Lowe Institute of Political Economy: Final Report

Visual Scanning Patterns in the Cross-Race Effect: Implications for Political Economy

Tomoe Kanaya   Anish Awasthi (CMC ‘28)

Abstract

The cross-race effect (CRE), a well-documented phenomenon where individuals are less accurate at remembering faces of races different from their own, has been implicated in wrongful convictions, which have significant social, political, and economic consequences. Eye-tracking research provides a powerful tool for uncovering the cognitive mechanisms underlying CRE by measuring where and how long individuals fixate and gaze on different areas of the face. Based on previous findings from expertise research, we expect to see subtle but significant differences in visual scanning patterns between same-race faces compared to other-race faces which we define as Cross Race Visual Scanning Patterns (CR-VSP). Using support vector machine (SVM) analyses, results revealed CR-VSP in fixation and gaze data when using nonlinear models, but not when using linear models. In addition, CR-VSP was more accurately captured using spatial binning approaches compared to K-Means clustering. These findings suggest that visual, perceptual encoding differences contribute to the CRE and highlight the importance of spatial granularity when analyzing face-scanning patterns. Recommendations for future research are provided.

Acknowledgments

The authors gratefully acknowledge Dr. Benjamin U. Marsh at the University of Tampa for providing the data used in these analyses.

Funding

This report was funded through the generous support of the Lowe Institute of Political Economy at Claremont McKenna College. The views expressed are those of the authors and do not represent the position of the Lowe Institute or Claremont McKenna College.

Introduction

The Cross-Race Effect

The social, political and economic impacts of wrongful convictions are substantial; since 1989, state compensations to exonerated victims have exceeded four billion dollars (National Registry of Exonerations, 2024), not including the cost of their trials or incarcerations. Victims are removed from the workforce and experience stigma and prejudices that make it difficult to re-enter the labor market after their release (Grounds, 2004). Wrongful convictions also undermine the public's trust in the judicial system and the political institutions that support it, which can carry a wide range of economic consequences including increased crime rates and reduced civic engagement and property values (Clow & Leach, 2015). Therefore, understanding the nature of eyewitness memory, and determining what factors lead to accurate eyewitness testimony, is a critical issue within the intersection of psychology, politics, and economics.

The cross-race effect (CRE), a well-documented phenomenon where individuals are less accurate at remembering faces of races different from their own, has been implicated in eyewitness testimony that leads to wrongful convictions (Meissner & Brigham, 2001). To date, an overwhelming majority of CRE research has focused on Black and White participants and faces, and significantly less is known about CRE in other races. To address this gap, Marsh (2021) has conducted CRE studies on Latino, Asian, White and Black participants and faces. The Asian participants in his sample did not display CRE, and the Latinos experienced a limited CRE. These findings underscore the importance of including individuals from multiple races and ethnicities in CRE research.

While CRE has been well documented, the underlying mechanisms remain unclear. One potential explanation is that individuals develop greater expertise with same-race faces compared to other-race faces due to more frequent exposure to individuals of their race throughout their daily lives (e.g., Sporer, 2001). This could lead to differences in eye tracking when looking at same-race faces compared to other-race faces, which we have labeled cross-race visual scanning patterns (CR-VSP).

Eye-tracking Data and Analyses

Eye-tracking tools provide useful data on the perceptual processes underlying face scanning patterns by recording the exact location on a face that observers look at, allowing researchers to infer which parts of the face are being processed. Gazes and fixations are reliably used as measures of visual processing (e.g., Pauszek, 2023). Gazes refer to the instantaneous point of regard recorded by an eye tracker and represent quick, visual attention and processing. Fixations, on the other hand, are clusters of consecutive gaze points that remain within a limited spatial area for a minimum duration, representing more involved visual processing and higher cognitive load.

A majority of the eye-tracking research on face scanning patterns relies on comparing fixation counts across different, pre-defined, areas-of-interest (AOI), such as the eyes, nose and mouth (e.g., Blais et al., 2008) using t-tests or Analyses of Variance analyses (e.g., Marsh, 2021) to detect significant differences between AOIs. While this approach provides valuable information, it may obscure subtle and meaningful differences in visual processing. For example, experts looking at same-race faces may look at different areas *within* an AOI, requiring analyses at a more granular, pixel level. Furthermore, experts may be able to remember same-race faces through short gazes, compared to novices who may require longer fixations to process different-race faces. Detecting such differences, however, requires analyses of high dimensional and high volume data.

