

The Proficiency for Distinguishing Faces is Independent of the Proficiency for Remembering Them

Catrina M. Hacker¹ & Irving Biederman^{1,2}

1. Program in Neuroscience, University of Southern California, Los Angeles, CA

2. Department of Psychology, University of Southern California, Los Angeles, CA

Corresponding author: bieder@usc.edu

Abstract

Tests of face recognition implicitly assume it to be an undifferentiated ability. However, several possible independent components could comprise face recognition proficiency, such as those for the perceptual discrimination of faces, face memory, and the ability to generalize across viewpoints. We assessed two possible components of face recognition ability: the proficiency for the perceptual discrimination of faces using a quantitative measure of face similarity and the proficiency for remembering those faces using a minimal delayed match-to-sample task. We document a strong cost of delay on error rates and mean correct reaction times in face matching. We further demonstrate that the slope of performance over increased perceptual difficulty is parallel for each of three levels of delay. These results provide evidence, by additive factors logic, that face memory and face perception are independent processes within subjects. Finally, we document an independence of the proficiencies for face memory and face perception across subjects.

Abstract word count: 151

Keywords: face recognition, face memory, face perception, prosopagnosia, delayed match-to-sample

Introduction

Face recognition is typically assumed to be an undifferentiated ability, yet a century of psychological research has firmly established that complex recognition tests, such as most popular tests of face recognition (Duchaine & Nakayama, 2006; Shah et al., 2015; Benton et al., 1994; Tanaka & Farah, 1993), engage separate capacities for perception, memory, and invariance to image transformations. The possible independence of these components has not been systematically studied in neurotypical subjects, but could have important implications for how face recognition proficiency is tested and for understanding its neural correlates. Here we investigate two possible components of face recognition proficiency—that for the perceptual discrimination of faces and that for face memory—using a minimal delayed match-to-sample task with computer-generated faces.

The developmental time courses of face perception and face memory proficiencies are distinct. In a study of 118 participants aged 5 years to adult, Weigelt et al. (2014) observed that memory specific to faces developed later than other forms of memory. On the other hand, face perception abilities developed earlier and at the same rate for faces as other object categories. The fact that the two proficiencies develop with different time courses could indicate that the two processes are independent in adulthood.

A behavioral dissociation between face memory and perception has been observed in individuals who have deficits in face processing, both in acquired and congenital prosopagnosia. Those with acquired prosopagnosia experience deficits in face perception and memory following lesions to face-selective cortex in late stages of the ventral visual stream. On occasion, these lesions are highly specific and patients develop *prosopamnesia*. Prosopamnesics show normal face perception, preserved memory for faces learned before acquiring the lesion and some show no memory deficits for other types of stimuli, but all have lost the ability to learn new faces (Tippett et al., 2000; Williams et al., 2007; Lopera & Ardila, 1992). These cases provide clear evidence for a dissociation between face perception and face memory on a biological and behavioral level in a population with brain lesions. Furthermore, they suggest that face memory is functionally distinct from other types of memory.

More recently, evidence for a dissociation between face memory and perception in congenital prosopagnosics has been documented (Dalrymple et al., 2014; Stollhoff et al., 2011).

Congenital prosopagnosics (CPros) report difficulty recognizing faces but, unlike acquired prosopagnosics, have no readily detectable underlying neurological damage. Dalrymple et al. (2014) tested a group of CPros with two face memory tests, the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006) and an old/new recognition task, and a face perception task, the Cambridge Face Perception Test (CFPT). They reported that while all subjects showed a deficit in face memory, only half of adults also showed a deficit in face perception. Cases of such a dissociation in both acquired and congenital prosopagnosics strongly suggests that a similar dissociation could be observed in neurotypical individuals. To our knowledge, no previous study has documented the independence of these two proficiencies in neurotypical subjects.

The Cambridge Face Memory Test (CFMT) is currently the most widely used instrument for the assessment of face memory. Rather than providing a pure measure of face memory, the CFMT conflates the three proficiencies of face memory, perception and invariance to viewpoint. The current study was designed to isolate task variations that affect face memory from variations affecting face perception in order to more rigorously assess whether the two proficiencies reflect independent processing stages. We used a quantitative measure of perceptual similarity between matching and foil faces in a 2AFC delayed match-to-sample task that allowed us to separate a purely perceptual component from the memory component. By assessing whether there was an interaction between the effects on RTs of increasing perceptual difficulty and the effects on RTs of an increase in the retention interval, we could apply additive factors logic to test whether these two experimental factors are independent and thus reflect separate stages in information processing (Sternberg, 1969; 2011). Two processes can be considered independent if the effect of each level of a given variable is equal across levels of the second. The application of this logic requires the ability to separate different levels of the two experimental components of interest—in this case face memory and face perception.

Many tests of face memory use a same-different task in which subjects are given a certain amount of time to learn a set of faces. After a delay, they are shown a new set of faces and asked to identify which are the faces they have previously learned (e.g., Duchaine & Nakayama, 2006; Tanaka & Farah, 1993). These tests reliably detect deficits in face processing

but do not allow for a scaling of perceptual difficulty given that there is more than one face to be held in memory. Without perceptual scaling of the similarity of the faces that are to be discriminated, this design does not readily allow quantitative assessment of the perceptual challenge on a given trial.

The match-to-sample paradigm avoids several major shortcomings of the same-different task. The same-different task forces the adoption of an arbitrary criterion as to whether a given stimulus is *identical* to the representation held in memory. The match-to-sample task allows the adoption of a relative criterion wherein the subject must decide which of the test faces *best* matches the original. With highly similar faces to be distinguished, as in the present study, a same different task can yield near chance—or even below chance—levels of performance as highly similar, but different faces, are judged to be the same. In a 2-alternative-forced-choice (2AFC) paradigm the delayed match-to-sample test also allows for the measurement of the perceptual difficulty of a given trial by measuring the similarity between the two test images. In the present study we employed the Gabor-jet model of similarity scaling (Lades et al., 1993; Yue et al., 2012; Margalit et al., 2016) to measure the perceptual difficulty of each trial. Such scaling has been shown to correlate extremely well—with *rs* in the mid .90s—with psychophysical performance in matching faces. These measurements allow us to systematically separate the effects of memory interval from variations in perceptual similarity within the same trial. Following their separation, we can assess the independence of face perception and face memory as reflected in both group and individual performance.

Methods

Subjects

51 undergraduates at the University of Southern California (average age 20.2 ± 1.5 years, 7 left-handed, 32 female) participated in the USC Face Memory Test (USC FMT) for course credit through USC's Psychology Department Subject Pool. Of the 51 subjects who completed the task, 49 subjects also completed the online USC Face Perception Test (USC FPT) (Yue et al., 2012) before completing the in-lab experiment. 7 of these 49 subjects were not included in subsequent analyses of USC FPT data because they had at least 20 trials with a reaction time greater than 5.0 sec, at least 20 trials with a reaction time less than 250 ms, or

pressed the same key more than 15 times in row, indicating that subjects were not paying attention to the task. All of the excluded subjects showed high error rates overall. Their data was included in the analysis of the USC FMT because the presence of an experimenter ensured diligent performance whereas the USC FPT was taken online at the subject's leisure. The work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All subjects gave informed consent in accordance with procedures approved by the University of Southern California's University Park Institutional Review Board.

