Vision Research 201 (2022) 108128

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Perceptual similarity modulates effects of learning from variability on face recognition

Tal Honig [b,1], Adva Shoham [a,1], Galit Yovel [a,b,*]

[a] School of Psychological Sciences, Tel Aviv University, Israel
[b] Sagol School of Neuroscience, Tel Aviv University, Israel

ARTICLE INFO

Keywords:
Face recognition
Face space
Variability
Familiar faces
Perception
Similarity

ABSTRACT

Face recognition is a challenging classification task that humans perform effortlessly for familiar faces. Recent studies have emphasized the importance of exposure to high variability appearances of the same identity to perform this task. However, these studies did not explicitly measure the perceptual similarity between the learned images and the images presented at test, which may account for the advantage of learning from high variability. Particularly, randomly selected test images are more likely to be perceptually similar to learned high variability images, and dissimilar to learned low variability images. Here we dissociated effects of learning from variability and study-test perceptual similarity, by collecting human similarity ratings for the study and test images. Using these measures, we independently manipulated the variability between the learning images and their perceptual similarity to the test images. Different groups of participants learned face identities from a low or high variability set of images. The learning phase was followed by a face matching test (Experiment 1) or a face recognition task (Experiment 2) that presented novel images of the learned identities that were perceptually dissimilar or similar to the learned images. Results of both experiments show that perceptual similarity between study and test, rather than image variability at learning per se, predicts face recognition. We conclude that learning from high variability improves face recognition for perceptually similar but not for perceptually dissimilar images. These findings may not be specific to faces and should be similarly evaluated for other domains.

Face identification is a computationally challenging classification task that requires successful discrimination of a homogeneous set of images to different identities and generalization across different appearances of the same identity (see Fig. 1). Although it is commonly argued that humans excel in this task (e.g., Rossion, 2018; Tanaka, 2001), recent studies have shown that this ability is superb for familiar faces but is prone to errors for unfamiliar faces (Jenkins, White, Van Montfort, & Burton, 2011; Ritchie et al., 2015; for reviews see Young & Burton, 2017, 2018). This gap between familiar and unfamiliar faces is primarily observed in identity matching tasks where face images are presented on the screen and participants are asked to determine if they belong to the same or different identities (Fig. 1). Performance on these tasks is significantly better for familiar than unfamiliar identities. For example, a study that presented UK and Australian celebrities to UK and Australian participants revealed that performance on an identity matching task for the same pairs of faces was about 90 % correct if they were familiar to the participants but only 70 % correct if they were not familiar to the participants (Ritchie et al., 2015). Another study presented 40 different images of 2 identities (20 images per identity) and asked participants to sort them according to their identity. Results showed that participants who were not familiar with the people in the images sorted them to an average of 7 different identities, whereas those who were familiar with them easily sorted them correctly to 2 identities. These findings indicate that participants failed to generalize across different appearances of unfamiliar identities (Jenkins et al., 2011). Based on these and other similar findings, Burton and colleagues suggested that learning the variability of one identity does not generalize to other identities because different people vary in different ways (Burton et al., 2016; Kramer, Young, & Burton, 2018) (see Fig. 1). They further claimed that face identification depends on learning how each identity varies across its different appearances (Burton et al., 2016; Burton, 2013; Dowsett, Sandford, & Burton, 2016; Young & Burton, 2017). This type of experience is available only for familiar but not for unfamiliar faces.

* Corresponding author at: School of Psychological Sciences, Tel Aviv University, Tel Aviv 69987, Israel.
E-mail addresses: gality@post.tau.ac.il, gality@tauex.tau.ac.il (G. Yovel).

[1] Equal contribution.

https://doi.org/10.1016/j.visres.2022.108128
Received 15 March 2022; Received in revised form 28 August 2022; Accepted 15 September 2022
Available online 19 October 2022

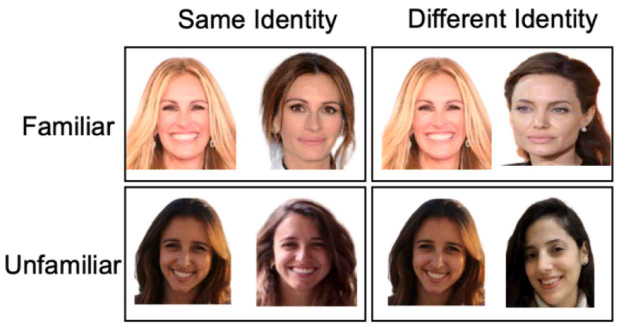

Fig. 1. Examples of pairs of same and different identity faces of familiar and unfamiliar identities. Performance on identity matching tasks is better for familiar than unfamiliar identities.

Based on this premise, several studies investigated the effect of learning identities from high or low variability images on face identification. In a study by Murphy, Ipser, Gaigg, & Cook (2015) image variability was manipulated across groups by either presenting a small number of appearances of each identity (i.e., 6 images per identity) and repeating them several times (low variability) or presenting a large number of appearances (i.e., 96 images per identity) and repeating each of them only once (high variability). Results show better recognition in an old/new memory test for novel images of the learned identities for the group who learned a larger number of appearances of each identity (i.e., high variability). The effect of learning faces from high vs low variability images was also compared directly (Ritchie & Burton, 2017). In this study, the low variability condition presented images that were cut from the same video of each identity, and therefore had similar characteristics and appearance (e.g., age, makeup, hairstyle). The high variability condition presented images that were taken from an internet search of each identity (e.g., taken at different times and places, from different cameras), creating a set of images that were perceptually dissimilar and more variable than the low variability set. A recognition task and an identity matching task with new images that were not presented at study showed better performance for identities that were learned from high than low variability (Ritchie & Burton, 2017, Exp 1a, 2).

Another study that compared the effect of learning from high vs low variability images defined variability based on the experimenter's similarity judgment, which was validated with a group of 10 participants. Results of a matching task showed that participants were more accurate and used a more liberal response criterion (i.e., more likely to make 'same' responses) when learned from high than low variability faces (Menon, White, & Kemp, 2015b). Similar findings were also reported in children that showed better identification of identities that were learned from 3 different videos than one video (Baker, Laurence, & Mondloch, 2017). Based on these and other findings Burton and colleagues have suggested that learning faces from highly variable appearances of the same identity is critical for face recognition (Burton, Jenkins, & Schweinberger, 2011; Young & Burton, 2017, 2018). It is noteworthy, that perceptual variability was never explicitly measured in these studies but was inferred from the media from which the images were taken (movie vs google images) or the number of images/movies that were presented in the learning phase.

Whereas results of the studies reported above suggest better performance for identities learned from high than low variability, a few studies also provided results that were inconsistent with it. In the study by Ritchie & Burton (2017, Exp1b) mentioned above, no differences were found between high and low learning conditions, when the test image was a novel image taken from the low variability set of images. Furthermore, in a study by Ritchie, Mireku, & Kramer (2020), participants were presented with either one face image or 4 face images per identity and were asked to match the face images to a real-life person. Results revealed similar performance in the two conditions, with no benefit for the 4 images compared to 1 image. Thus, learning from variability may not always lead to better performance in face identity tasks.

