OMB Control No. 0693-0043 – NIST Generic Clearance for Usability Data Collections

Measuring the accuracy of facial forensics comparisons

FOUR STANDARD SURVEY QUESTIONS

1. Explain who will be surveyed and why the group is appropriate to survey.

Previously, a study was conducted to measure the accuracy of facial forensic examiners at forensic comparisons of faces. In this forensic comparison study, the accuracy of facial forensic examiners was compared against the accuracies of face super-recognizers and fingerprint examiners. However, matching faces is only one aspect of face recognition; there are other aspects such as memory, other race effect, different poses, and disguises.

In order to better understand the face processing system of facial forensic examiners as well as people with superior face recognition ability, facial forensic examiners, face super-recognizers, and fingerprint examiners who participated in the forensic comparison study will be asked to take five experiments online in order to connect their results from the experiments as well as the forensic comparison study.

2. Explain how the survey was developed including consultation with interested parties, pre-testing, and responses to suggestions for improvement.

The experiments were designed by experts in order to test different aspects of face recognition: memory, other race effect, different poses, and disguises.

Matching faces involves looking at two face images side-by-side. In facial recognition, the memory aspect refers to remembering faces seen before and being able to identify the faces when seen at a later time. Duchaine & Nakayama (2006) created the Cambridge Face Memory Test (CFMT) to test face memory. The Glasgow Face Matching Test (GFMT), developed by Burton, White, & McNeill (2010), is a standard benchmarking test used throughout the literature to assess performance on matching faces. The GFMT is also used to compare performances across studies. Performance on the CFMT and GFMT allows us to compare our results to other works.

For clarification, the following list describes the names of the attachment's described in this supporting statement.
Attachment D: Publication: The Glasgow Face Matching Test
Attachment E: Publication: The Cambridge Face Memory Test
Attachment F: Publication: Face Recognition in Challenging Situations

Attachment G: Publication: An Other-Race Effect for Face Recognition Algorithms
Attachment H: (FES) Additional Questions
Attachment I: Cambridge Face Memory Test (CFMT) Collection Instrument
Attachment J: Glasgow Face Matching Test (GFMT) Collection Instrument
Attachment K: Other Race-Effect (ORE) Collection Instrument
Attachment L: Pose Experiment Collection Instrument
Attachment M: Disguised Faces Collection Instrument

The other race effect (ORE) looks at the change in accuracy when recognizing faces of different races. The ORE experiment will measure the effect of race on the accuracy of forensic examiners; Phillips, Jiang, Narvekar, Ayyad, & O'Toole (2011) created an ORE test using Caucasian and East Asian faces. In the forensic comparison study, all faces images were facing the camera. The Pose experiment will measure accuracy when the faces are not always facing the camera. The Disguised Faces experiment will measure accuracy when people are in disguise; Noyes (2016) developed a test to measure the effect of disguised faces on forensic examiners.

To conduct these experiments, additional information needs to be collected (see Attachment H); three questions will be asked over the phone. Specifically, the questions relate to the CFMT, GFMT, and ORE experiments.

The CFMT and GFMT are standard benchmarking tests and repeated trials may result in better results as participants may memorize the faces; therefore, it is necessary to know if the participant has taken one or both of these.

Measuring the other race effect requires the experimenters to know race of both the faces images presented to the subject and the race of the subject. We know the race of the faces in the images. The questionnaire will let us know the race of the participant. The race question is restricted to the minimum information needed to determine the race categories relevant to this experiment.

3. Explain how the survey will be conducted, how customers will be sampled if fewer than all customers will be surveyed, expected response rate, and actions your agency plans to take to improve the response rate.

The experiments are part of a study to investigate different aspects of face recognition: memory, other race effect, pose, and disguises. The descriptions of the five experiments below state how many questions are asked. Each of the five experiments will present face images and the participants are required to compare or remember the faces. For scientific rigor, the same question with different faces is asked multiple times. In our experiments this varies from 20-156 repetitions. In each experiment, the faces are selected from a set of face images. These sets of images are owned by different organizations and were collected with human subjects approvals that limit their distribution. The complete set of images are not attached due to the license terms with regard to distribution and copyright and the US Government Human Subjects Regulations. Because of this, attached are mock-ups of the five experiments with only a handful of questions (Attachments I-M); the images shown in the mock-ups are publishable from the experiments and are representative of the images shown in the complete experiments. Attachments D-G explain the experiments in detail.