Stimuli

20 unique computer-generated face identities were constructed using FaceGen Modeller (Singular Inversions, Toronto, Canada), a 3D face modelling program. The base face was a bald, Caucasian, gender-neutral face on a black background. Each identity was created by slight variations in the shape of the face parts and the relations between the parts. This resulted in 20 faces that could only be distinguished by metric and largely ineffable differences that are characteristic of natural face recognition (Figure 1a; Supplemental Figure 1). The use of computer-generated faces eliminated local distinguishing features such as beauty marks or moles present in photographs that might allow subjects to match faces based on local features, thus circumventing holistic face processing that is also characteristic of typical face recognition. Additionally, the uniformity of the faces allowed for the use of the Gabor Jet model for similarity scaling.

In addition to the 20 computer-generated faces used in the discrimination task, 133 celebrity and 133 non-celebrity headshots were used as distractor images during the delay period. The photos were high-quality professional headshots of celebrities, models or professionals and were acquired through a Google image search. Each image was cropped to a square with only the face and tops of the shoulders showing. After this, the background was removed so as to eliminate any easy cues to celebrity status, and the image was gray-scaled and centered on a homogenous grey background. This image set was used in a previous experiment in which it was shown that the celebrities were well-known by undergraduates and

that the quality of the celebrity images did not differ from the non-celebrity images (Hacker & Biederman, 2019).

The objects used in the task were geons (Biederman, 1987) which could be metrically varied so that their discrimination closely matched the type of processing required to discriminate two faces (Margalit et al., 2016). Eleven exemplar geons differing in nonaccidental properties were created using Blender (Blender Foundation, Blender Institute, Amsterdam, Netherlands) and three additional metric variations were applied to each so that there were 44 unique geons (Figure 1b; Supplemental Figure 1). The two geon to be discriminated in each geon trial were always from the same class. Again, the uniformity of the computer-generated geons allowed for the use of the Gabor jet model for dissimilarity scaling.

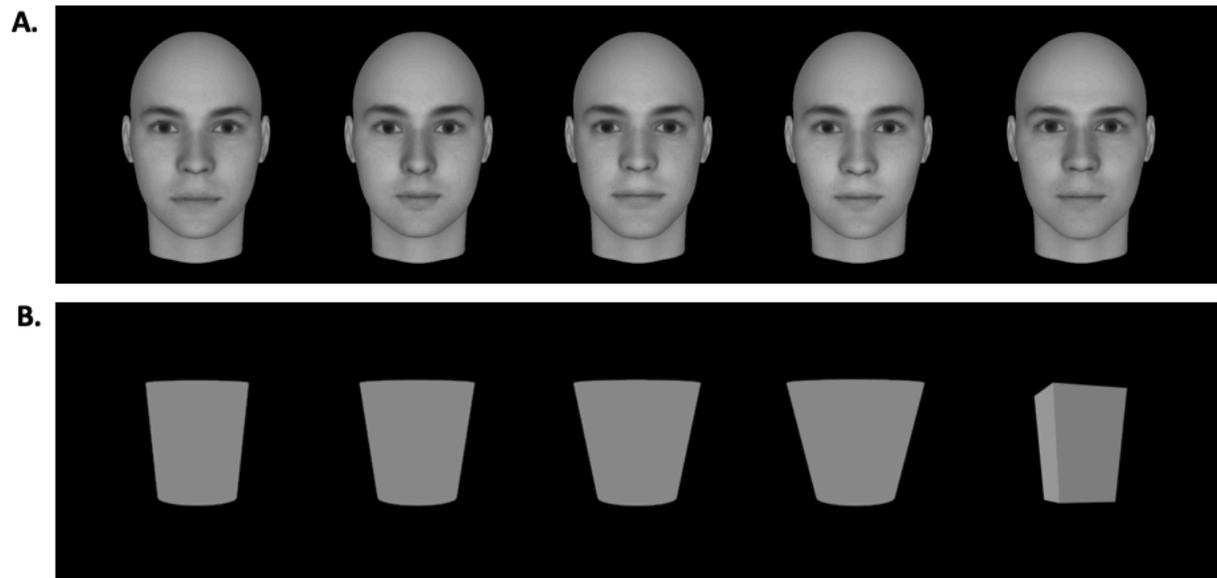


Figure 1. Examples of the stimuli used in the USC Face Memory Test A) Five of the twenty experimental face identities. The Gabor dissimilarity between the first and second face is 174 and between the first and fifth face is 183. B) Four geons of the same type varying metrically and a different geon differing from the other four in nonaccidental properties (right). Geons within a given class varied metrically. The Gabor dissimilarity between the first and second geon is 354 and between the first and fourth geon is 536. The range of matching to foil dissimilarities in the experiment was 92.4-339.7 for the faces and 64.3-656.2 for the geons. Trials only included comparisons between metric variations of the same geons. Exemplars of all experimental stimuli are in Supplemental Figure 1.

Gabor Jet Dissimilarity Scaling

The Gabor jet model (Lades et al., 1993; app and didactic: Margalit et al., 2016) is a model based on V1, simple-cell hypercolumn filtering that yields a similarity value for a pair of faces that almost perfectly predicts their psychophysical discriminability (Yue et al., 2012). To calculate a dissimilarity value, a 10x10 grid of Gabor jets is placed over the image and each point is evaluated at five scales, eight orientations and two phases (Figure 2). This results in an 8,000 x 1 vector of values for each face. The dissimilarity is calculated by finding the Euclidean distance between these two vectors. The higher the dissimilarity value, the more discriminable the faces. Gabor dissimilarity values correlate in the mid .90s with accuracy in discriminating faces in a 2AFC match-to-sample task (Yue et al., 2012).