Here we propose that effects of image variability at learning on identity matching tasks may be modulated by the perceptual similarity between the images of the learned identities that are presented at study and test, which was not explicitly measured in previous studies. The role of perceptual similarity in face and object recognition has been demonstrated in studies that have shown a gradual drop in recognition for faces and objects as a function of the difference in view angle between the learned and tested images. For example, Tarr and Gauthier (1998) demonstrated that identifying an object from a novel angle in an old-new recognition task is easier following familiarity with objects shown from a similar view angle. Similarly, a study that tested recognition for objects across different views showed better matching performance for objects presented from the same view (Lawson & Humphreys, 1996). Another study that trained participants with faces from different viewing angles revealed that generalization to the unlearned faces decreased as the angle of rotation between study and test increased (Hill, Schyns, & Akamatsu, 1997; see also, Schwartz & Yovel, 2019). Based on these and other studies, Tarr and colleagues suggested that a view-invariant representation depends on learning faces/objects from multiple views with some interpolation to perceptually similar views of the learned identities (Bülthoff, Edelman, & Tarr, 1995; Tarr, 1995; Tarr & Gauthier, 1998).

Recent studies that examined the role of image variability in face identity have used ambient face images (Jenkins et al., 2011) rather than the well-controlled face images that vary in head-view, which were common in earlier studies (e.g., Jeffery, Rhodes, & Busey, 2006, O'Toole, Edelman, & Bülthoff, 1998). Whereas the ecological validity of ambient stimuli is a significant advantage, their perceptual similarity was not explicitly quantified. Nevertheless, the effect of image similarity between learning and test that were found for the controlled images are expected to apply also to the ambient face images. Thus, it is critical to provide a quantitative measure of the variability between the learned images as well as their perceptual similarity to the novel images presented at test.

To demonstrate the possible interaction between perceptual variability of face images at learning and their similarity to faces at test in a quantitative manner, we describe face images as dots in a multidimensional space, in which perceptually similar face images are in closer locations and perceptually different face images are in distant locations. Whereas cognitive face space models have primarily considered the representation of images of different identities in a face space (Valentine, 1991), the same principle can be applied to different images of the same identity, as typically done in computational models of face recognition (e.g., Abudarham et al., 2021; Hill et al., 2019; O'toole et al., 2018). Fig. 2 shows an illustration of the locations of different images of the same identity in a face space. Low variability images occupy a smaller area in face space. Thus, a randomly selected novel image of this identity is more likely to be perceptually different from the low variability images, resulting in lower performance in matching the identity of the novel image to each of the learned images (Fig. 2B). High variability images occupy a larger area in face space. Thus, a randomly selected novel image of the learned identity is more likely to be closer to at least one of the images, which will result in higher performance in a face matching task (Fig. 2C). We therefore presume that previous studies that reported better performance following high than low variability at learning, selected images that are more similar to high variability images and less similar to the low variability images. However, the perceptual similarity between learned and novel images can be manipulated in a way that will generate different outcomes. Fig. 2A shows a novel face image that is perceptually similar to low variability learned images and will be easy to match to the learned images. Fig. 2D shows a face image at test that is perceptually different from high-variability learned images and will be hard to match to the learned images even

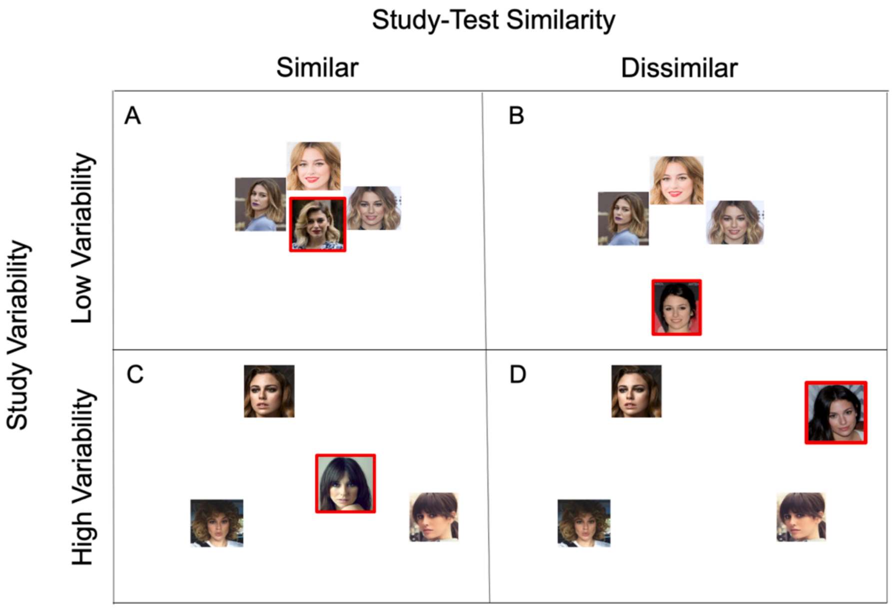

Fig. 2. An illustration of the locations of different images of the same identity in a face space. Distance between faces indicates their perceptual similarity. Novel faces presented at test are marked with a red frame. A. Low variability faces at study and a perceptually similar face at test. B. Low variability faces at study and a perceptually dissimilar face at test. C. High variability faces at study and a perceptually similar face at test. D. High variability faces at study and a perceptually dissimilar face at test.

though they were learned from high variability. Thus, the perceptual similarity between the learned and the novel images should be considered when assessing the effects of learning from low or high-variability images.

In the current study, we asked participants to rate the perceptual similarity between face images of the same identity. These similarity ratings enabled us to quantitatively measure both image variability at study and perceptual similarity between study and test images. This adds to previous studies in two ways. First, it provides a quantitative measure of perceptual similarity that was not explicitly measured in previous studies and is expected to play an important role in performance on face recognition tasks (O'Toole, Edelman, & Bülthoff, 1998; Tarr & Gauthier, 1998). Second, it enabled us to systematically allocate the face images to each of the four conditions displayed in Fig. 2 and evaluate the effects of image variability at learning, the similarity between the face images at study and test and the interaction between them, on face recognition. Accordingly, we manipulated both study variability and study-test similarity in a matching task and a recognition task. We predict better performance for images that are perceptually similar than perceptually dissimilar to the learned images above and beyond effects of image variability at learning.

1. Experiment 1

1.1. Methods

A pre-registration of the hypotheses, design and analysis of the study is available here: https://aspredicted.org/ZDG_VG3.

1.1.1. Participants

Eighty Psychology students (19 males, mean age = 23) participated in the pre-test identity rating task and 80 additional students participated in the main experiment, in exchange for course credit. The participants were randomly assigned to four experimental conditions (see design). This sample size was determined based on a factorial ANOVA with an effect size of $\eta^2p = 0.1$, $p < 0.05$ and power = 80 %. Two participants were excluded from the analysis because more than 25 % of their responses were longer than 10 s, resulting in 78 participants (6 males, mean age = 23 years). All the experiments were administered online using a code written in JavaScript that presented the images and collected responses. The experiment was approved by the ethics committee of Tel Aviv University and all participants gave their informed consent to participate in the study.

1.1.2. Stimuli

The stimuli included images of 10 celebrity identities from different European countries (5 females). All identities were unfamiliar to university students in Israel. For each identity, 15 photos were selected from Google images. They were chosen to represent a large perceptual variability of appearances of each identity. Identity decision ratings were collected to provide a quantitative measure of their perceptual similarity.

