Research Proposal:
Detecting Trust: Visual Search for Social Dimensions in Faces

Harini Sankar
Department of Psychology, University of Illinois Urbana-Champaign

PSYC_496/BCOG_200

Dr. Stefan Uddenberg

May 19, 2025

Detecting Trust: Visual Search for Social Dimensions in Faces

People make numerous instantaneous judgments about a person based on their faces. These judgments can be objective (e.g., a person's race or gender) or subjective (e.g., how trustworthy or smart). These judgments often have real-world consequences, such as determining electoral success, criminal sentencing, and even hiring decisions (Blair et al., 2004; Rule & Ambady, 2008; Todorov et al., 2005).

Trustworthiness is a well-established social dimension that people perceive and evaluate others on. Oosterhof & Todorov (2008) showed that valence (a cue that signals whether one should approach a person or not) and dominance (a cue that signals the physical strength of a person) are the two main dimensions that underlie social traits such as trustworthiness. They built a model of facial trustworthiness and dominance that can generate faces that vary along these traits. More recently, Peterson et al. (2022) have improved upon that model and have created a model that can generate highly realistic faces that vary on a wide variety of both subjective and objective traits, as well as predict the ratings of given faces on these traits.

It is established that visual search can be performed on dimensions such as race and emotion, which are rather objective and visually discriminable characteristics of faces (Levin, 2000; Otten, 2016; Sun et al., 2013). However, race and emotion are just two dimensions or attributes that one can infer from other's faces during everyday interaction. For instance, it would be ideal to find someone friendly and outgoing to talk to at a party you were invited to, whereas looking out for faces that look untrustworthy at Times Square is a good way to remain vigilant to keep your personal belongings safe. Ostensibly, humans are able to search across social dimensions that are more subjective, but to our knowledge, it has not been empirically investigated.

Earlier visual search experiments using face stimuli have shown that search can be performed over more objective traits such as race and gender. (Levin, 2000) conducted a visual search study to understand the persistence of cross-race recognition. The cross-race recognition deficit is the phenomenon where individuals are less accurate in recognizing and discriminating between members of a different race than their own. He demonstrated that individuals who show a cross-race recognition deficit during discrimination tasks did not have slower search times for cross-race faces and were accurate in discriminating among cross-race faces within a visual search paradigm.

Search across emotion and its interaction with race have also been studied. Otten (2016) used Black and White faces that displayed either a frightened or angry emotion and demonstrated that Black faces were detected more quickly than White faces, and angry faces were detected quicker than frightened faces, and that Black angry faces were found quicker than their White counterpart. To further probe if this finding was due to implicit associations between race and negative emotions, Otten paired the search task with an IAT measure involving images of Black and White faces and words that were either positive or negative. He found that individuals who scored higher on the IAT measure, indicating a stronger anti-Black/pro-White bias, had less steep search slopes (i.e. Black targets captured attention so much so that increasing the display size of search did not influence search times).

Social identities, such as artificially induced ingroup and outgroup identities, can capture attention (Brosch & Van Bavel, 2012). The researchers briefly trained participants on faces that they arbitrarily assigned to an ingroup or outgroup (same or different college affiliation as the participant) by presenting them in different colored backgrounds. Then they presented the learned faces randomly as one of the cues in a dot probe task, along with a neutral face as the second cue. Immediately after this cue presentation, they presented the dot probe at one of the two locations and showed that the reaction time to the probe was smaller when the target was preceded by an outgroup face cue. They subsequently demonstrated that this effect generalized to new, unfamiliar outgroup face images shown during the task. This demonstrates that social information extracted from faces can influence attention.

Chwe & Freeman (2024) demonstrated that trustworthiness judgments can also be made at a group level. Participants shown a display of multiple faces were able to determine the trustworthiness of the ensemble, even when flashed for half a second. Additionally, the participants' ratings of the trustworthiness of ensembles were more accurate than the ratings for the individual faces within the ensemble, indicating that participants were more sensitive to group-level trustworthiness ratings than individual faces. Similarly, other social dimensions such as attractiveness have been shown to be extracted from ensemble displays of faces (Luo & Zhou, 2018).