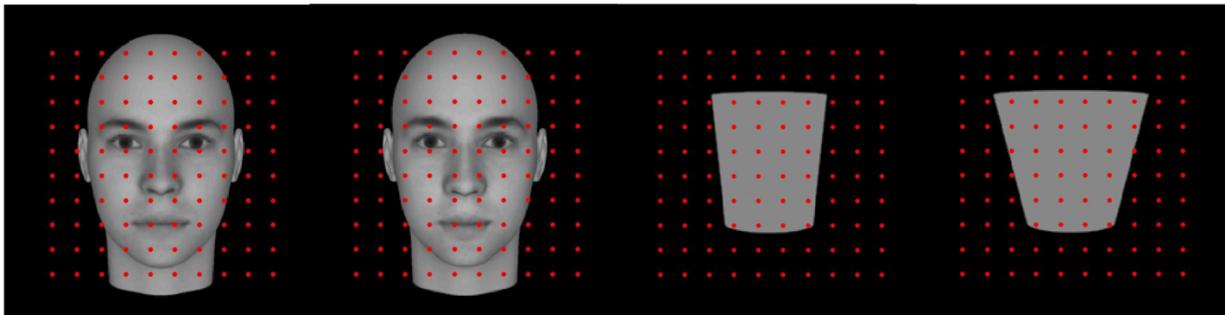


Figure 2. Example of the Gabor jet grid placement on two faces and two geons used in the experimental task. Each of the hundred (10 x 10) points represents the position of one jet where the image is filtered through five scales, eight orientations and two phases. The Gabor dissimilarity between the faces is 219 and between the geons is 397.

Experimental Task

Subjects performed the USC FMT in a dedicated experimental room on a 27" iMac. The core task was a 2AFC delayed match-to-sample task illustrated in Figure 3. Subjects were given 3 s to remember one 6° computer-generated face. Presentation of the sample face was followed by a retention interval of 0, 4, or 12 sec before the two test faces were displayed. All three faces—the sample and two test faces—were never simultaneously on the screen (Figure 3). In the 0 sec delay condition, presentation of the sample face was immediately followed by presentation of the two test faces, the match and the foil. The three test images were positioned so that if they were presented simultaneously they would be in a triangular arrangement. Such an arrangement encourages holistic face processing by preventing local

feature matching of a face presented in the identical location as well as preventing horizontal or vertical alignments from which observers could more readily distinguish the non-matching face (the foil) from the matching face based on pixel differences. Subjects were instructed to indicate as quickly and as accurately as possible which of the two test faces was an exact match to the first (the sample) using the left and right arrow keys. The test faces stayed on the screen for three seconds before disappearing. In the rare case that subjects failed to answer within three seconds, they heard a low beep and that trial was counted as an error. After choosing the test face, subjects heard a high beep if their choice was incorrect and nothing if their choice was correct. To suppress rehearsal of the sample and to maintain attention during the 4 and 12 sec delay periods, subjects judged whether each of the series of professional headshots, each presented for 800 msec, was a celebrity by pressing the space bar. Headshots had a 50% chance of being a celebrity and appeared in a randomized order. After the delay, the trial continued just as in the 0 sec delay trials.

There were 303 total trials, composed of 243 face and 60 geon trials. The geon trials were randomly distributed throughout the session. These trials proceeded in the same manner as the face trials (Figure 3), but the stimuli were simple geons such as those illustrated in Figure 1b (all geons shown in Supplemental Figure 1). Subjects were presented with one geon for three seconds to store in memory, followed by a delay of 0, 4, or 12 sec. When there was a delay of 4 or 12 sec, several geons were presented for 800 ms each and participants were instructed to press the space bar for those with a curved axis as a task to suppress explicit rehearsal of the stored geon and to maintain attention to the task. After the delay, the identical geon and another geon from the same class which varied metrically from the original were simultaneously presented. Subjects had three seconds to indicate which of the two geons was an exact match to the original. The duration of the session, including instructions, was approximately one hour.

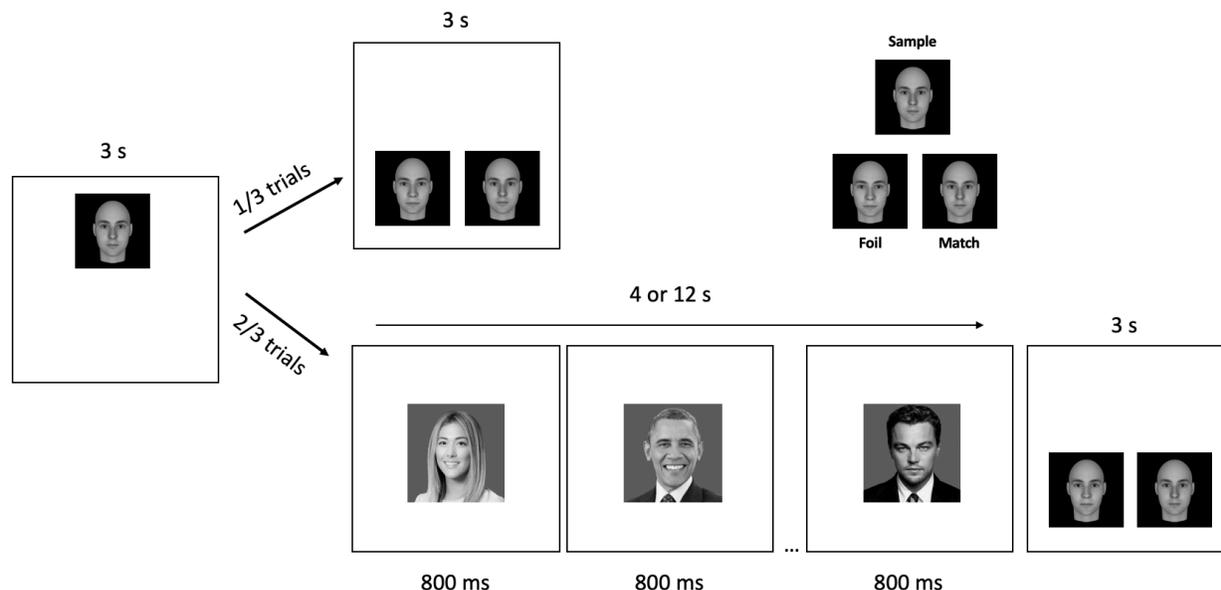


Figure 3. Sample trial illustrating the three possible delay periods (0, 4, or 12 sec). Subjects started each trial by taking three seconds to memorize one identity (the sample). This was followed by a 0, 4 or 12 sec retention interval. When there was a 4 or 12 sec retention interval (bottom half of figure), the interval was filled by a celebrity recognition task in which subjects had to judge if each 800 ms headshot was that of a celebrity. The celebrity recognition task was designed to suppress rehearsal and required that attention be maintained throughout the entirety of each trial. Following the retention interval, a pair of computer generated faces was presented and the subject designated (with the left or right arrow key) which of the faces was a match to the sample.

Additional Tests of Face Recognition Proficiency

Before taking the USC FMT in the lab, subjects completed two additional tests of face recognition proficiency outside of the lab on their personal computers: the USC Face Perception Test (USC FPT) and the Prosopagnosia Index (PI20; Shah et al., 2015). The USC Face Perception Test is a 2AFC match-to-sample task that tests proficiency for face perception. Subjects view one sample face and two test faces in a triangular arrangement and they indicate which of the two test faces is an exact match to the sample (Yue et al., 2012). Unlike the USC FMT in the present experiment, all three images are simultaneously displayed so that there is no memory requirement. The PI20 is a widely employed measure of an individual's self-reported assessment of their face recognition proficiency. Subjects rate their agreement with 20 statements such as "I am better than most people at putting a 'name to a face'" on a scale of 1-5. Higher scores indicate greater difficulty with face recognition. Subjects averaged a PI20 score

of 41.2 (S.D. = 9.8, range = 27 – 67), similar to the mean reported for controls in larger samples (Shah et al., 2015).