Pre-test - Perceptual similarity rating. To assign face images to the different conditions, we used data that was collected from 80 participants who were asked to decide whether each pair of images belong to the same identity or to different identities using a scale between 1 – definitely different people to 6 – definitely the same person. Each image pair was rated by 10 participants. Of a total of 5550 (2775 for each gender) image pairs, each participant rated 693 or 694 pairs. These ratings were used for the initial allocation of the stimuli to the different learning and test conditions (Fig. 3). The perceptual similarity of the selected images was then re-assessed by asking a different group of 80 participants (61 females, mean age: 23), 20 for each condition, to rate how perceptually similar are they on a scale between 1 and very different to 6 – very similar. The correlation between the identity ratings and the perceptual similarity ratings was 0.91, indicating that both methods can be used to estimate perceptual similarity between faces. To measure inter-subject reliability of the perceptual rating task, we computed the correlations between each participant's similarity ratings with the average similarity rating of the other participants in the same experimental group. The average reliabilities were r = 0.53 (range: 0.22–0.72) for *study high test dissimilar*, r = 0.65 (range: 0.39–0.76) for *study high test similar*, r = 0.83 (range: 0.76–0.88) for *study low test dissimilar* and r = 0.88 (range: 0.72–0.95) for *study low test similar*.

1.1.3. Design

The design of the experiment is presented in Fig. 3. The experiment included a study phase in which participants learned face identities based on a subset of different appearances. During the test phase, they were presented with pairs of images of the same or different identities. The image pairs at test were either both presented in the study phase (learned image pairs) or were pairs of an image that was presented at study paired with a novel image that was not presented at study (i.e., *unlearned pairs*). These unlearned pairs enabled us to directly assess the

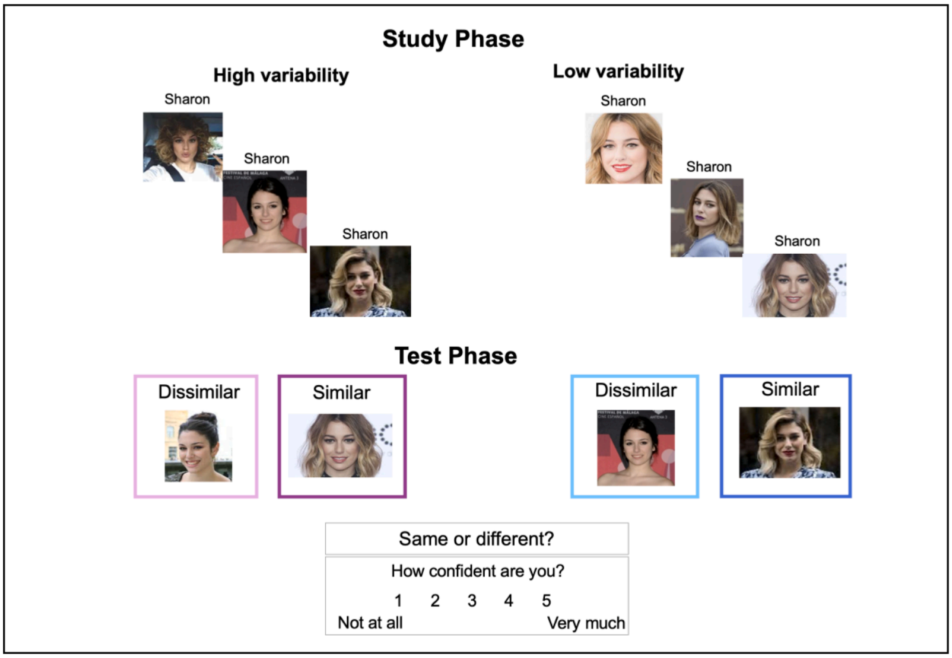

Fig. 3. A between-subjects experimental design was used to study effects of image variability at learning and study-test similarity. Left. A high variability study phase presents three perceptually different images of a learned identity with the same name label. Right. A low variability study phase presents three perceptually similar images of a learned identity with the same name label. The study phase was followed by a test phase that presented perceptually similar or dissimilar images of the learned identities. In Experiment 1 the test images were presented with each of the learned images in an identity matching task. In Experiment 2 they were presented with new identities in an old/new recognition task.

effect of perceptual similarity between the learned images and a novel image of the same identity on identity matching.

In a between-subjects design, we manipulated two factors: the perceptual variability among different images of the same identity during the study phase and the perceptual similarity between different images of the same identity in the study and test phase. Each condition had two levels: perceptual variability at study was high or low and perceptual similarity between study and test images was similar or dissimilar, which resulted in a 2x2 experimental design (Fig. 3).

For the low variability study phase, we selected the 3 most similar face images out of the 15 images of each identity, according to the pretest ratings. The high variability faces were the 3 most different images out of the 15 images of each identity. For each identity, we selected a perceptually similar and dissimilar test face. The similar face was the image that was most similar to all the three images (i.e., the minimal average distance from the 3 face images) that were selected to the study phase. The dissimilar face was the image that was the most different from all three images that were selected to the study phase (i.e., the maximal average distance from the 3 face images). We selected only 3 faces for the learning phase, to maximize the differences in image similarity at learning and test in the four different conditions. Including additional faces would decrease the differences between the experimental conditions and make it harder to dissociate between variability at learning and study-test similarity. Table 1 reports the average within-identity similarity rating among the study image pairs in the high and low variability conditions and the study-test similarity of each condition, which is the average similarity between the unlearned face pairs

Table 1
The average (standard deviation) perceptual similarity scores of the face images at study (top row) and the novel image of the learned identities at test (bottom row).

Study Variability	High variability (Low similarity)		Low variability (High similarity)	
	3.39 (0.83)		5.66 (0.31)	
Study-Test Similarity	Dissimilar	Similar	Dissimilar	Similar
	3.67 (0.89)	4.54 (0.73)	3.18 (0.9)	5.31 (0.45)

that are consisted of a learned image of a novel image of the learned identities. This enables us to assess how well participants can generalize between learned and novel images of learned identities. As we explained above (see Fig. 2), we expect that learning from high variability with a similar test image (Fig. 2C) will be better than learning from low variability with a dissimilar test image (Fig. 2B). We suggest that previous studies that overlooked effects of study-test perceptual similarity, only presented these two conditions, which show an advantage for high than low variability. The complementary conditions (Fig. 2A, 2D) will enable us to disentangle effects of learning from variability and study-test similarity.

The face images used for the different identity pairs were selected randomly from the same set of images that were used for the same identity pairs. The different identity learned pairs were images of different identities that were presented in the study phase. The different identity unlearned pairs comprised of one image from the study and another image of a different identity that was novel. The task included 10 identities that were presented in separate study-test sessions. The 5 female identities were presented in one session and the 5 male identities were presented in another session. There was a 24-hour interval between the two sessions and the order was randomized across participants.

1.1.4. Procedure

The experiment consisted of study and test phase as follow:

Study phase Participants were instructed to learn all the face images of the 5 identities that were presented during the study phase. Each identity included three different images that were shown sequentially. All three images of the same identity were presented in mini-blocks (one after the other) with the same name label above the face (see an example trial in Fig. 3). This sequence of 15 faces (5 identities × 3 images per identity) was repeated 3 times. The order of images within identity (within mini-block) and between identities was randomized across participants. Each image was presented for 1 s with a 1-second interval between images. Each identity mini-block included either high or low variability images. This was manipulated between participants.