Analyzing eye-tracking data at this level of granularity is computationally taxing and requires machine learning techniques such as Support Vector Machines (SVM). SVM is a class of supervised learning algorithms used for classification and regression. They identify the "optimal hyperplane" that separates different classes of data with the maximum possible margin. SVMs maximize the margin between the two classes, leading to a more robust model that is less prone to overfitting and learning the noise instead of the actual patterns. As a result, the model performs better at classifying new, unseen data. SVMs can be implemented with different kernel functions, allowing researchers to test whether underlying patterns are linearly separable or require more complex modeling, such as the radial basis function (RBF), which allows for more flexible and localized decision boundaries (Hastie et al., 2009).

K-Means clustering is a commonly used technique used with SVM. K-Means creates groups for gaze or fixation points into a fixed number of clusters based on spatial proximity, producing a lower-dimensional representation for classification. Using this approach, Liu et al. (2016) found that children diagnosed with Autism Spectrum Disorder (ASD) have different face-scanning fixation patterns compared to their neuro-typical counterparts with a RBF kernel. The K-Means clusters, however, can still represent large areas of space and, therefore, would not detect more fine-grained differences at the pixel level. To address this, Zhao et al. (2023) examined gaze patterns by dividing the stimulus screen into a fixed grid and assigned each gaze to a bin based on the specific pixel location of that gaze. A histogram of the bins were calculated and used in the SVM to compare face-scanning patterns between ASD and non-ASD participants.

In addition to their important clinical implications, these studies also serve as conceptual and methodological foundations for the current study. Specifically, we analyzed eye-tracking data on face-scanning patterns of participants and faces from multiple races to determine if CR-VSP can be an underlying contributor to CRE. We utilized multiple machine learning approaches to identify group differences including SVM with K-Means clustering and spatial bin histograms on gazes and fixations using linear and nonlinear models.

Participants and Procedure

The experiment was advertised to participants as "memory for diverse faces" at the University of Tampa. A total of 205 participants (Female = 122; Male = 83) completed the experiment. The majority of participants were racially/ethnically White (n = 155) with a smaller portion being Black (n = 22), Latino (n = 22), and Asian (n = 6).

Participants were seated at a desktop computer with a 22 inch screen equipped with Tobii X2-60 Eye Tracker. The experiment was designed and conducted using iMotions biometric software. Participants first performed an eye tracking calibration phase wherein they visually followed a dot as it moved to five spots on the screen. Afterwards, the screen changed immediately to the instructions for the experiment.

Ninety-six faces were selected from the Chicago Face Database (Ma et al., 2015). An equal number of male and female faces were included across four racial groups. The faces were divided into two sets of 48 faces, each containing six faces for each of the twelve race x sex combinations. During the study phase, participants were presented with the faces from one set and were instructed to study each face carefully, as their memory would be tested later. Each face was presented individually in random order for six seconds.

In the subsequent test phase, participants completed a recognition memory task in which the previously presented faces were intermixed with the other set, resulting in all 96 faces presented in random order. Each face remained on the screen until the participant indicated whether they had seen the face during the study phase. Furthermore, the face was displayed on the left-hand side of the screen and the memory questions on the right-hand side of the screen. After the test phase, participants took a demographic questionnaire.

Participants whose eye tracker calibrations were labeled as “poor” by the software were excluded from the dataset. The Asian participants and Asian faces were also removed from the current analyses due to findings by Marsh (2021), where Asians participants did not display CRE. To reduce potential sex driven differences and due to their small sample, data collected from male participants and male faces were also excluded from the present analyses. Finally, a discrepancy was discovered between the demographic information provided by some of the participants through an online survey and the ID numbers used in their eye-tracking data that could not be resolved within the timeline of the project. Therefore, we analyzed the eye-tracking data from 114 females (76 White, 18 Black and 20 Latina) participants on 18 Black, White and Latina female faces during the study phase.

Analyses

The overall analytical approach examined whether participants exhibited CR-VSP by transforming gaze and fixation data into two distinct feature sets: (1) K-Means clustering and (2) spatial binning. Elbow plots, which identify the point at which increases in *k* produce diminishing reductions in within-cluster variance, indicated that a five-cluster solution was optimal for both gaze and fixation data (see Figures 1 and 2). In addition to the clustering approach, spatial bins were created by dividing the portion of the screen displaying the face stimulus into a 54 × 54 grid, with each bin representing a 20 × 20 pixel region. Each gaze and fixation point was then assigned to the corresponding cluster or spatial bin based on its screen coordinates. (See Figures 3 and 4 for histograms of spatial bins of gaze and fixation bins across all participants and faces.)