Results

Error rates and reaction times increased with retention interval on face and geon trials

There was a strong cost of delay on reaction times and error rates in both face and geon matching. As shown in Figure 4a-b, imposing a 12 sec filled delay between the presentation of sample and test faces increased error rates on face trials by 19.5%, $F(2, 147) = 146.94$, $p < 0.001$, and mean correct reaction times by 300 msec, $F(2, 147) = 46.7$, $p < 0.001$. As shown in Figure 4c-d, imposing the same duration of filled delay between the presentation of sample and test geons increased error rates on geon trials by 8.43%, $F(2, 147) = 14.3$, $p < 0.00001$, and mean correct reaction times by 228 ms, $F(2, 147) = 46.7$, $p < 0.00001$. For both kinds of stimuli, faces and geons, most of the cost—77% and 117% for error rates and 77% and 80% for reaction times on face and geons trials respectively—was observed with just a 4 sec delay.

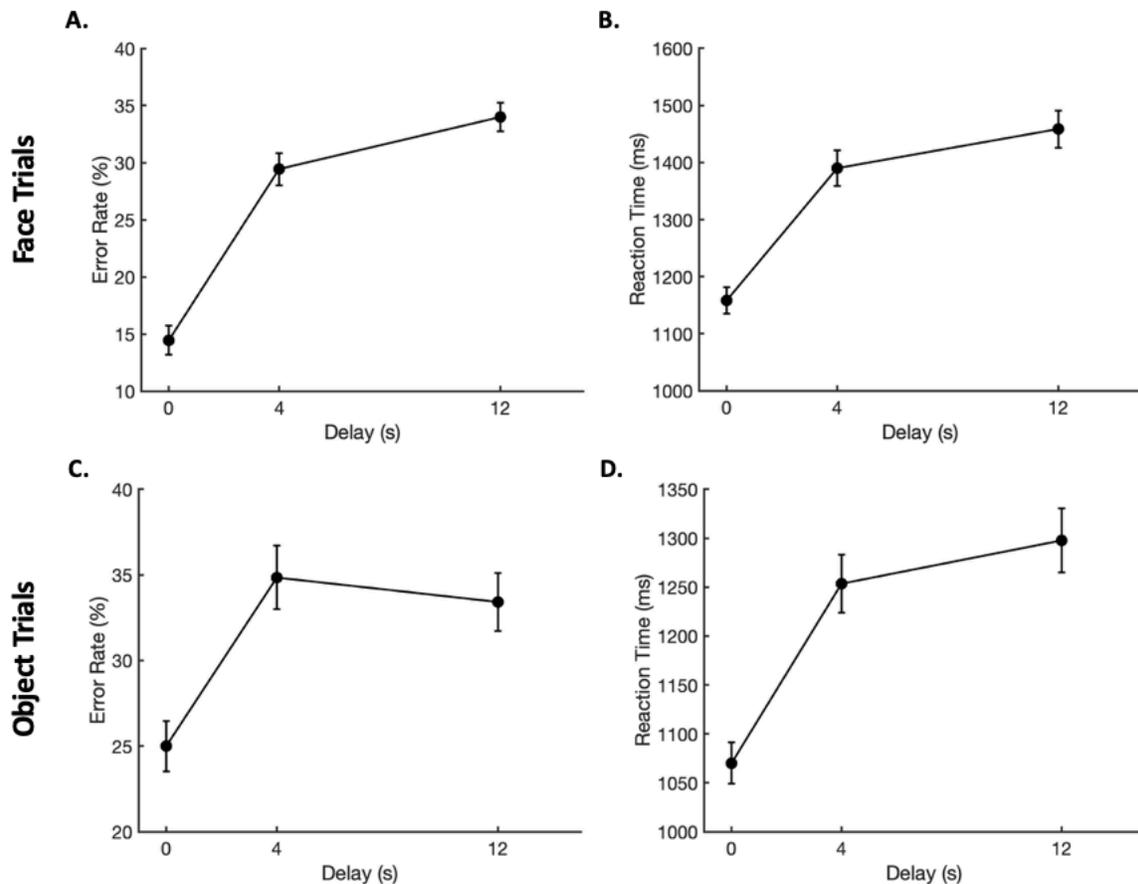


Figure 4. Effect of increased delay period on a) error rates and b) mean correct reaction times for face trials and c) error rates and d) mean correct reaction times for geon trials in the USC FMT. For both faces and geons, increasing delay imposed a strong cost on error rates and reaction times over the first 4 sec with much smaller costs evident between delays of 4 sec and 12 sec.

Group level evidence for an independence of the processes of face perception and face memory

There was a shallow slope of error rates and reaction times over Gabor dissimilarity between the match and foil faces on face trials, with both declining slowly as a function of increased dissimilarity (Figure 5). The shallow slope in Figure 5 was likely a consequence of the small range of Gabor face dissimilarity values in the present task which spanned approximately 123 Gabor units yielding an overall decline in error rates of approximately 5%. In Yue et al. (2012), the range of matching to foil dissimilarity in their match-to-sample task with similar computer-generated faces was approximately 300 Gabor units which yielded a range of error rates of approximately 27%. In the following analysis, the slope of error rates or reaction times over Gabor dissimilarity provides a direct and pure measure of the effect of increasing difficulty

of face perception, while separating the data by duration of the delay period separately accounts for increasing difficulty in holding a face in memory. The core question is whether the effect on RTs of variation in these two variables is additive. This would imply by additive factors logic (Sternberg, 1969) that they affect separate stages. If the slopes are interactive, then the inference is that they are affecting the same stage. To test the independence of the slopes, we ran an analysis of the covariance (ANCOVA) on the binned data points at each duration of the delay. The first step of this analysis, reported here, tests the independence of the slopes of several lines, each formed by a limited number of data points. The slope of error rates and reaction times over Gabor dissimilarity was independent of duration of the delay, error rates: $F(1, 8) = 0.0006$, $p = 0.98$, reaction times: $F(1,8) = 0.004$, $p = 0.95$. If the proficiency for face memory was not independent of the proficiency for face perception, then we would expect that increasing the difficulty of the memory portion of the task would have affected the difficulty of the perceptual task, i.e. the slope. However, perceptual difficulty was unaffected by increased difficulty in the memory portion of the task. By additive-factors logic (Sternberg, 1969; 2011), this result implies that the two processes are independent.