Test phase A total of 120 face pairs were presented during the test phase. Half of the test face pairs were same identity faces and half were different identity faces. Half of the face pairs were face images that were presented during the study phase (learned face pairs) and half were face pairs in which one of the faces was presented in the study phase and the other was novel (unlearned face pairs). The same identity unlearned pairs were either similar or different (see Fig. 3). This was manipulated between participants.

The participants were asked to decide, for each pair, whether the two faces belong to the same identity or to different identities (by using the keyboard). Each pair was presented on the screen until the decision was made. After making their decision, the participants were asked to rate how confident they were in their decision on a scale of 1 (not at all) to 5 (very much). The time interval between the confidence ranking and the next stimulus was 1 s.

1.1.5. Data analysis

To assess participants' performance level we computed the area under the curve (AUC) of the ROC for learned and unlearned face pairs in the four different conditions. ROC was computed by combining the identity decision and its confidence rating by multiplying the confidence ratings after a 'different' identity decision by − 1, resulting in a confidence scale (−5)–(+5) (Pashler, 2002). Reaction time were measured during the test phase on each trial from the onset of the face pair till response.

Prior to the analysis, we excluded from the data trials in which the response time was longer than 10 s (0.02 % of the trials) or shorter than 200 ms (2 % of the trials). Two participants were excluded from the analysis as more than 25 % of their trials were removed due to the response time exclusion criteria.

Raw data can be found in this OSF link: https://osf.io/5epxb/?view_only = c75a96b7130b4e96b467d6586220706f.

1.2. Results & Discussion

Table 2 reports the proportion of Hit and FA rates, AUC and reaction times for the unlearned image pairs across the four different conditions.

Table 2
The average (standard deviation) Hit rate, FA rate, AUC and reaction times (RTs, in milliseconds) for the unlearned test images of the learned identities in the four experimental conditions.

	High		Low	
	Dissimilar	Similar	Dissimilar	Similar
Hit	0.47 (0.23)	0.6 (0.2)	0.31 (0.17)	0.91 (0.12)
FA	0.10(0.1)	0.15(0.12)	0.12(0.12)	0.03(0.03)
AUC	0.70(0.14)	0.78(0.12)	0.63(0.13)	0.97(0.03)
RTs	3654(1176)	3241(949)	3252(807)	2175(379)

To examine the effect of image variability and perceptual similarity on identity matching, we measured performance level on the identity matching task for unlearned face pairs in each of the four conditions. We performed a 2x2 between-subjects ANOVA with study variability (high, low) and test similarity (similar, dissimilar) as independent variables, and the AUC as a dependent variable. A main effect was found for study variability, $F(1,74) = 5.92$, $p = 0.01$, $\eta^2p = 0.07$, indicating better performance for identities that were presented at study in low ($M = 0.8$) than high ($M = 0.74$) variability images. A main effect of test similarity, $F(1,74) = 68.11$, $p < 0.001$, $\eta^2p = 0.47$, indicates better performance when the unlearned face pairs were similar to the learned identities than when they were dissimilar. In addition, we found a significant interaction between study variability and test similarity, $F(1,74) = 26.22$, $p < 0.001$, $\eta^2p = 0.26$ (see Fig. 4). Post hoc tests showed an advantage for learning faces from high variability than low variability only when the test was similar to the high variability faces (Fig. 4 dark pink bar: $M = 0.78$) and dissimilar to the low variability faces (Fig. 4 light blue bar – $M = 0.63$) $t(74) = 4.12$, $p < 0.001$, Cohen's $d = 1.31$, CI for mean difference [0.05,0.24]. As mentioned above, previous studies that selected the test image randomly were more likely to choose a similar image to high variability learned images and a dissimilar image to low variability learned images, which led to better performance for high than low variability conditions. Our findings therefore replicate this effect.

However, there was no significant difference between learning faces from high variability (Fig. 4: light pink – $M = 0.7$) than low variability (Fig. 4: light blue) when the test face was dissimilar in both conditions $t(74) = 1.9$, $p = 0.37$, Cohen's $d = 0.6$, CI for mean difference [−0.03 – 0.16]. Finally, performance was best when faces were learned from low variability and test faces were similar (Fig. 4: dark blue – $M = 0.97$) relative to all other conditions: study low test dissimilar, $t(74) = 9.58$, $p < 0.001$, Cohen's $d = 3.02$, CI of mean difference [0.25,0.48], study high test dissimilar, $t(74) = 7.56$, $p < 0.001$, Cohen's $d = 2.42$, CI for mean difference [0.18,0.37] and to study high test similar, $t(74) = 5.34$, $p < 0.001$, Cohen's $d = 1.71$, CI[0.1,0.29].

A similar analysis with proportion correct on same and different unlearned face pairs as a dependent measure revealed similar findings of an interaction between the variability at learning and the similarity of study and test $F(1,74) = 49.9$, $p < 0.001$, $\eta^2p = 0.4$, a main effect for test similarity $F(1,74) = 86.24$, $p < 0.001$, $\eta^2p = 0.54$ and for study variability $F(1,74) = 8.26$, $p = 0.005$, $\eta^2p = 0.1$.

Overall, these findings show that the perceptual similarity between faces presented at study and test determines performance on a face

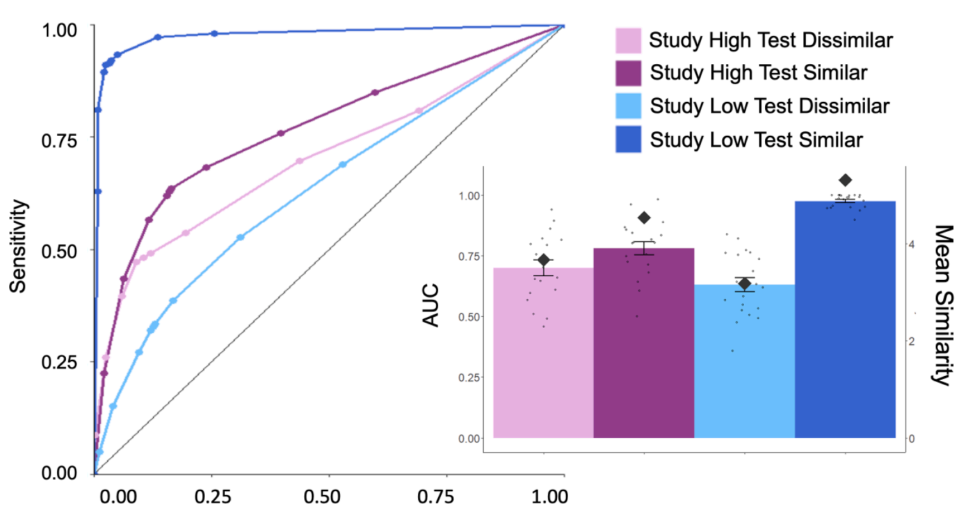

Fig. 4. ROC curves for each condition. The diagonal line indicates chance performance. The bar plot in the bottom right corner shows the mean and individual AUCs across subjects. The black diamonds indicate the mean perceptual similarity of the face pairs presented at test in each condition. Error bars show the standard error of the mean.