Supervised classification was performed on the K-Means cluster and spatial bins using SVMs with linear and RBF kernels. Model evaluation was conducted using grouped five-fold cross-validation, with grouping defined by participant to prevent participant-level leakage (Hastie et al., 2009). All models included feature standardization, such that each feature was rescaled based only on the training data within each cross-validation fold, resulting scaling parameters were applied to the corresponding test data. This procedure ensured that information from the test set did not influence preprocessing, preventing feature-level data leakage (Kaufman et al., 2012). Finally, permutation testing using 500 permutations was conducted to assess the statistical significance for all models (Hastie et al., 2009).

Results

For the gaze data, K-Means cluster features yielded a linear SVM mean accuracy of 45.79% (± 4.87%, p = *n.s.*), while the RBF SVM achieved a mean accuracy of 51.57% (± 1.48%, p = *n.s.*). Using spatial bin features, the linear SVM produced a mean accuracy of 51.76% (± 1.52%, p = *n.s.*), whereas the RBF SVM improved to 64.82% (± 1.19%, $p < .05$).

For the fixation data, K-Means cluster features resulted in a linear SVM mean accuracy of 53.92% (± 7.16%, p = *n.s.*) and an RBF SVM mean accuracy of 53.96% (± 5.70%, p = *n.s.*).

In contrast, spatial bin features yielded a linear SVM mean accuracy of 50.26% (± 1.42%, p = *n.s.*), while the RBF SVM achieved a mean accuracy of 67.09% (± 3.71%, $p < .01$).

Discussion

The results show that spatial binning combined with an RBF kernel produced the strongest and only significant classification performance for CR-VSP with gazes and fixations. Overall, the findings indicate that classification performance depends primarily on spatial granularity and nonlinear decision boundaries. These findings have important implications for CRE. More specifically, they suggest that CRE may, in part, be driven by differences in early visual sampling strategies when first viewing a face. The stronger performance of the spatial bins indicates that these scanning differences are localized rather than broadly distributed regions. These results are consistent with expertise research, which suggest that individuals develop more efficient and specialized processing strategies for faces they encounter frequently (Meissner & Brigham, 2001).

This study extends prior eye-tracking research by building on AOI analyses, which summarize fixations across meaningful facial regions such as the eyes, nose, and mouth (e.g., Blais et al., 2008). The results contribute to the growing literature using machine learning approaches to identify group differences with more spatial granularity and provide insight into the relative effectiveness of different feature representations (e.g., K-Means clustering vs. spatial binning), models (linear vs. RBF kernels), and data types (e.g., gaze vs. fixation measures) within this emerging field (Liu et al., 2016; Zhao et al., 2023). Finally, this study is among the few to apply these approaches to a non-clinical population.

Overall, the study established a strong analytical foundation through multiple complementary approaches and produced promising preliminary results and direction for future methodological development. In addition, we began developing spatial heatmaps as an exploratory visualization tool. Heatmaps are graphical representations that show the intensity or frequency of values across a spatial grid, allowing regions of elevated activity to be identified more easily than in tabular outputs (Wilkinson & Friendly, 2009). Early results were promising (see Figures 5 and 6) and suggested this approach has strong potential for future analyses.

Despite these strengths, several limitations should be noted. The study was exploratory, and several analytical choices were not utilized, including a different number of K-Means clusters, different pixel dimensions for the spatial bins, and the number of permutations used for significance testing. Future research should explore different parameter settings to see if model accuracy can be improved. Additional analyses should also include forward feature selection, as well as alternative performance metrics such as Receiver Operating Characteristic and Area Under the Curve, to provide a more comprehensive evaluation of classification performance (e.g., Liu et al., 2016).

Additionally, the analyses did not include data from the male and Asian participants and faces. Including these groups in the future will be important for understanding how CR-VSP may vary across gender and race, as well as possible interactions between these factors. Finally, we did not analyze data from the test phase, where participants completed the recognition task used to identify CRE. As a result, while this work uncovers race-related differences in face-scanning behavior, it does not directly examine how these differences relate to CRE performance.

