/10.1016/j.visres.2018.03.003

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

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*. Retrieved from https://doi.org/10.1080/09541440340000439

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology. General*, 115(2), 107–117. Retrieved from https://doi.org/10.1037//0096-3445.115.2.107

Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *Quarterly Journal of Experimental Psychology* , 69(1), 1–10. Retrieved from https://doi.org/10.1080/17470218.2015.1017513

Dwyer, D. M., & Vladeanu, M. (2009). Perceptual learning in face processing: comparison facilitates face recognition. *Quarterly Journal of Experimental Psychology* , 62(10), 2055–2067. Retrieved from https://doi.org/10.1080/17470210802661736

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology*, 11(2), 219–224. Retrieved from https://doi.org/10.1016/s0959-4388(00)00200-2

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews. Neuroscience*, 15(8), 536–548. Retrieved from https://doi.org/10.1038/nrn3747

Hill, M. Q., Parde, C. J., Castillo, C. D., Ivette Colón, Y., Ranjan, R., Chen, J.-C., … O'Toole, A. J. (2019). Deep convolutional neural networks in the face of caricature. *Nature Machine Intelligence*,

1(11), 522–529. Retrieved 18 May 2020 from https://doi.org/10.1038/s42256-019-0111-7

Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science*, 319(5862), 435. Retrieved from https://doi.org/10.1126/science.1149656

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*. Retrieved from https://doi.org/10.1016/j.cognition.2011.08.001

Kemp, R., Towell, N., & Pike, G. (1997). When Seeing should not be Believing: Photographs, Credit Cards and Fraud. *Applied Cognitive Psychology*. Retrieved from https://doi.org/3.0.co;2-o">10.1002/(sici)1099-0720(199706)11:3<211::aid-acp430>3.0.co;2-o

Klare, B. F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., ... & Jain, A. K. (2015). Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *Proceedings of the IEEE conference on computer vision and pattern recognition,* 1931-1939.

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*. Retrieved from https://doi.org/10.1167/15.4.1

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, 172, 46–58. Retrieved from https://doi.org/10.1016/j.cognition.2017.12.005

Kramer, R. S. S., Young, A. W., Day, M. G., & Burton, A. M. (2017). Robust social categorization emerges from learning the identities of very few faces. *Psychological Review*, 124(2), 115–129. Retrieved from https://doi.org/10.1037/rev0000048

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology*. Retrieved from https://doi.org/3.0.co;2-7">10.1002/1099-0720(200101/02)15:1<101::aid-acp697>3.0.co;2-7

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*. Retrieved from https://doi.org/10.1038/82947

Leopold, D. A., Rhodes, G., Müller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to

faces. *Proceedings. Biological Sciences / The Royal Society*, 272(1566), 897–904. Retrieved from https://doi.org/10.1098/rspb.2004.3022

Lewis, M. (2004). Face-space-R: Towards a unified account of face recognition. *Visual Cognition*, *11*(1), 29-69. Retrieved from https://doi.org/10.1080/13506280344000194

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology. Human Learning and Memory*, 5(3), 212–228. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/528913

Menon, N., Kemp, R. I., & White, D. (2018). More than a sum of parts: robust face recognition by integrating variation. *Royal Society Open Science*. Retrieved from https://doi.org/10.1098/rsos.172381

Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of Experimental Psychology. Animal Behavior Processes*, 33(2), 124–138. Retrieved from https://doi.org/10.1037/0097-7403.33.2.124

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology. Human Perception and Performance*, 41(3), 577–581. Retrieved from https://doi.org/10.1037/xhp0000049

Noyes, E., & Jenkins, R. (2017). Camera-to-subject distance affects face configuration and perceived identity. *Cognition*, 165, 97–104. Retrieved from https://doi.org/10.1016/j.cognition.2017.05.012

Noyes, E., & Jenkins, R. (2019). Deliberate disguise in face identification. *Journal of Experimental Psychology. Applied*, 25(2), 280–290. Retrieved from https://doi.org/10.1037/xap0000213