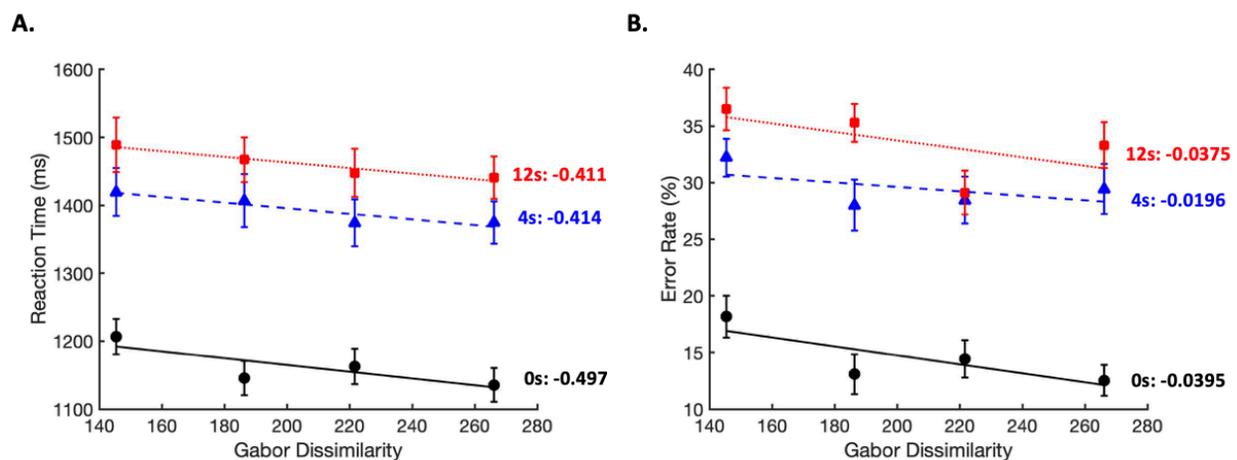


Figure 5. a) Mean correct reaction times and b) error rates for match-to-sample performance on face trials as a function of Gabor dissimilarity between the match and foil faces at three retention intervals. Units for the slope values are a) $\Delta ER/\text{unit Gabor Dissimilarity}$ and b) $\Delta RT/\text{unit Gabor Dissimilarity}$.

The processes of geon perception and memory are also independent

To determine the extent to which the observed independence between perception and memory was unique to faces, we applied an analysis identical to that detailed above to the geon trials. The key question was whether the effect on RTs of variation in Gabor dissimilarity and delay was additive. We again employed an ANCOVA demonstrating that the slope of error rates and reaction times over Gabor dissimilarity was independent of duration of delay, error rates: $F(1,8) = 0.0009$, $p = 0.98$, reaction times: $F(1,8) = 0.0074$, $p = 0.93$. By additive factors logic this additivity implies that the two processes are independent. If the difficulty of the perceptual task, as measured by the steepness of the slope of error rates over Gabor dissimilarity, was dependent upon the difficulty of the memory task, as measured by duration of delay, we would observe an interaction between delay and dissimilarity, but we did not. This suggests that the separate stages of processing for face perception and face memory is not unique to faces but may be a general property of geon, if not more generally, object memory.

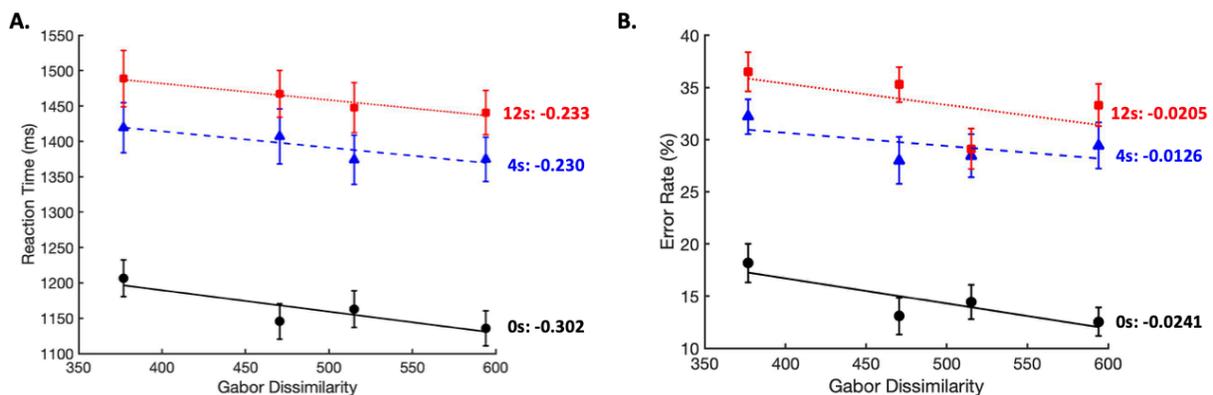


Figure 6. a) Mean correct reaction times and b) error rates for match-to-sample performance on geon trials as a function of Gabor dissimilarity between the match and foil geons at three retention intervals. Units for the given slope values are a) $\Delta RT/\text{unit Gabor Dissimilarity}$ and b) $\Delta ER/\text{unit Gabor Dissimilarity}$.

Individual proficiency for face perception is independent of proficiency for face memory

Having found group level evidence that face perception and face memory are independent processes, we next assessed whether these two proficiencies might be independent across individual subjects. We sought to determine to what extent subjects with a higher proficiency for face perception also had a higher proficiency for face memory.

For each subject, separate proficiency scores were calculated for face perception and face memory. In order to avoid conflation with possible speed for accuracy tradeoffs, we calculated the inverse efficiency (IE) measure for each subject (Townsend & Ashby, 1983). Inverse efficiency is calculated as $RT/(1-ER)$ and thus combines RT and error rate data to avoid ambiguous correlations that may be caused by a speed for accuracy tradeoff. Consequently, the inverse efficiency measure provides a clearer picture of what correlations may exist between proficiency for face memory and face perception than either error rates or mean correct reaction times considered alone or together. Recreating the analyses in Figure 4a-b and Figure 5 with inverse efficiencies reveals the same general population trends, further validating the use of inverse efficiencies (Figure 8a-b).

For face memory, the proficiency score was defined as the mean inverse efficiency with a 12 sec delay minus the same measure with a 0 sec delay on the USC FMT. This measure quantifies how much an individual subject was affected by lengthening the filled retention interval of the memory portion of the task and subtracts out baseline ability to isolate a cost of delay for each individual subject. Those with a higher proficiency for face memory are defined as having a smaller cost of delay. The proficiency for face perception was calculated using data from the USC FPT in which all faces are simultaneously displayed such that there is no memory component (see methods section). Correlations between performance on the USC FPT and USC FMT were moderate to strong and highly significant (Figure 7), accuracy: $r(39) = 0.633$, $p < 0.00001$, 95% CI = [0.41, 0.79], reaction times: $r(39) = 0.420$, $p < 0.01$, 95% CI = [0.13, 0.64]. Individual proficiency was calculated as the slope of inverse efficiencies over Gabor dissimilarity in the USC FPT. This measure quantifies how much an individual subject was affected by increasing the perceptual difficulty of the trials, and again subtracts out baseline ability to isolate the effect of increased perceptual difficulty. Subjects with a greater proficiency for face perception are defined as having a shallower slope, i.e., less negative value. Slopes are generally negative because inverse efficiencies (and error rates and RTs) decrease with increasing Gabor dissimilarity.