This study provides an important step in uncovering CR-VSP and highlights their potential role in understanding the perceptual mechanisms underlying the CRE. Given the social, political, and economic significance of the cross-race effect (e.g., Winston, 2025), further research in this area has immediate implications for political economy. A better understanding of the mechanisms underlying CRE could inform improvements in areas such as eyewitness identification and other decision-making contexts where face recognition in multi-racial contexts play a critical role.

References

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022

Clow, K. A., & Leach, A.-M. (2015). After innocence: Perceptions of individuals who have been wrongfully convicted*.* *Legal and Criminological Psychology, 20*, 147–164. https://doi.org/10.1111/lcrp.12018

Grounds, A. (2004). Psychological consequences of wrongful conviction and imprisonment. *Canadian Journal of Criminology and Criminal Justice, 46*(2), 165–182. https://doi.org/10.3138/cjccj.46.2.165

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data, 6*(4), 1–21. https://doi.org/10.1145/2382577.2382579

Liu, W., Li, M., & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research, 9*(8), 888–898. https://doi.org/10.1002/aur.1615

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*, 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Marsh, B. U. (2021). The cost of racial salience on face memory: How the cross-race effect is moderated by racial ambiguity and the race of the perceiver and the perceived. *Journal of Applied Research in Memory and Cognition, 10*(1), 13–23. https://doi.org/10.1016/j.jarmac.2020.09.008

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

National Registry of Exonerations. (2024). *Annual report 2024*. https://www.law.umich.edu/special/exoneration/Documents/2024_Annual_Report.pdf

Pauszek, J. R. (2023). An introduction to eye tracking in human factors healthcare research and medical device testing, *Human Factors in Healthcare, 3*, 100031 https://doi.org/10.1016/j.hfh.2022.100031

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7*(1), 36–97. https://doi.org/10.1037/1076-8971.7.1.36

Wilkinson, L., & Friendly, M. (2009). *The history of the cluster heat map*. *The American Statistician, 63*(2), 179–184. https://doi.org/10.1198/tas.2009.0033

Winston, A. R. (2025). What are the financial & social costs of police misconduct? *Express Legal Funding*. https://expresslegalfunding.com/police-misconduct-costs/

Zhao, Z., Wei, J., Xing, J., Zhang, X., Qu, X., Hu, X., & Lu, J. (2023). Use of oculomotor behavior to classify children with autism and typical development: A novel implementation of the machine learning approach. *Journal of Autism and Developmental Disorders, 53*(3), 934–946. https://doi.org/10.1007/s10803-022-05685-x

