OMB Control No. 0693-0043
Expiration Date: 6/30/2025
NIST Generic Clearance for Usability Data Collections

Measuring the difference in accuracy of forensic facial comparisons of different races

FOUR STANDARD SURVEY QUESTIONS

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

Forensic facial examiners are highly trained to perform forensic comparisons of faces in images and determine if they are of the same person or not. They document their findings in detailed reports which can be used in court.

Forensic facial reviewers are trained to perform quicker forensic comparisons of faces in images in order to generate leads for investigators.

Super-recognizers are people who have naturally high ability at face recognition but do not have any forensic training.

A control group will consist of people with no forensic training and have not been identified as super-recognizers.

In a cross-race comparison, the race of the facial examiner and the face in the images is different (this definition applies to facial reviewers, super-recognizers, and the general population). The aim of this study is to understand the effect of cross-race comparisons on forensic facial examiners, reviewers, and super-recognizers when the image subject is of a race different than their own.

Phillips *et al* (2018) investigated the accuracy of forensic facial examiners, reviewers, and super-recognizers compared to untrained control participants and computer algorithms which were state-of-the-art at the time of publishing. Forensic facial examiners and reviewers were allowed to have access to their tools and methods and had up to three months to perform the twenty comparisons. Phillips *et al* (2018) demonstrated that all three groups were equally accurate on a challenging face recognition test and that all three were more accurate than untrained participants. In a follow-up study, Yates *et al* (2023) looked at the effect on accuracy at cross-race comparison for examiners and reviewers. The study found that the accuracy of examiners and reviewers dropped for cross-race comparisons. In this study, examiners and reviewers only had limited time to complete their comparisons, and they did not have access to their tools or methods normally used forensic facial comparisons.

Given their role in the Justice system, it is imperative to understand how the accuracy of examiners and reviewers is impacted by cross-race forensic comparisons. Since Phillips *et al*

(2018) showed that super-recognizers perform equally to examiners and reviewers with face comparisons, measuring the impact of cross-race comparisons on super-recognizers will help us better understand the face processing system of forensic facial examiners and reviewers.

Additional Background if needed:

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

Phillips, P.J., Yates, A.N., Hu, Y., Hahn, C.A., Noyes, E., Jackson, K., Cavazos, J.G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J.-C., Castillo, C.D., Chellappa, R., White, D., & O'Toole, A.J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences, 115*(24):6171–6176. DOI: 10.1073/pnas.1721355115

Yates, A.N., Cavazos, J.G., Jeckeln, G., Hu, Y., Noyes, E., Hahn, C.A., O'Toole, A.J., & Phillips, P.J. (2023). Who Is That? Perceptual Expertise on Other-Race Face Comparisons, Disguised Face Comparisons, and Face Memory. *PsyArXiv Preprints*. DOI: 10.31234/osf.io/s87na (preprint)

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

The survey tools were developed based on previous work in Phillips et al (2018) and Yates et al (2023) as well as in consultation with experts in facial forensics. The procedures are the same as those in Phillips et al (2018). The Eligibility and Demographic Questionnaire was developed from multiple sources. The eligibility questions (#1-5) were based off the ones used in Phillips et al (2018). The questions on familiarity with Black and White individuals (#8-11) were based off standard background questions used in psychology literature on cross-race face comparisons. The demographic questions (#6-7, 12-13) are standard for reporting population statistics. Questions #1-5 are needed to determine if a participant is eligible to take part in the study and, if so, which category (i.e., examiners, reviewers, super-recognizers, or control) the participant is in. Question #7 is needed to perform analysis between participant races and image subjects. All other questions are optional in the questionnaire.

As in Phillips et al (2018), 20 pairs of images will be given to participants, who will have 3 months to complete their comparisons and send their responses to NIST. They will send their comparisons to NIST using the Comparison Questionnaire, which will be hosted on Survey Monkey. The questionnaire utilizes the same questions and response scales as in Phillips et al (2018), which was developed in consultation with the forensic facial community.

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.

Participants will come from forensic facial and super-recognizer communities. Recruitment emails asking for potential participants will go to professional groups such as the Facial Identification Scientific Working Group (FISWIG), the Facial Identification subcommittee of the Organization of Scientific Area Committees (OSAC), and the European Network of Forensic Science Institutes (ENFSI). Recruitment emails will also go to super-recognizer researchers who have tested and validated groups of super-recognizers.

Potential participants will have an interview over a phone or video call with a NIST researcher to go over the study and information sheet. Volunteers who choose to participate will proceed to the screening questionnaire. Participants will be verbally asked the Eligibility and Demographic Questionnaire by the NIST researcher. The questions determine if the participant fits into one of the tested groups, determine the analysis performed based on race, and are standard demographic questions for reporting in psychology studies.

After it is determined if the participant fits into one of the four groups, the participant will be sent a secure link to download the images. Participants will have up to 3 months to complete their comparisons and select their responses for all image pairs. To submit their responses, NIST will also send a link to a web-based survey (hosted on Survey Monkey) for completing the Comparison Questionnaire. The attached questionnaire illustrates the questions and presentation. In order not to release the images before the study, the attached Comparison Questionnaire only has pages for two mock-up pairs that are not included in the test. The final questionnaire will contain 40 questions for 20 pairs of images; the questions for each pair will be the same, demonstrated in the attached mock-up.

Participants will have the option to request their responses. Scores will be de-identified with an alphanumeric study code. The participant's name and email will be connected to their study code while data is being collected and analyzed so that participants may request their responses. Concurrent with the publishing of the first paper, the link between names and study code will be destroyed, and participants will not be able to request their responses afterward.

We estimate that we will receive up to 300 responses across all the four groups.

- A *forensic facial examiner* is a person who performs facial comparisons as part of their job and has been trained to perform forensic facial comparisons. Examiners perform one-to-one comparisons as part of their job duties.
- A *forensic facial reviewer* is a person who performs facial comparisons as part of their job and has been trained to perform forensic facial comparisons. Reviewers do not perform one-to-one comparisons and instead perform other comparisons, such as one-to-many.

- A *super-recognizer* is a person who has been acknowledged as having superior performance in face recognition compared to the general population. This acknowledgement may be established either by having taken a psychology test confirming them or by having been recruited to work using their skill as a face super recognizer.
- A person who has no forensic training and is not a super-recognizer will belong to the *control group*. These participants will be a proxy for the general public.

All potential participants are expected to complete the Eligibility and Demographic Questionnaire.

personally identifiable information (PII) will be collected for administering the study. The respondents' PII will be protected in the following manner. Each subject enrolled in this study will be assigned a study number and this number will identify all data, including results of facial comparisons. A master list linking the PII to the study number will be kept on an encrypted disk located in a locked office or a locked cabinet at NIST. Responses to the edibility and demographic questionnaire will be treated as PII and will not be shared outside of NIST. The exception being the race of the participant, which will be shared in order to replicate cross-race analysis. Personal privacy information would be withheld from release to the public under the Freedom of Information Act, 5 U.S.C. 552, exemption (b)(6): personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

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.

Using methods from signal detection theory, analysis of variance, and correlation theory, the groups will be compared across and within groups.