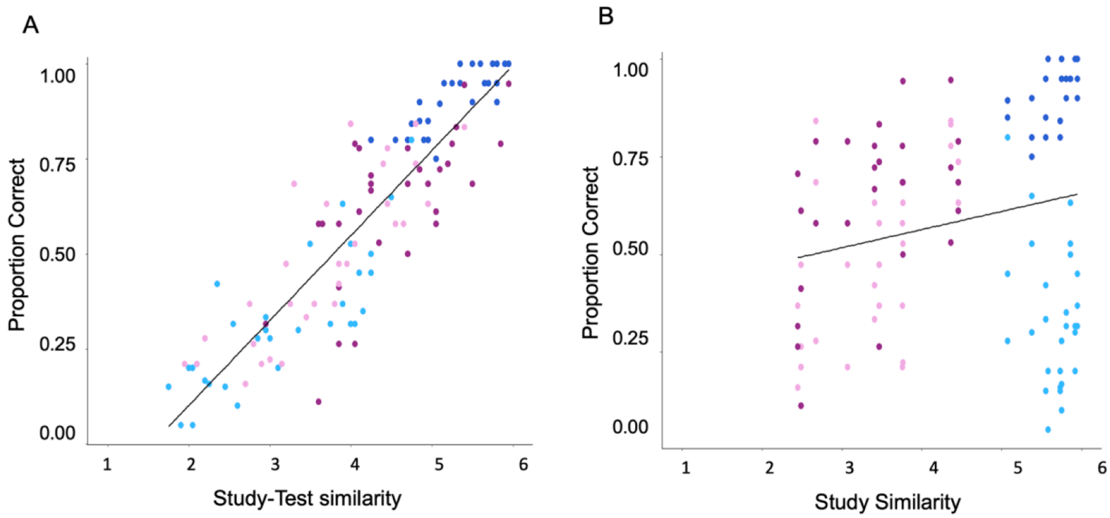

Fig. 5. A scatterplot of the proportion correct by mean similarity scores. Each dot represents a pair of faces from the unlearned same identity condition. The color indicates the experimental condition (see Fig. 4). A strong correlation was found between study-test similarity and performance on the face matching task. B. A relatively low correlation was found between proportion correct and mean similarity scores of the study phase faces (variability at learning).

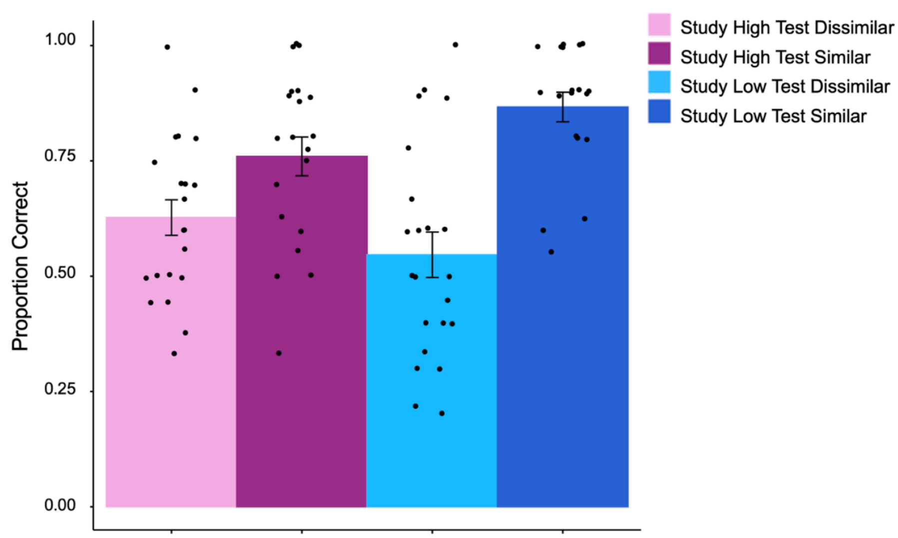

Fig. 6. Proportion correct on face recognition task as a function of image similarity between study and test and image variability at study. Results show that study-test similarity modulates effects of variability at learning. Error bars show the standard error of the mean.

identity task. To further explore the relationship between the perceptual similarity between the images at study and test and performance on an identity matching task, we ran an additional, unplanned analysis in which we examined the correlations between the perceptual similarity score that were collected in the pre-test and the performance on the identity matching task of the unlearned same identity face pairs across all four conditions. Fig. 5A shows a strong relationship between perceptual similarity and performance on a face matching task ($r(118) = 0.89$, $p < 0.001$, CI(95 %) = 0.85 – 0.92) indicating better performance when the study and test image pairs are similar (darker points) than when they are dissimilar (brighter points) for both the high and low similarity conditions.

We next display the relationship between the variability of the images at study and performance in the identity matching task for each face pair. Fig. 5B shows a weaker relationship between image variability at study and performance on the identity matching task ($r(118) = 0.2$, $p = 0.02$, CI(95 %) = 0.021–0.37).

Finally, we also examined the effect of learning from variability on performance of the learned pairs (pairs of faces that were both presented in the study phase). We averaged the performance for the learned pairs for the high variability and low variability conditions. A one-way ANOVA [study low ($M = 0.98$, $SD = 0.02$) and study high ($M = 0.81$, $SD = 0.14$)] as an independent variable and the AUC as a dependent variable revealed a main effect for learning variability $F(1,76) = 60.14$, $p < 0.001$, $\eta^2p = 0.44$, indicating a significantly better performance for face pairs learned from low than high variability images. A similar analysis with proportion correct as a dependent measure revealed similar findings of a main effect of learning variability $F(1,76) = 75.3$, $p < 0.001$, $\eta^2p = 0.49$. These findings indicate that learning faces from high variability is challenging also for the learned face images.

We performed a similar 2x2 ANOVA on reaction times (RTs) with study variability (high, low) and test similarity (similar, dissimilar) as independent variables. A main effect was found for study variability, $F(1,72) = 13.45$, $p < 0.01$, $\eta^2p = 0.16$, indicating faster RT for low than high variability images. A main effect for test similarity was found as well, $F(1,72) = 13.84$, $p < 0.001$, $\eta^2p = 0.16$, indicating better performance when the face pairs were similar to the learned identities than when they were dissimilar. The interaction between these two variables was not significant.

Taken together, results show that performance on a face identification task is better when the test image is similar to the learned images than when they are dissimilar. Whereas our findings seem to differ from previous studies that reported better performance for learning from high than low variability images, the advantage of learning from high variability is found also in our study when high variability images are tested with perceptual similar images compared to low variability images with perceptually dissimilar images. We propose that previous studies that selected images at test randomly included only these two conditions. Our study includes the complementary conditions that enable to dissociate between variability at learning and similarity with test. Our results show that study-test similarity modulates effects of image variability at learning.

The task that we used in Experiment 1 examines the effects of learning from variability and study-test similarity with an identity matching task of learned and novel images. This task was not used in previous studies that examined learning from variability. Whereas this task does measure generalization, it remains to be seen whether the same findings will be also found in a more standard old-new recognition task. In Experiment 2, we therefore ran the same learning procedure followed by an old-new face recognition task.

2. Experiment 2

A pre-registration of the hypotheses, design and analysis of the study is available here: https://aspredicted.org/blind.php?x=RCZ_BMF.

2.1. Methods

2.1.1. Participants

A total of 135 US participants were recruited in the Prolific platform. We excluded participants that 25 % or more of their trials were removed. This left 82 participants (females = 37, mean age = 30) that were included in the analysis, similar to the sample used in Experiment 1. All the experiments were administered online using a code written in JavaScript that presented the images and collected responses. The experiment was approved by the ethics committee of Tel Aviv University and all participants gave their informed consent to participate in the study.