Figure 1

Elbow Plot of K-Means Clusters for Gaze Data

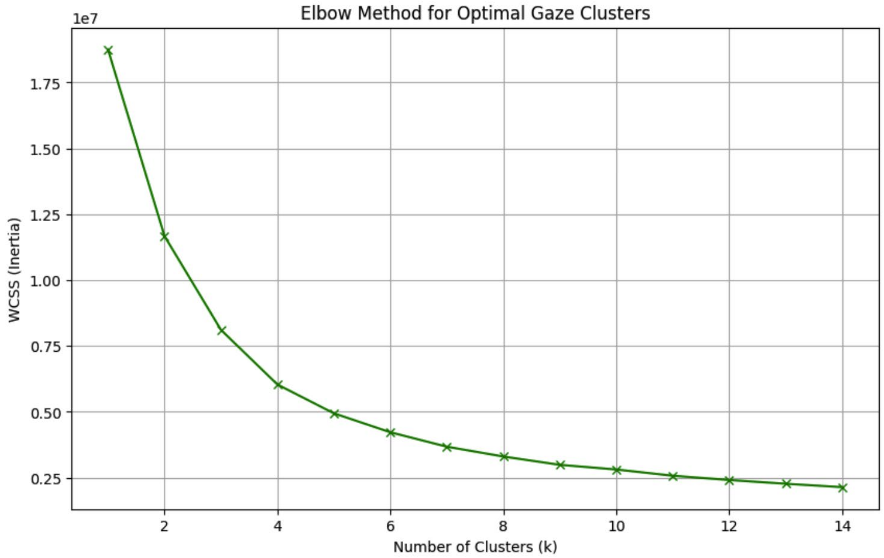

Figure 2

Elbow Plot of K-Means Clusters for Fixation Data

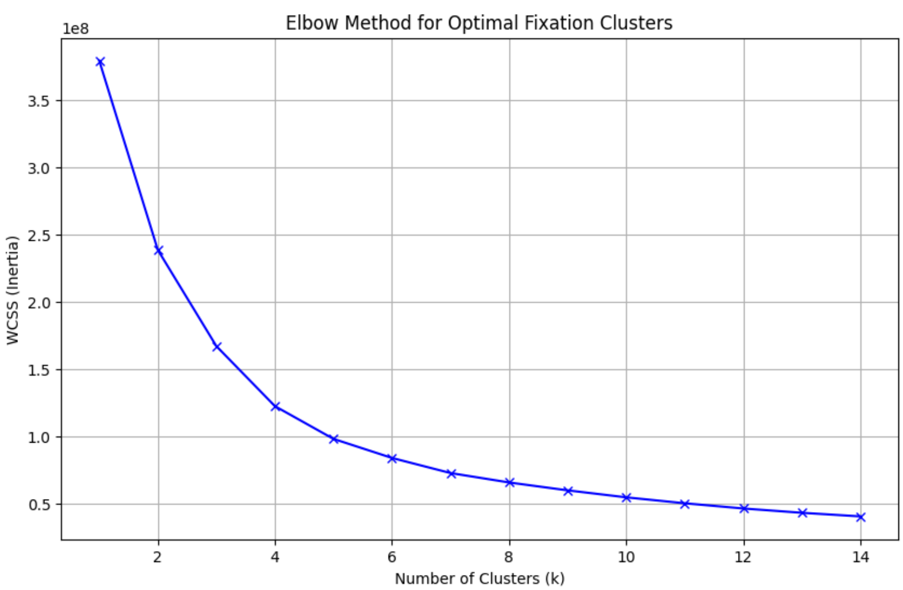

Figure 3

Histogram of Most Frequently Fixated Bins across all Participants and Faces

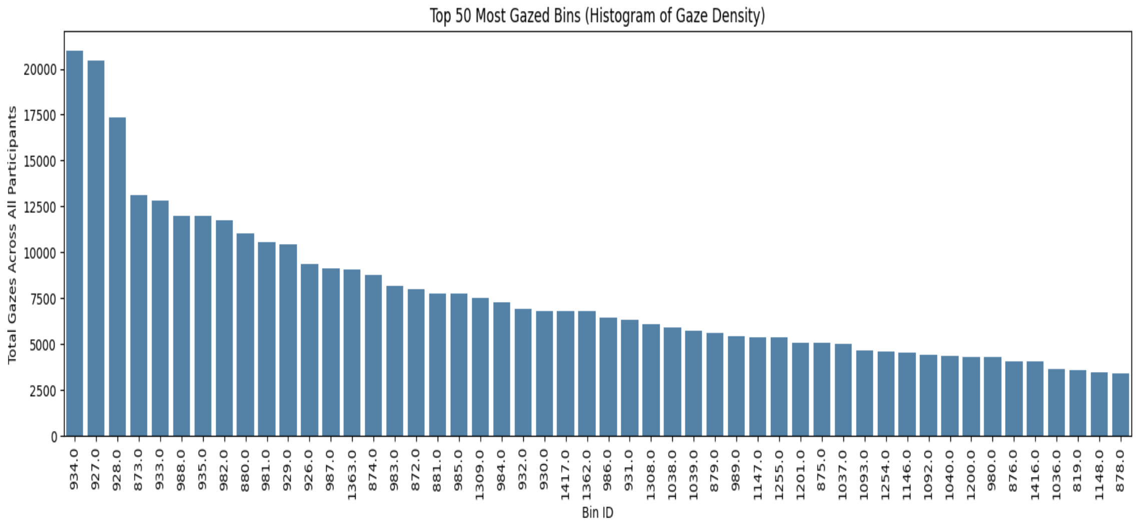