Taken together, this indicates that subjective information such as social characteristics can be extracted rapidly from not just individual faces, but also in group settings. Information such as learned ingroup/outgroup identities, which cannot be disambiguated based on visual information, like color or contrast, can influence attention. Therefore, it is possible that trustworthiness is a dimension that can be searched on in a visual search paradigm.

In this study, participants will be shown a display of artificially synthesized faces that vary on whether they appear trustworthy or untrustworthy and are required to search for a trustworthy target amongst untrustworthy distractor faces or vice versa. If trustworthiness is a dimension that can be searched across, participants should be able to accurately identify the target. It would be interesting to see whether search asymmetries emerge when the target is a trustworthy face versus an untrustworthy one, as it is unclear if observers can classify trustworthiness as a feature that is either present or absent in a very explicit manner.

Method

Participants

As this study will be run in person, the subjects will be recruited from the SONA participant pool. They will be required to have normal or corrected-to-normal vision. Typically, visual search experiments recruit a minimum of 30–50 participants.

Stimuli and Experiment Setup

In each trial, subjects will view a display of faces. Each face is created by averaging 10 randomly selected faces from the Chicago Face Database (Ma et al., 2015) after being encoded into the GAN model (Peterson et al., 2022) . Random noise is added to their identities so each face looks a little different from one another. Twenty unique faces were created for this experiment. Currently, only white male faces are used. In the final iteration of the project, faces from other available races and both sexes from the CFD will be used. The model created by Griffiths et al., which contains the latent space distribution corresponding to a variety of social dimensions, is used to transform the averaged face to vary in how trustworthy they appear. The model is set to control for factors such as age, smiling, and gender. For each identity, both trustworthy and untrustworthy versions were generated. The background of all the images was removed using a remove background tool library (Gatis, 2025) that automatically detects the edges of the face and removes the background. From this set of images, the target and distractor faces are randomly selected for each trial display.

The face marker locations for each face are obtained using the dlib Python library. All 68 face marker coordinates, along with the valence (here, I use valence as the term to describe whether the transformed face is trustworthy (+1), untrustworthy (-1), or neutral (0); this is different from the way that valence is defined in Oosterhof and Todorov, 2008), are stored in a CSV file (stim_data.csv). The dlib-generated pixel coordinates are transformed into degrees of visual angle within PsychoPy.

The face marker locations that correspond to the mid-ear location on both sides of the face will be used to determine the position of the target cross in the final search display. Once the random set of images is selected for the experiment, the program places the target cross a distance of 0.5 degrees outward from the selected markers (markers 2 and 15) so that the cross appears at a consistent distance from each face across the display. Only one marker appears next to each face in the display, and the location (left or right side of the face) is randomized.

The target valence is randomized across blocks, and the display set size is randomized within each block and can consist of either 1, 3, 6, or 12 images. Each participant will complete a total of 240 trials (30 trials per set size (4) per target valence (2)). Within the current experiment files, there are a total of 32 trials, for the sake of brevity.

The experiment will be created and run using PsychoPy (v2024.2.4)(Peirce et al., 2019). PsychoPy has built-in functions that allow for precise records of reaction time measures and has a graphical user interface setup that is optimized for psychophysical experiments.

One concern with this experiment is how the subjectivity of trustworthiness judgments might influence the accuracy and reaction time measures. It is possible that in any given trial, the target trustworthy or untrustworthy face is not the face that the participant might perceive or rate as being the most trustworthy, and therefore there is a possibility of erroneously responding to an untrustworthy face that they perceive as the trustworthy target. A proposed variant of this task to work around this subjectivity issue is to set the target in each block to be one fixed face and instruct participants to just find this set target instead of instructing them to find the trustworthy or untrustworthy target.

Procedure

After completing the standard consent procedures, participants are presented with instructions before starting the task. Their task is to search for the target face that is either trustworthy or untrustworthy amongst the distractors that are of opposite valence (e.g., if the target is trustworthy, the distractors are untrustworthy). If the alternate version of the experiment is run, the instructions will contain the target face that they will be searching for, and the target will remain constant across the whole block. Once they locate the target face within the display, they are required to respond to the location of the blue target cross, which is present either on the right or left side of the face. They should press the right arrow key on the keyboard if the blue target is on the right, and the left arrow key if the blue target is on the left. They will be given short breaks frequently throughout the experiment.