2.1.2. Stimuli

The faces at study were the same faces that were presented in Study 1. The faces at test were the same unlearned images of the same identity faces that were presented at test in Study 1 and were selected based on their high or low similarity to the learned images (Fig. 3). Each of the images of the learned identities was paired with an image of a new identity that was randomly selected from a set of 10 new identities.

2.1.3. Design

The design of the Experiment is similar to the design that was used in Study 1 and presented in Fig. 3, with a few differences. Experiment 1 presented the male and female identities in separate study-test sessions, and in Experiment 2 participants were presented with one session that included all 10 male and female identities. Exactly like in Experiment 1, the study phase included 3 images of each of 10 identities that repeated 3 times. In the test phase new (unlearned) images of the 10 learned identities were presented paired with images of new identities. The new identities were randomly paired across participants with a same gender learned identity image.

Similar to Experiment 1, in a between-subjects design, we manipulated two factors: the perceptual variability among different images of the same identity during the study phase and the perceptual similarity between different images of the same identity in the study and test phases. Each condition had two levels: perceptual variability at study was high or low and perceptual similarity between study and test images was similar or dissimilar, which resulted in a 2x2 experimental design (Fig. 3).

2.1.4. Procedure

Study phase The study phase was similar to the study phase in Experiment 1, except that it included one session that presented all 10 identities.

Test phase The test phase presented 10 face pairs. Each pair included a new (unlearned) image of one of the learned identities and an image of a new identity. The participants were asked to determine by pressing a right or left arrow key on the keyboard whether the face on the left or the right was an identity that was presented in the study phase. The test images were presented until response. After making their decision, the participants were asked to rate how confident they were in their decision on a scale of 1 (not at all) to 5 (very much). The time interval between the confidence ranking and the next stimulus was 1 s. After completing all trials, the participants were asked to rate for each of the 20 identities presented at test if they were familiar with this person before the experiment or not.

2.1.5. Data analysis

Accuracy was measured by computing the proportion of trials in which participants selected the correct old image. Reaction times were measured from the onset of the stimuli until response.

Prior to the analysis, we excluded trials in which the response time was longer than 10 s (2 % of the trials) or shorter than 200 ms (0 % of the trials). We also excluded trials in which the participants were familiar with one of the identities. The participants that 25 % or more of their trials were excluded, they were excluded from the analysis completely.

Raw data can be found in this OSF link:
https://osf.io/5epxb/?view_only = c75a96b7130b4e96b467d65862 20706f.

2.2. Results & Discussion

Accuracy Fig. 6 shows the proportion correct recognition of learned identities from novel images across the four different conditions. The results are very similar to results of Experiment 1 (Fig. 3). The effect of image variability at learning was qualified by the similarity between images of the same identity at study and test. Performance was better when the test images were similar than dissimilar to the learned images both when faces were learned from low variability and high variability (see Table 2). A 2x2 between-subjects ANOVA with study variability (high, low) and test similarity (similar, dissimilar) as independent variables, and proportion correct as a dependent variable revealed a main effect for test similarity $F(1,78) = 41.48$, $p < 0.001$, $\eta^2p = 0.35$ as well as a significant interaction $F(1,78) = 9.2$, $p < 0.005$, $\eta^2p = 0.10$. The effect of variability at learning was not significant $F(1,86) = 0.006$, $p = 0.94$, $\eta^2p = 0.0$. Whereas these findings seem in consistent with previous reports, we believe that previous studies that found an advantage for high variability at learning but did not measure perceptual similarity between study and test images, were more likely to select dissimilar images in the low variability condition (Fig. 2B) and similar images in the high variability condition (Fig. 2C). When these two conditions are compared in our data, we also reveal that recognition for faces that were learned from high variability and were tested on similar images was significantly better than faces that were learned from low variability and tested with perceptually dissimilar images ($t(39) = 4.55$, $p < 0.001$, Cohen's $d = 1.31$, CI (95 %) = 0.62–1.99).

Similar to Experiment 1, we also examined the correlations between image similarity at test and proportion correct across images and found very high correlations (Fig. 7A), similar to Experiment 1 (see Fig. 5A). The correlation between proportion correct and test similarity was very high ($r(39) = 0.68$, $p < 0.001$, CI(0.48–0.82)) (see Fig. 7B), whereas no correlation was found with study variability ($r(39) = 0.04$, $p = 0.79$, CI (−0.27–0.35)), in line with results of Experiment 1 (Fig. 5B). This findings indicate that image variability alone does not determine face recognition, when study-test similarity is also considered.

Reaction time A 2x2 ANOVA on reaction time of correct responses

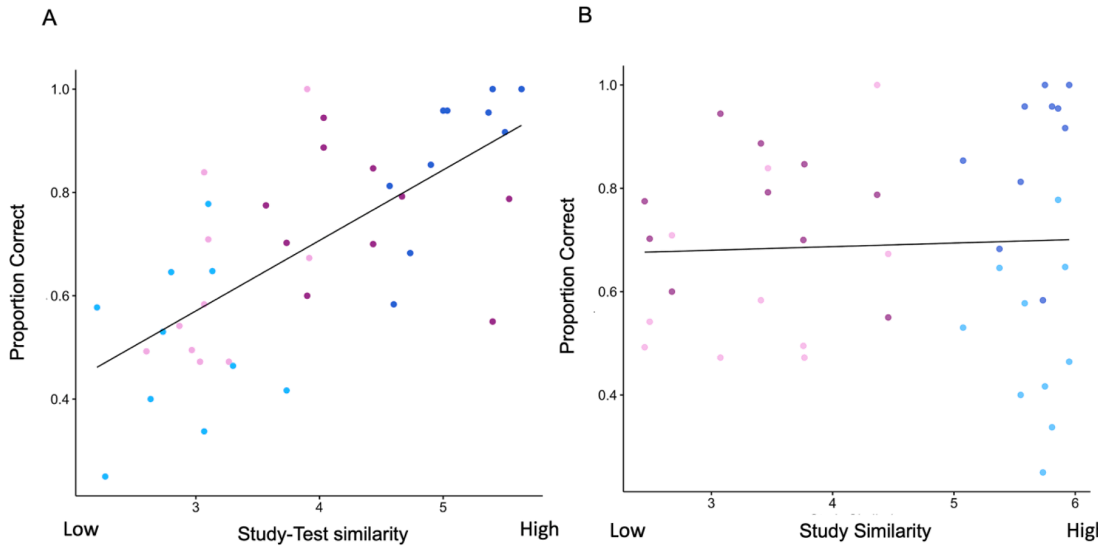

Fig. 7. Scatterplots of proportion correct in the face recognition task by mean similarity scores. Each dot represents an identity. The color indicates the experimental condition (see Fig. 6). A. a strong correlation was found between study-test similarity and performance on the face recognition task. B. No correlation was found between proportion correct on the face recognition task and mean similarity scores of the study phase face images, indicating that variability at learning does not predict face recognition.

revealed similar findings of a main effect of study-test similarity $F(1,78) = 5.12$, $p < 0.05$, $\eta^2p = 0.06$ and a significant interaction $F(1,78) = 8.29$, $p < 0.01$, $\eta^2p = 0.10$. The effect of variability at learning was not significant $F(1,78) = 0.16$, $p = 0.68$, $\eta^2p = 0.002$. Table 3 reports average and standard deviation of RTs of the 4 conditions.