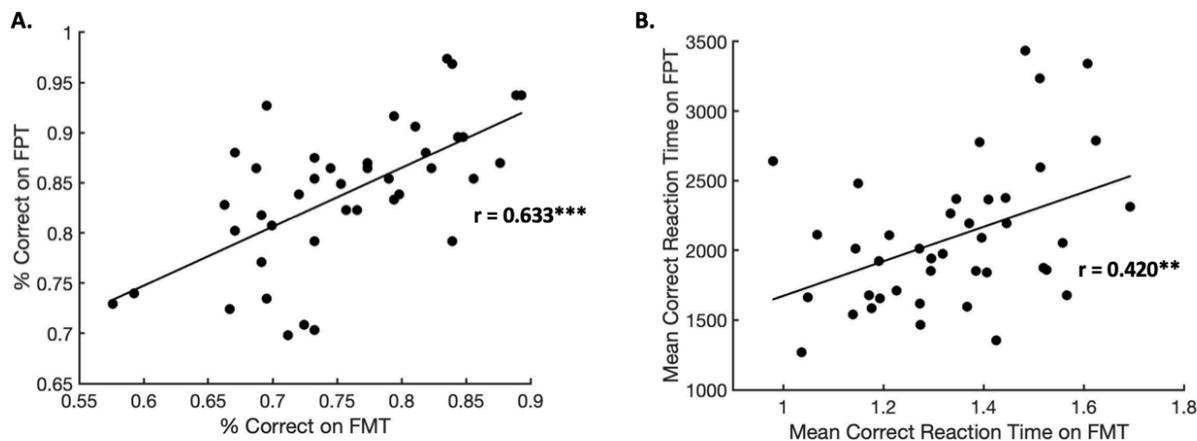


Figure 7. Correlations across subjects for overall a) accuracy (% Correct) and b) mean correct reaction times between the USC FPT and the USC FMT.

As seen in Figure 8c, the proficiency for face perception is uncorrelated with the proficiency for face memory, $r(39) = -0.092$, $p = 0.57$, 95% CI = [-0.39, 0.22]. Correlations calculated using error rates and mean correct reaction times yielded inconsistent results, likely due to a speed for accuracy tradeoff as explained above (Supplemental Figure 2). Odd/even trial reliability for error rates and reaction times were high for both the USC FMT and the USC FPT, USC FMT error rates: $r(39) = 0.719$, reaction times: $r(39) = 0.946$, USC FPT error rates: $r(39) = 0.847$, reaction times: $r(39) = 0.963$. Consequently, the different pattern of correlations across subjects is likely not due to variations in the reliability of the RT and error rate measures. These results suggest that an individual's proficiency for face perception is unrelated to their proficiency for face memory such that subjects with greater proficiency for face perception do not necessarily have greater proficiency for face memory.

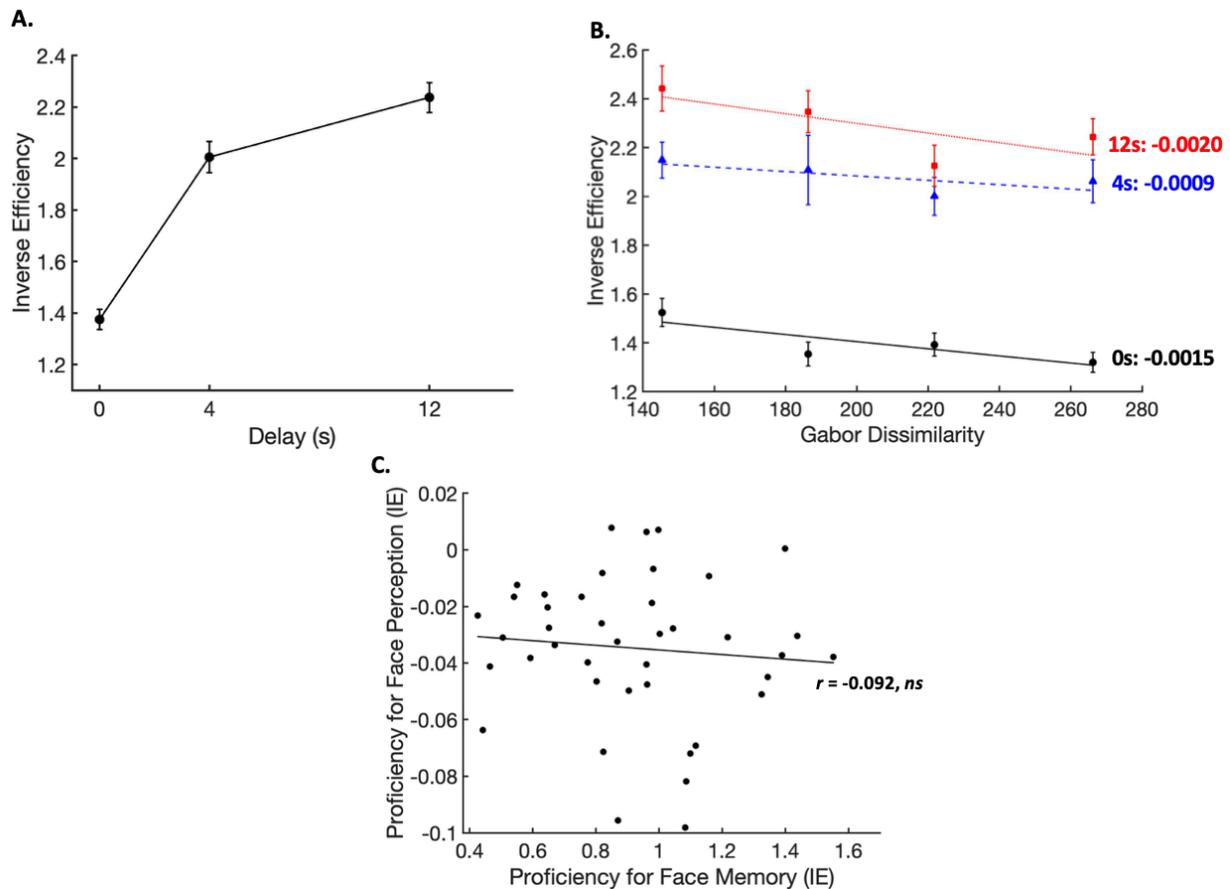


Figure 8. Inverse efficiency results. a) Inverse efficiency at each retention interval in the USC FMT b) Inverse efficiency across Gabor dissimilarity between the match and foil faces at three retention intervals c) Correlation between the proficiency for face memory and the proficiency for face perception. The proficiency for face memory is calculated as the difference between inverse efficiencies with a 12 sec delay versus a 0 sec delay on the USC FMT. The proficiency for face perception is calculated as the slope over Gabor dissimilarity of inverse efficiencies on the USC FMT.