Figure 4

Histogram of Most Frequently Fixated Bins across all Participants and Faces

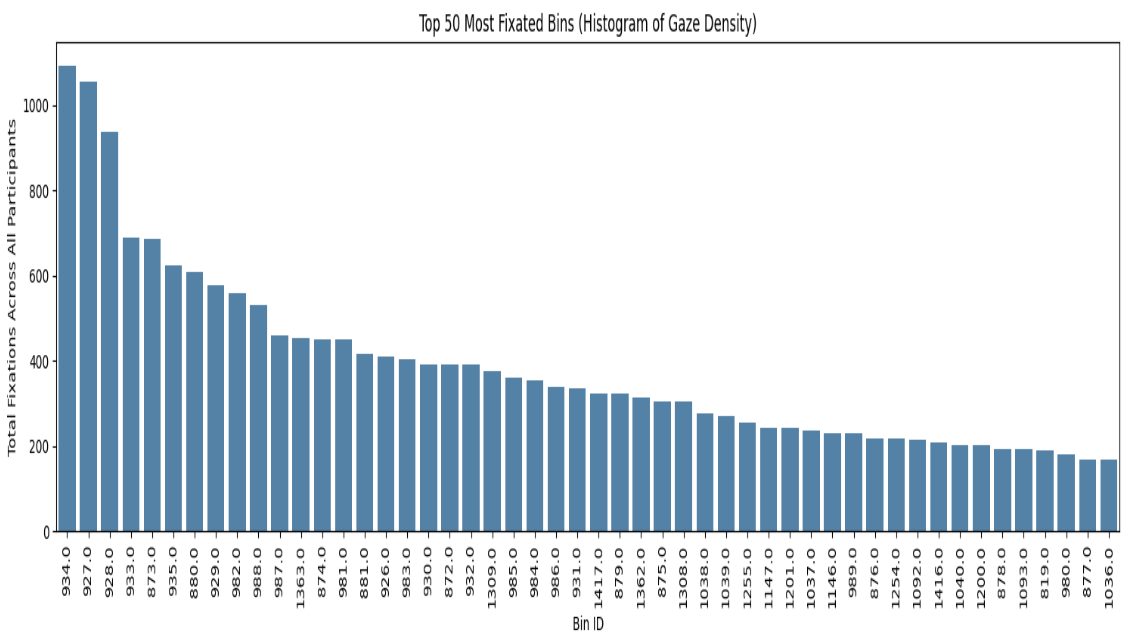

Figure 5

Preliminary Heatmaps of all Participants by Race of Face

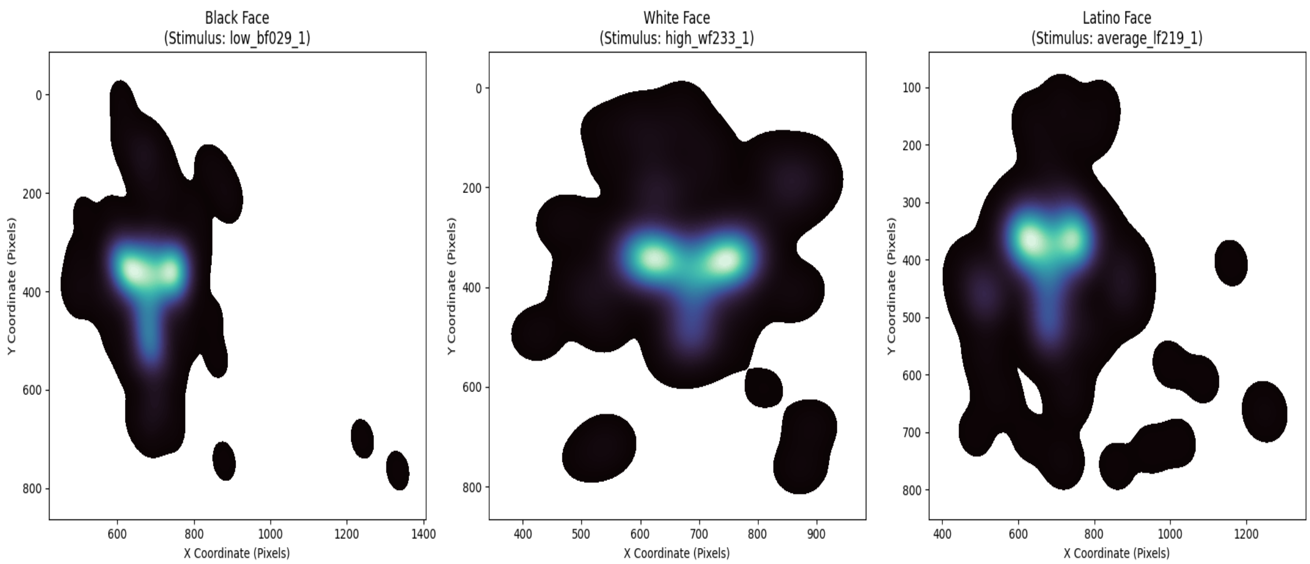

Figure 6

Preliminary Heatmaps of White Participants by Race of Face