Overall, these findings replicate results of Experiment 1 in a face recognition task, which is similar to previous studies that examined effects of image variability on face recognition (e.g., Ritchie & Burton, 2017, Baker et al., 2017). These findings are consistent with our suggestion that study-test similarity modifies effects of learning from variability. These findings highlight the importance of quantitatively measure and independently manipulate these two factors and consider the critical contribution of perceptual similarity between the learned images and the images at test on face recognition.

3. General discussion

The importance of learning from high variability images has been recently emphasized in many studies of face recognition (Burton, 2013; Jenkins et al., 2011; Menon, White, & Kemp, 2015a; Ritchie & Burton, 2017). Indeed, high variability images are more likely to be perceptually similar to a larger set of novel images of learned identities relative to low variability images (Fig. 2). Thus, the idea that learning from high variability is beneficial to face identity is in line with results of better recognition for faces that are perceptually similar to the learned images (Bülthoff, Edelman, & Tarr, 1995; Tarr, 1995; Tarr & Gauthier, 1998, Schwartz & Yovel, 2019). Nevertheless, previous studies did not explicitly assess the possible interaction between image variability of the learned images and their perceptual similarity to the images at test. To test the effects of both factors, we measured the variability of the learned images and the perceptual similarity between the study and test images. Our findings show an interaction between the effects of perceptual similarity between the learned and test images and the variability of the learned images. Performance was better following learning from high than low variability only when the test images were similar to the learned high-variability images and dissimilar to the learned low-variability images. Thus, learning from high variability is advantageous because perceptually different images of the same identity are more likely to be similar to the learning set. When this condition is not met, high variability at learning may not necessarily lead to better performance than low variability, but the reverse effect may occur. Indeed, performance was best when the image at test was perceptually similar to low variability learned images (Figs. 4 & 6). These findings indicate that perceptual similarity between the study and test images should be considered when effects of image variability at learning are investigated.

To study the mechanisms of face identification, many recent studies have used unconstrained images of faces, which greatly vary in illumination, color, pose and expression. Whereas these images enable to study face identification in a more ecological manner, it is not possible to systematically vary their perceptual similarity as typically done in studies that present faces that vary in head pose or illumination in a controlled manner. This limitation can be addressed by measuring the perceptual similarity among the ambient images by an independent group of participants. Based on these similarity ratings, face images can be assigned to high and low variability conditions and their similarity to unlearned images can be quantitatively assessed. Previous studies that did not measure the perceptual similarity between faces, selected the face images to be presented at test in an unsystematic manner (Andrews, Jenkins, Cursiter, & Burton, 2015; Baker et al., 2017; Matthews & Mondloch, 2018; Menon, Kemp, & White, 2018; Ritchie & Burton, 2017). As explained above, novel images are more likely to be similar to the learned identities in the high variability than the low variability condition (See Fig. 2) and lead to better performance for high than low variability face images as we also found in the current study (the dark pink and light blue bars in Figs. 4 & 6). By quantitatively measuring the perceptual similarity among the face images, we were able to include the complementary conditions showing how image similarity may modulate effects of learning from variability.

Examination of the different studies that manipulated image

Table 3
Average (standard deviation) proportion correct and reaction times (RTs, in milliseconds) for the test images of the learned identities in the four experimental conditions.

	High		Low	
	Dissimilar	Similar	Dissimilar	Similar
Proportion Correct	0.62 (0.18)	0.76 (0.19)	0.51 (0.20)	0.88 (0.14)
Reaction Time	3288 (1043)	3457 (1448)	4187 (1469)	2781(918)

variability at study and examined identity matching at test reveal inconsistent findings with some studies showing better performance for high than low variability conditions (Baker et al., 2017; Matthews & Mondloch, 2018; Menon, Kemp, & White, 2018; Murphy et al., 2015; Ritchie & Burton, 2017, Exp 1a) and others that did not support this hypothesis (Ritchie & Burton, 2017, Exp 1b, Ritchie et al., 2020; Sandford & Ritchie, 2021). We suggest that perceptual similarity between the learned faces and the faces at test, which was not explicitly measured in these studies, may account for some of these inconsistent findings. For example, in the study by Murphy et al., (2015), participants learned identities from either 6 images per identity (low variability) or 96 images per identity (high variability). We propose that the better performance that was found in the high variability condition is due to the higher likelihood that the test picture is perceptually similar to one of the 96 images compared to the 6 images. The same applies also to the studies that used images from a Google search for the high variability condition and images taken from a single movie for low variability condition. Perceptual similarity between the learned and test faces may also account for studies that did not find an advantage for learning from high variability. For example, when images were compared to a real-life person (Ritchie et al., 2020), a larger number of images in a set did not improve identification relative to a single image probably because the four images were not similar enough to the real-life representation of the person. Manipulation of perceptual similarity between the images and the person seen in real life, may modify this effect.

Another reason that was recently proposed for this discrepancy between studies that examined effects of learning from variability is that effects of variability are only found when a memory component is involved in the task (Ritchie et al., 2021). This was based on reported findings that when images are presented simultaneously, high variability does not improve face identification. In our study we found the effect both in a memory and a matching task. Notably, our matching task did involve a learning condition and is therefore different from previous studies that examined face matching of several images presented simultaneously. We suggest that results of the face matching tasks should be re-examined with a design similar to the current study that considers effects of perceptual similarity with the target image/person.

The number of different images per identity that we presented in the learning phase was relatively low. Each identity was represented by 3 images that repeated 3 times during the learning phase. Previous studies that manipulated the number of images at study found better performance when a larger number of images of each identity was learned (e.g., Murphy et al., (2015)). However, image variability and number of images in these studies were not independently dissociated. A recent study presented similar results to the findings reported in the current experiment. Sandford & Ritchie (2021) manipulated the variability of both the learned images and test images as well as the number of images per identity (1 /2 / 3). Their design included both low and high variability images. The test faces were taken from low variability data set or from the high variability data set. Similar to our findings, their results showed no consistent benefit of exposure to high variability compared to low variability, which was explained by the similarity between the target and the array. However, a few differences between our and Sandford & Ritchie (2021) study should be noted. First, the way variability levels were manipulated was the same as in Ritchie & Burton (2017). Low variability images were taken from the same video whereas high variability images were taken from Google image search, with no explicit measure of perceptual similarity. Furthermore, their definition of low and high variability target was based on the dataset from which the target was taken rather than the similarity between the array and the test faces. Therefore, their high variability study, low variability target condition was in fact parallel to what we named high variability test dissimilar condition. Their design therefore did not include a critical condition included in our study in which the test image is perceptually similar to at least one of the faces in the high variability array. Taken together, we suggest that the different results that were reported in previous studies that manipulated image variability at learning may be accounted for by the similarity between the face pairs at test, which was never explicitly measured. Indeed, we found a strong relationship between perceptual similarity and identity matching across all face pairs (Fig. 5A, Fig. 7A).