O'Toole, A. J., An, X., Dunlop, J., Natu, V., & Phillips, P.J. (2012). Comparing face recognition algorithms to humans on challenging tasks. *ACM Transactions on Applied Perception*. Retrieved from https://doi.org/10.1145/2355598.2355599

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face Space Representations in Deep Convolutional Neural Networks. *Trends in Cognitive Sciences*, 22(9), 794–

809. Retrieved from https://doi.org/10.1016/j.tics.2018.06.006

O'Toole, A. J., Phillips, P.J., Jiang, F., Ayyad, J., Penard, N., & Abdi, H. (2007). Face Recognition Algorithms Surpass Humans Matching Faces Over Changes in Illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Retrieved from https://doi.org/10.1109/tpami.2007.1107

Parde, C. J., Castillo, C., Hill, M. Q., Colon, Y. I., Sankaranarayanan, S., Chen, J. C., & O'Toole, A. J. (2017, May). Face and image representation in deep CNN features. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 673-680. IEEE.

Parde, C. J., Colón, Y. I., Hill, M. Q., Castillo, C. D., Dhar, P., & O'Toole, A. (2020). Single Unit Status in Deep Convolutional Neural Network Codes for Face Identification: Sparseness Redefined. *arXiv Preprint*. Retrieved from https://doi.org/arXiv:2002.06274.

Phillips, P. J., Hill, M. Q., Swindle, J. A., & O'Toole, A. J. (2015). Human and algorithm performance on the PaSC face Recognition Challenge. *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. Retrieved from https://doi.org/10.1109/btas.2015.7358765

Phillips, P. J., & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*. Retrieved from https://doi.org/10.1016/j.imavis.2013.12.002

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Retrieved from https://doi.org/10.1109/34.879790

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*. Retrieved from https://doi.org/10.1016/j.visres.2006.03.002

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology* , 70(5), 897–905. Retrieved from https://doi.org/10.1080/17470218.2015.1136656

Rossion, B. (2018, June). Humans Are Visual Experts at Unfamiliar Face Recognition. *Trends in*

cognitive sciences. Retrieved from https://doi.org/10.1016/j.tics.2018.03.002

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1701-1708. IEEE.

Sankaranarayanan, S., Alavi, A., Castillo, C. D., & Chellappa, R. (2016). Triplet probabilistic embedding for face verification and clustering. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. Retrieved from https://doi.org/10.1109/btas.2016.7791205

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 815-823. IEEE.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71–86. Retrieved from https://doi.org/10.1162/jocn.1991.3.1.71

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 43(2), 161–204. Retrieved from https://doi.org/10.1080/14640749108400966

Vapnik, V., Golowich, S. E., & Smola, A. (1997). Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. *Advances in Neural Information Processing Systems*, (9), 281–287. Retrieved from https://doi.org/10.1007/978-3-642-33311-8_5

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS One*, 9(8), e103510. Retrieved from https://doi.org/10.1371/journal.pone.0103510

Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014). Learning Face Representation from Scratch. Retrieved from https://doi.org/http://arxiv.org/abs/1411.7923

Young, A. W., & Burton, A. M. (2017). Recognizing Faces. *Current Directions in Psychological Science*. Retrieved from https://doi.org/10.1177/0963721416688114

Young, A. W., & Burton, A. M. (2018, June). What We See in Unfamiliar Faces: A Response to Rossion.

Trends in cognitive sciences. Retrieved from https://doi.org/10.1016/j.tics.2018.03.008

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, 16(6), 747–759. Retrieved from https://doi.org/10.1068/p160747

Zhao, W., Krishnaswamy, A., Chellappa, R., Swets, D. L., & Weng, J. (1998). Discriminant Analysis of Principal Components for Face Recognition. *Face Recognition*. Retrieved from https://doi.org/10.1007/978-3-642-72201-1_4