Predicted Results

The accuracy in finding the target will determine whether trustworthiness is a dimension that can be searched for. A two-way repeated-measures ANOVA should be conducted on the reaction time data across the different set sizes (1,3,6, and 12) and target conditions (trustworthy and untrustworthy). If reaction times for finding the trustworthy target are consistently higher across all set sizes compared to the untrustworthy target, this would indicate a search asymmetry—suggesting that trustworthiness is perceived as a dimension that is either present or absent.

If the alternate proposed version is run, we would examine the reaction time differences between target conditions. If trustworthy targets among untrustworthy distractors take longer to find than untrustworthy targets among trustworthy distractors—or vice versa—it could indicate that the trustworthiness of faces in the display influences search times. This would suggest that trustworthiness functions as a categorical feature in visual search, where the presence of a trustworthy face is detected more readily than an untrustworthy face (or vice versa), indicating a search asymmetry.

Discussion

If participants can accurately search for the target, this would indicate that trustworthiness is a dimension that can be used as a feature in visual search. Furthermore, this would provide the first evidence that subjective social dimensions inferred from faces can be used in a visual search paradigm.

As mentioned earlier, one challenge encountered while designing this study is the highly subjective nature of trustworthiness judgments. Additionally, the current model used to manipulate facial trustworthiness is not able to keep identity perfectly constant; it inadvertently introduces subtle changes in facial identity, which may add low-level visual information that could influence search behavior.

A potential solution to this issue is to create custom stimuli for each participant based on their own trustworthiness ratings, collected in a face-rating task conducted prior to the visual search task. The faces displayed during the search would then be manipulated in real time using new latent values derived from participants' ratings. This approach would ensure that each participant sees a target that is, according to them, the most or least trustworthy face in the set. If successful, this method would not only allow for more reliable conclusions about visual search using the trustworthiness dimension but also help us understand individual differences in visual search patterns. It could further extend this paradigm to other subjective social attributes inferred from faces.

In future studies, it would be valuable to explore how race and gender (of both the stimuli and the participant) interact with trustworthiness in visual search. This would help us better understand how visual search and social perception operate in more ecologically valid, real-world settings.

References

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*(10), 674–679. https://doi.org/10.1111/j.0956-7976.2004.00739.x

Brosch, T., & Van Bavel, J. J. (2012). The flexibility of emotional attention: Accessible social identities guide rapid attentional orienting. *Cognition*, *125*(2), 309–316. https://doi.org/10.1016/j.cognition.2012.07.007

Chwe, J. A. H., & Freeman, J. B. (2024). Trustworthiness of Crowds Is Gleaned in Half a Second. *Social Psychological and Personality Science*, *15*(3), 351–359. https://doi.org/10.1177/19485506231164703

Gatis, D. (2025). *Rembg* (Version 2.0.66) [Computer software]. 'https://github.com/danielgatis/rembg'

Levin, D. T. (2000). Race as a Visual Feature: Using Visual Search and Perceptual Discrimination Tasks to Understand Face Categories and the Cross-Race Recognition Deficit. *Journal of Experimental Psychology: General*, *129*(4), 559–574. https://doi.org/10.10371/0096-3445.129.4.559

Luo, A. X., & Zhou, G. (2018). Ensemble perception of facial attractiveness. *Journal of Vision*, *18*(8), 7. https://doi.org/10.1167/18.8.7

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Otten, M. (2016). Race Guides Attention in Visual Search. *PLOS ONE*, *11*(2), e0149158. https://doi.org/10.1371/journal.pone.0149158

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peterson, J., Griffiths, T., Uddenberg, S., Todorov, A., & Suchow, J. W. (2022). Deep models of superficial face judgments. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4041458

Rule, N. O., & Ambady, N. (2008). The Face of Success: Inferences From Chief Executive Officers' Appearance Predict Company Profits. *Psychological Science*, *19*(2), 109–111. https://doi.org/10.1111/j.1467-9280.2008.02054.x

Sun, G., Song, L., Bentin, S., Yang, Y., & Zhao, L. (2013). Visual search for faces by race: A cross-race study. *Vision Research*, *89*, 39–46. https://doi.org/10.1016/j.visres.2013.07.001

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of Competence from Faces Predict Election Outcomes. *Science*, *308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589