Our findings are consistent with studies that showed that face and object recognition are view specific (Hill et al., 1997; O'Toole et al., 1998; Schwartz & Yovel, 2019; Tarr, 1995; Tarr & Gauthier, 1998). These studies indicate that our ability to generalize to novel images that are perceptually different from the learned images is quite limited. This conclusion is also consistent with the Direct fit model, which indicates that generalization to new examples is expected based on interpolation to similar examples rather than extrapolation to distant examples (Hasson, Nastase, & Goldstein, 2020). Indeed, in real life, people do not change radically across temporally close encounters and therefore in most cases interpolation among perceptually similar appearances enables intact social interactions. When people do radically change across encounters, either due to a large gap between encounters (e.g., a class reunion) or radical changes in appearance (e.g., removal of facial hair), supervision is typically available as part of a normal social interaction among people, either explicitly by indicating the identity of the person or by using other identity cues including voice or the content of a conversion. Thus, familiar face recognition is enabled by generalization to images that are perceptually similar to previously learned appearances together with supervision to perceptually different appearances (Yovel & Abudarham, 2021).

The role of exposure to different appearances in face identification also raises the question of the definition of familiarity. Because face learning is view specific, perceptual experience is undoubtedly critical to reach the superb recognition that humans show for familiar faces. Nevertheless, even recognition of the people who are most close to us is limited to the appearances that we had experience with. For example, most of us will not recognize a picture of our colleagues and even of our parents or grandparents when they were 4 years old (as long as we have never seen their childhood pictures). This anecdotal evidence is consistent with results of the Before they were famous (BTWF) test, where identification of famous people from images at childhood is very poor (Russell, Duchaine, & Nakayama, 2009). This indicates that even extensive exposure to highly variable appearances of familiar identities is limited to images that are perceptually similar to the appearances we had experience with. The question of whether this effect is generalized to personally familiar faces with which we have much more experience should be examined in future studies.

Finally, the benefit of variability at learning to generalization at test is not unique to faces. In a recent review on the way variability shapes learning and generalization, Raviv and colleagues (2022) indicated that whereas learning from high variability is challenging during training, it does lead to better generalization in various tasks including motor learning, language, categorization, reasoning, and problem solving. Our findings are consistent with this claim. Performance for high variability learned images was overall lower but better from low variability learning, when dissimilar images were presented at test, therefore improving generalization.

In sum, the current experiment re-evaluates the role of image variability at learning for face recognition. We conclude that perceptual similarity between face images should be considered to evaluate the role of learning images from high variability. We highlight the importance of collecting quantitative measures of perceptual similarity when assessing the contribution of experience in general, and image variability and image similarity, in particular, to performance on classification and recognition tasks. This is critical particularly when uncontrolled face images are used as rightly done in many recent studies. More generally, effects of image variability at learning and image similarity at test discussed here are not limited to face recognition and are likely to apply to the learning and classification of any other perceptual stimuli (e.g., voices, objects) as well as other cognitive domains (Raviv et al., 2022).

CRediT authorship contribution statement

Tal Honig: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Adva Shoham:** Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – review & editing. **Galit Yovel:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Itai Linzen for creating the face database, Adi Fisher for collecting the data for the pre-test similarity rating study and Naphtali Abudarham for comments and discussion on this manuscript. This work was funded by the Israeli Science Foundation grant # 971/21 to GY.

References

Abudarham, N., Grosbard, I., & Yovel, G. (2021). Face recognition depends on specialized mechanisms tuned to view-invariant facial features: Insights from deep neural networks optimized for face or object recognition. *Cognitive Science, 45*(9), Article e13031.

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. https://doi.org/10.1080/17470218.2014.1003949

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition, 161*. https://doi.org/10.1016/j.cognition.2016.12.012

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex, 5*(3), 247–260. https://doi.org/10.1093/cercor/5.3.247

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology, 102*(4), 943–958. https://doi.org/10.1111/j.2044-8295.2011.02039.x

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. https://doi.org/10.1111/cogs.12231

Burton, M. A. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology, 66*(8), 1467–1485. https://doi.org/10.1080/17470218.2013.800125

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. https://doi.org/10.1080/17470218.2015.1017513

Hasson, U., Nastase, S. A., & Goldstein, A. (2020). Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron, 105*(3), 416–434. https://doi.org/10.1016/j.neuron.2019.12.002

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. In *Cognition* (Vol. 62).

Hill, M. Q., Parde, C. J., Castillo, C. D., Colon, Y. I., 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.

Jeffery, L., Rhodes, G., & Busey, T. (2006). View-specific coding of face shape. *Psychological Science, 17*(6), 501–505.

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313–323. https://doi.org/10.1016/j.cognition.2011.08.001

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

Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance, 22*(2), 395.

Matthews, C. M., & Mondloch, C. J. (2018). Improving identity matching of newly encountered faces: Effects of multi-image training. *Journal of Applied Research in Memory and Cognition, 7*(2), 280–290. https://doi.org/10.1016/j.jarmac.2017.10.005

Menon, N., Kemp, R. I., & White, D. (2018). *More than a sum of parts : robust face recognition by integrating variation Subject Category : Subject Areas : Author for correspondence :* 0–6.

Menon, N., White, D., & Kemp, R. I. (2015a). Identity-level representations affect unfamiliar face matching performance in sequential but not simultaneous tasks. *Quarterly Journal of Experimental Psychology, 68*(9), 1777–1793. https://doi.org/10.1080/17470218.2014.990468

Menon, N., White, D., & Kemp, R. I. (2015b). Variation in photos of the same face drives improvements in identity verification. *Perception, 44*(11), 1332–1341. https://doi.org/10.1177/0301006615599902

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). *Exemplar variance supports robust learning of facial identity*. https://doi.org/10.1037/xhp0000049.

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research, 38*(15–16), 2351–2363. https://doi.org/10.1016/S0042-6989(98)00042-X

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.

Pashler, H. (2002). *Stevens' handbook of experimental psychology, methodology in experimental psychology*. John Wiley & Sons.

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*.

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

Ritchie, K. L., Kramer, R. S., Mileva, M., Sandford, A., & Burton, A. M. (2021). Multiple-image arrays in face matching tasks with and without memory. *Cognition, 211*, Article 104632.

Ritchie, K. L., Mireku, M. O., & Kramer, R. S. S. (2020). Face averages and multiple images in a live matching task. *British Journal of Psychology, 111*(1), 92–102. https://doi.org/10.1111/bjop.12388

Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015). *Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox*. https://doi.org/10.1016/j.cognition.2015.05.002.

Rossion, B. (2018). Humans are visual experts at unfamiliar face recognition. *Trends in Cognitive Sciences, 22*(6), 471–472. https://doi.org/10.1016/j.tics.2018.03.002

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin and Review, 16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Sandford, A., & Ritchie, K. L. (2021). Unfamiliar face matching, within-person variability, and multiple-image arrays. *Visual Cognition, 29*(3), 143–157. https://doi.org/10.1080/13506285.2021.1883170

Schwartz, L., & Yovel, G. (2019). Independent contribution of perceptual experience and social cognition to face recognition. *Cognition, 183*, 131–138. https://doi.org/10.1016/j.cognition.2018.11.003

Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130*(3), 534.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2*.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition, 67*(1–2), 73–110. https://doi.org/10.1016/S0010-0277(98)00023-7

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 43*(2), 161–204.

Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Current Directions in Psychological Science, 26*(3), 212–217. https://doi.org/10.1177/0963721416688114

Young, A. W., & Burton, A. M. (2018). February 1. Are we face experts? *Trends in Cognitive Sciences, 22*, 100–110. https://doi.org/10.1016/j.tics.2017.11.007

Yovel, G., & Abudarham, N. (2021). From concepts to percepts in human and machine face recognition: A reply to Blauch, Behrmann & Plaut. *Cognition, 208*(2020). https://doi.org/10.1016/j.cognition.2020.104424