
Electronic appendices are refereed with the text. However, no attempt has been made to impose a uniform editorial style on the electronic appendices.

This document provides more details on the experimental design described in the main paper.

1 Experiment 1: Selection of Face Pairs

Experiment 1 was devoted to the collection of a well-controlled yet substantially large stimulus set for Experiments 2 and 3. To consistently measure discrimination impairment as a result of inversion, this stimulus set needed to consist of images for which subjects were equally good at detecting configural and featural changes when upright. Any significant difference in upright detection would be a confounding variable since it could also affect how featural and configural cues were processed when inverted.

1.1 Subjects

A total of thirty adults were randomly assigned to one of two subject groups. Thus, each subject group contained 15 mutually exclusive participants. All subjects were naive as to the exact purpose of the experiment. Subjects had normal or corrected-to-normal vision.

1.2 Stimuli

Two hundred original male and female face images (the face prototypes from [1]) were manipulated to compile 2,428 pairs of gray-scale, frontal faces used in this experiment. To avoid providing easy cues for identification, the faces were cut at the hairline, and did not include distinctive features such as beards or spectacles. The 200 images were divided into two groups: source faces and target faces. Using a customized morphing program written in MATLAB (The Mathworks, MA), image manipulations were applied as follows: Each source face was duplicated and prepared in 10 configural versions and 10 featural versions, differing in the position (for the configural set) or shape (for the featural set) of the eye and/or mouth region [2, 3, 4]. Configural changes involved spatial shifts of the eyes and/or mouth such that the basic symmetry of a face was preserved. Eyes were
displaced a maximum of two pixels along the x-axis (in or out) and/or two pixels along the y-axis (up or down). The mouth was displaced a maximum of two pixels along the y-axis (up or down). Both magnitude and direction of movement were randomly selected. Featural changes were accomplished by replacing the eyes and mouth of a source face with the corresponding features from a randomly selected target face. Target features used on a given source face were selected from two different faces (for instance, source A received eyes from Y and mouth from X), and the combination was unique. Source faces were matched with target features of the same gender. Note that all eye alterations treated the eyes and eyebrows as a single unit.

Each of the 20 created images were then paired once with themselves, and once with another image that underwent the same type of manipulation. The above procedure was repeated for each of the 100 source faces, resulting in a total of 4,000 unique pairs. These pairs were then screened for morph artefaces. The resulting stimulus set was comprised of 2,428 image pairs, with an equal number of those pairs (607) belonging to each of the four stimulus categories: pairs with the same features (SF), same configuration (SC), different features (DF), and different configuration (DC).

1.3 Procedure

Two subject groups underwent the exact same experiment, with the exception that the stimulus set presented to each group consisted of different images. This design allowed us to collect a large stimulus set while avoiding noise due to subject exhaustion. The resulting data consisted of 15 different responses for each of the 2,428 image pairs. Prior to beginning the experiment, both written and oral instructions were given. Subjects were told that the goal of the experiment was to investigate how well they could distinguish between similar faces, and warned about the possible/potential similarity of the two faces (like identical twins with different expressions).

Subjects were presented with approximately 10 blocks, each with 100 image pairs (numbers differed somewhat for each subject groups). The order of presentation was randomly selected at the beginning of each experiment. In a given trial, a selected pair was presented as a sequence of two images using the following design: cross (300ms) → image 1 (300ms) → random noise mask (700ms) → image 2 (300ms) → blank screen until subjects respond by pressing one of two labeled keys. Image selection, presentation, and the recording of subject input were controlled using a MATLAB (The Mathworks, MA) program.
2 Experiment 2: Face Discrimination (Unblocked)

In this experiment, featural and configural image pairs were randomly presented both upright and inverted. The aim was to determine how well subjects could judge two inverted pictures to be the same or different depending upon whether the images differed by a featural or a configural change.

2.1 Subjects

Fifteen subjects, ranging in age from 19 to 56 (mean age 33), participated in the study for payment.

2.2 Stimuli

Through Experiment 1, the initial stimulus set of 2,428 image pairs was reduced to 80 for Experiment 2 by choosing only among those that were correctly judged to be the same or different two images approximately 80% of the time (78.0% on average for the entire set). The 80 pairs were evenly divided into each of the four stimulus categories: SC, DC, SF, and DF.

2.3 Procedure

Each subject viewed all 80 image pairs in the stimulus set both upright and inverted, thus leading to 160 trials in the entire experiment. Both the presentation order of stimulus type (SC, DC, SF, or DF) and orientation (whether upright or inverted) was randomized and counterbalanced.

In each trial, subjects were shown the two pictures in a face pair in a sequence as before (see Expt. 1), either both upright or both inverted (see Fig. 1 in the paper). Each subject received the same oral instructions from the same experimenter who ran all the subjects in Experiments 2 and 3. Subjects were told that the goal of the experiment was to measure their abilities to differentiate between two similar faces. They were warned that the two pictures in each face pair were going to be very similar (like pictures of two twins or different expressions). Their task was simply to determine whether the two pictures were identical or not. They were explicitly instructed that they did not have to judge whether the two images were of the same face or person, just whether there was any perceptible difference between the two images. Subjects were also informed that there were 160 trials in all with a break halfway through the experiment for them to rest if they liked.

Images were displayed on a 21” Trinitron monitor, at a true size of 186 x 186 pixels (the remainder of the screen remained neutral black throughout the experiment). The screen resolution was set to 1024 x 768 pixels, with a refresh rate of 60 Hz. Image selection, presentation, and the recording of subject input
were automated and controlled using a MATLAB program. Subjects freely viewed the monitor from a distance of about 60 cm.

3 Experiment 3: Face Discrimination (Blocked)

Experiment 3 was identical to Experiment 2 in all the details of its stimuli and procedure except that featural trials (SF,DF) and configural trials (SC,DC) were grouped together in different stimulus blocks corresponding to the first and second halves of the experiment. Within each block, presentation order of stimulus type and orientation was randomized. If the featural block was first, for example, subjects would see 80 featural trials, half upright, half inverted and half with the same features (SF), half different (DF) in the first half of the experiment. Whether any particular subject saw the featural or configural block in the first half of the experiment and the other in the second half was randomized. Subjects received no indication that the trials were blocked by stimulus type and were told, as in Experiment 2, that the break halfway through the experiment was intended for them to rest if they liked.

3.1 Subjects

Twenty-five subjects, ranging in age from 18 to 52 (mean age 24), with normal or corrected-to-normal vision, participated in the study for payment. Subjects were randomly placed in two non-overlapping groups corresponding to the two blocked presentation orders (featural or configural block first). The mutual exclusion was enforced to prevent any transfer of information from one condition to another. Twelve subjects were tested in the configural block first condition with the remaining thirteen viewing the featural block first.

3.2 Results

Subject performance for the two different groups is shown in Figure 3.
Figure 3: Subject performance in the blocked discrimination experiment (Experiment 3). The upper plot shows the performance of subjects (n=12) in the “configural first” group, while the lower plot shows the performance of subjects (n=13) in the “featural first” group. The bars show subjects’ performance in each group for the different trial types (“conf.” refers to trials where the two images differed by a configural change, “feat.” to trials with featural changes, and “same” to trials with two identical images; “up” refers to upright images, “inv.” to inverted images). Error bars show standard error of the mean.

References