Discussion

This experiment documents a strong cost of an increase in a filled delay period on performance in a delayed match-to-sample task with faces and geons. Despite the strong effect of delay, slopes of mean correct reaction times for each delay period over Gabor dissimilarity were parallel as shown in Figure 5 for faces and Figure 6 for geons, providing evidence that face and geon perception and face and geon memory are independent processes and affect separate processing stages. Such independence in neurotypical subjects is consistent with a segregated face representation in cortex wherein increased difficulty of different aspects of face recognition—face perception and face memory—are processed in separate stages, as

modeled by the additive factors method (Sternberg, 1969). Here we evaluated two possible independent components of face recognition – proficiency for face memory and proficiency for face perception—but there are many other possible components. One possibility is invariance to rotation in depth or expression, which is an important part of our ability to generalize learned identities in real-world face recognition. This independence could be tested in a manner similar to that described in the present investigation.

Some biological evidence for a dissociation between the proficiencies for face memory and face perception comes from a recent neuroimaging study by Ramot et al. (2019). These authors isolated eight face-selective ROIs per subject as areas showing a larger BOLD response to a face than to a scene: bilateral fusiform face area (FFA), occipital face area (OFA) and amygdala and the right anterior temporal lobe (ATL) face patch and right posterior Superior Temporal Sulcus (pSTS). They reported that these face-selective loci are tightly coupled at rest, but that the strength of these connections could not predict face memory as measured by performance on the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006). To determine which connections were most predictive of face memory, they correlated ROIs and measured their correlation with CFMT scores, finding that correlations between medial temporal lobe and face-selective ROIs were most predictive of CFMT scores and correlations between face-selective ROIs were not. Consequently, they concluded that face memory engages a distinct network from face perception.

By performing the same analysis for face and geon trials, we demonstrate that the independence of the proficiency for memory and perception is true of geon, and perhaps more generally, object trials and is not specific to faces. The present study employed simple objects, geons, which are much less complex than faces. Congenital prosopagnosics (CPros) who struggle to match faces, show no deficits in matching metrically varying geons, which do not resemble biological entities, nor do they have the complexity or the ineffability of the subtle metric variation as faces (Biederman & Kalocsai, 1997). CPros do show deficits, relative to non-prosopagnosic participants, in matching blobs--harmonics of a sphere--which more closely resemble biological entities in matching the smooth complex 3D sculpting of faces and can vary metrically with the same Gabor dissimilarity of a set of faces (Yue et al., 2012). Indeed, the

prediction of psychophysical similarity from the Gabor-jet scaling is, like it is for faces, almost perfect although the blobs do not activate FFA above baseline. We demonstrate that the independence across and within subjects apparent in the perception and memory for simple geon matching is also characteristic of the processing of complex faces.

We also provide the first rigorous evidence that behavioral measures for the proficiency for face perception and face memory are independent across subjects such that neurotypical subjects with a greater proficiency for face perception do not necessarily also have a greater proficiency for face memory. The independence of face memory and face perception across neurotypical subjects may also be true of CPros as suggested by the findings of Dalrymple et al. (2014). CPros have difficulty recognizing faces, but do not have any observable lesions. Dalrymple et al. (2014) found a dissociation between face memory and face perception in adult CPros by first identifying CPros and then comparing their deficits on two face memory tests (old/new and the Cambridge Face Memory Test) with the Cambridge Face Perception Test (DFPT). They found that some CPros had a deficit in face perception and face memory while other CPros only showed a deficit in face memory. This suggests that although congenital prosopagnosics have traditionally been considered a homogenous population, there might be two or more subtypes of prosopagnosia characterized by deficits to different combinations of components of face recognition. This would be consistent with subdivisions observed in acquired prosopagnosia wherein prosopagnosics show general face recognition difficulties while prosopamnesics have deficits specific to face memory (Tippett et al., 2000; Williams et al., 2007; Lopera & Ardila, 1992). The observed dissociation between proficiency for face memory and face perception in neurotypical and prosopagnosic populations strongly suggests that future tests of face recognition proficiency should assess these two components separately in evaluating general proficiency and deficits.

Acknowledgements

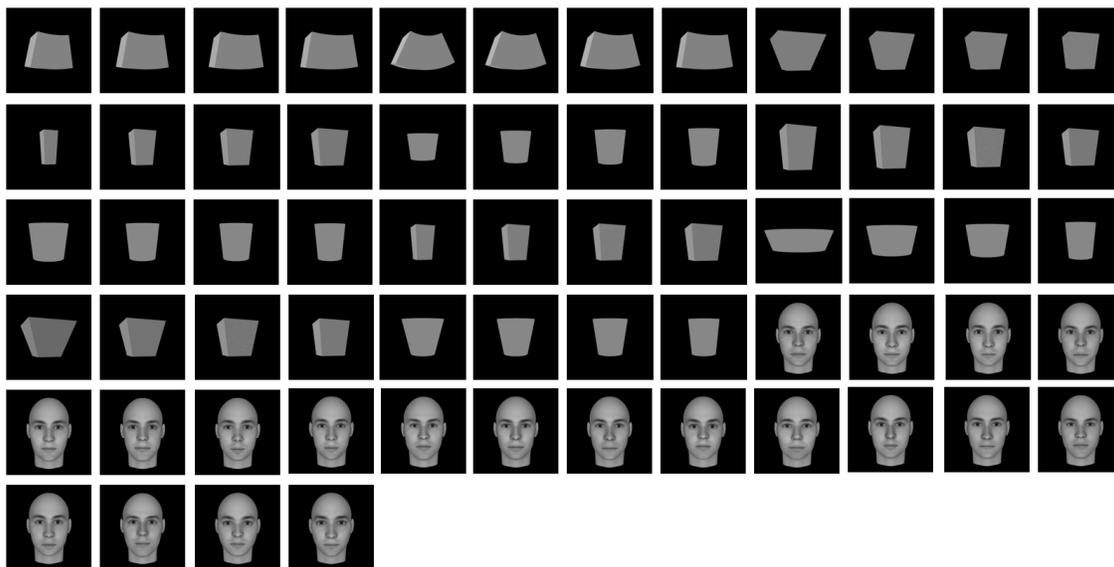
We thank Wendy Qi for creating the computer-generated face images used in the USC FMT. This research was supported by the Harold W. Dornsife Research Fund.