In the CFMT experiment, participants are shown images of faces and then asked to select the face that they were asked to memorize from a set of three to six images. In the practice and first block, participants are shown three views of an individual each for 3 seconds before being asked to choose the face they just memorized among different groups of three. In the other blocks, participants are asked to memorize six faces for 20 seconds before being asked to choose a face they just saw among different groups of three. In total, there are 102 questions.

In the GFMT experiment, participants are shown a pair of faces for up to 30 seconds and asked if they are of the same person or of different people; if the participant does not respond within 30 seconds, the faces will disappear from the screen, and they will need to enter a response before being able to continue to the next pair of images. In total, there are 40 questions.

In the ORE, Pose, and Disguised Faces experiments, participants will be shown a pair of faces for up to 30 seconds and asked to rate the similarity of a pair of faces on a 5-point scale; if the participant does not respond within 30 seconds, the faces will disappear from the screen, and they will need to enter a rating before being able to continue to the next pair of images.

In the ORE experiment, the pairs will be of Caucasian and East Asian faces. In total, there are 80 questions.

In the Pose experiment, the pairs of faces are not all frontal. In total, there are 20 questions.

In the Disguised Faces experiment, the pairs of faces include a face that is disguised. In total, there are 156 questions.

Participants will be asked the additional questions over the phone (see Attachment H: [FES] Additional Questions). The five experiments (see Attachments I-M) will be conducted over the web.

Participants will come from the facial forensic, face recognition, and fingerprint communities. Emails asking for potential participants (recruitment announcement) will go to subjects who completed the previous forensic comparison study.

We estimate that we will receive up to 153 responses from the four groups described above. We estimate we will receive up to 87 responses from facial forensic examiners, 13 responses from face super-recognizers, and 53 responses from fingerprint examiners.

> Facial forensics examiners are people who perform forensic facial comparisons as part of their jobs and have been trained to perform facial forensic comparisons.
>
> Face super-recognizers are people who have been acknowledged as one by at least one of two ways: has taken a test establishing them as one or was recruited to work using their skills as a face super-recognizer.
>
> Fingerprint examiners are people who perform forensic fingerprint comparisons as part of their jobs and have been trained to perform fingerprint comparisons. These examiners must not perform or have performed facial forensic comparisons as part of their jobs. They must also have been trained to perform facial forensic comparisons.

We expect all participants to finish at least one experiment.

No personally identifiable identification (PII) will appear in publication. Only contact information will be obtained from subjects. Any data linked to PII will be stored on encrypted devices. To de-identify the data, each participant is assigned a random study number. All research is performed on de-identified data with study numbers linking the results; therefore, none of the analysis links to participants. PII will only be kept to contact the participants. The master list linking the names of the subjects to their scores will be destroyed when the first paper is published (estimated to be in 2018). Once the link is destroyed, it will not be possible to link the analyses to the PII.

4. Describe how the results of the survey will be analyzed and used to generalize the results to the entire customer population.

This study will address concerns raised in the National Academy of Sciences (NAS) forensic study *Strengthening Forensic Science in the United States: A Path Forward*. The NAS study called for more scientific rigor and the need to establish an underlying forensics science, particularly in forensic pattern analysis such as comparing two faces. Facial forensics examiners are trained to compare faces that appear in images and videos to determine if the two faces are the same person or not, providing justification for their conclusions. They are prepared to testify

in court and to defend their conclusions.

The aim of this study is to understand the processing system of facial forensics examiners and people with superior face recognition ability.

Using methods from signal detection theory, analysis of variance, and correlation theory, the results of these experiments will be compared with each other and the results from the previous black-box study.

Burton, A.M., White, D., & McNeil, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42*(1). doi:10.3758/BRM.42.1.286

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*(4). doi:10.1016/j.neuropsychologia.2005.07.001

National Research Council. (2009). *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press.

Noyes, E. (2016). *Face Recognition in Challenging Situations*. (doctoral dissertation, University of York).

Phillips, P.J., Jiang, F., Narvekar, A., Ayyad, J., & O'Toole, A.J. (2011). An other-race effect for face algorithms. *ACM Transactions on Applied Perception, 8*(2). doi:10.1145/1870076.1870082