References

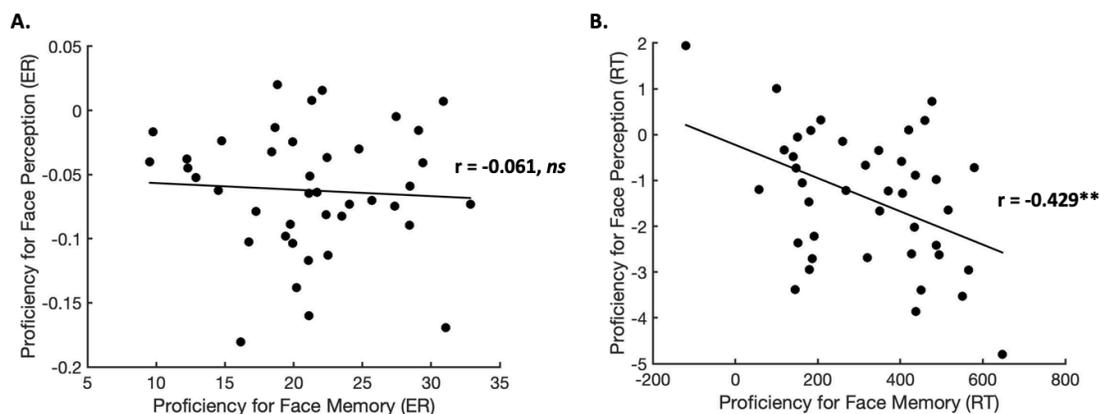
- Benton, A. L., Sivan, A. B., Hamsher, K. De S., Varney, N. R. & Spreen, O. (1994). *Contribution to Neuropsychological Assessment*. New York: Oxford University Press.
- Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94, 115-147.
- Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society London: Biological Sciences*, 352, 1203-1219.
- Biederman, I., Margalit, E., Maarek, R., Meschke, E., & Shilowich, B. (2017). What is the perceptual deficit in congenital prosopagnosia? *Journal of Vision*, 17, 619. Doi:<https://doi.org/10.1167/17.10.619>.
- Dalrymple, K. A., Garrido, L. & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience*, 10, 10-20. doi: 10.1016/j.dcn.2014.07.003.
- Duchaine, B. & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44, 576-585. doi: 10.1016/j.neuropsychologia.2005.07.001.
- Hacker, C. M. & Biederman, I. The invariance of recognition to the stretching of faces is not explained by familiarity or warping to an average face. *PsyArxiv*. doi: 10.31234/osf.io/e5hgx.
- Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P. & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42, 300-311. doi: 10.1109/12.2101.
- Lopera, F. & Ardila, A. (1992). Prosopamnesia and visuolimbic disconnection syndrome: A case study. *Neuropsychology*, 6(1), 3-12. doi: 10.1037/0894-4105.6.1.3.
- Margalit, E., Biederman, I., Herald, S. B., Yue, X. & von der Malsburg, C. (2016). An applet for the Gabor scaling of the differences between complex stimuli. *Attention, Perception, & Psychophysics*, 78(8), 2298-2306. doi: 10.3758/s13414-016-1191-7.
- Margalit, E., Shah, M. P., Tjan, B. S., Biederman, I., Keller, B. & Brenner, R. (2016). The lateral occipital complex shows no net response to object familiarity. *Journal of Vision*, 16, 1-8. doi: 10.1167/16.11.3.
- Ramot, M., Walsh, C. & Martin, A. (2019). Multifaceted integration- memory for faces is subserved by widespread connections between visual, memory, auditory and social networks. *Journal of Neuroscience*, 39, 4976-4985. doi: 10.1523/JNEUROSCI.0217-19.2019.
- Shah, P., Gaule, A., Sowden, S., Bird, G. & Cook, R. (2015). The 20-item prosopagnosia index (PI20): a self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science*, 2:140343. doi: 10.1098/rsos.140343.
- Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. *Acta Psychologica Attention and Performance*, 30, 276-315. doi: 10.1016/0001-6918(69)90055-9.

- Sternberg, S. (2011). Modular processes in mind and brain. *Cognitive Neuropsychology*, 28, 3-4. doi: 10.1080/02643294.2011.557231.
- Stollhoff, R., Jost, J., Elze, T. & Kennerknecht, I. (2011). Deficits in long-term recognition memory reveal dissociated subtypes of congenital prosopagnosia. *PLoS One*, 6, e15702. doi: 10.1371/journal.pone.0015702.
- Tanaka, J.W. & Farah, M.J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46A, 225-245. doi: [10.1080/14640749308401045](https://doi.org/10.1080/14640749308401045).
- Tippett, L. J., Miller, L. A. & Farah, M. J. (2000). Prosopamnesia: A selective impairment in face learning. *Cognitive Neuropsychology*, 17, 241-255. doi: 10.1080/026432900380599.
- Townsend, J. T., & Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.
- Yue, X., Biederman, I., Mangini, M. C., von der Malsburg, C. & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. *Vision Research*, 55, 41-46. doi: 10.1016/j.visres.2011.12.012.
- Weigelt, S., Koldewyn, K., Dilks, D. D., Balas, B., McKone, E. & Kanwisher, K. (2014). Domain-specific development of face memory but not face perception. *Developmental Science*, 17, 47-58. doi: 10.1111/desc.12089.
- Williams, M. A., Berberovic, N. & Mattingley, J. B. (2007). Abnormal fMRI adaptation to unfamiliar faces in a case of developmental prosopamnesia. *Current Biology*, 17, 1259-1264. doi: 10.1016/j.cub.2007.06.042.

Supplemental Figures



Supplemental Figure 1. All of the geons and faces that appeared in the discrimination portion of the USC Face Memory Test. There were 11 classes of geons, each with four levels of metrically varying properties. There were 20 unique face identities, each of which differed from all other faces only by moderate metric variations.



Supplemental Figure 2. Correlation between the proficiency for face perception and the proficiency for face memory calculated using a) error rates, $r(39) = -0.061, p = 0.71, 95\% \text{ CI}: [-0.362, 0.251]$ and b) mean correct reaction times, $r(39) = -0.429, p < 0.01, 95\% \text{ CI}: [-0.651, -0.140]$. The error rate and reaction time data yielded an inconsistent picture for the proficiencies for face perception and face memory. Odd/even trial reliability for error rates and reaction times were high for both the USC FMT and the USC FPT, USC FMT error rates: $r(39) = 0.719$, reaction times: $r(39) = 0.946$, USC FPT error rates: $r(39) = 0.847$, reaction times: $r(39) = 0.963$. Consequently, the different pattern of correlations across subjects is likely not due to variations in the reliability of the RT and error rate measures. More likely, based on the lack of correlation when inverse efficiency is used (Figure 8), this inconsistency is due to a speed for accuracy tradeoff.