APPENDIX A

*** Difference between the different perceptible ethnic groups ***

(a) Acceptance rate for each user
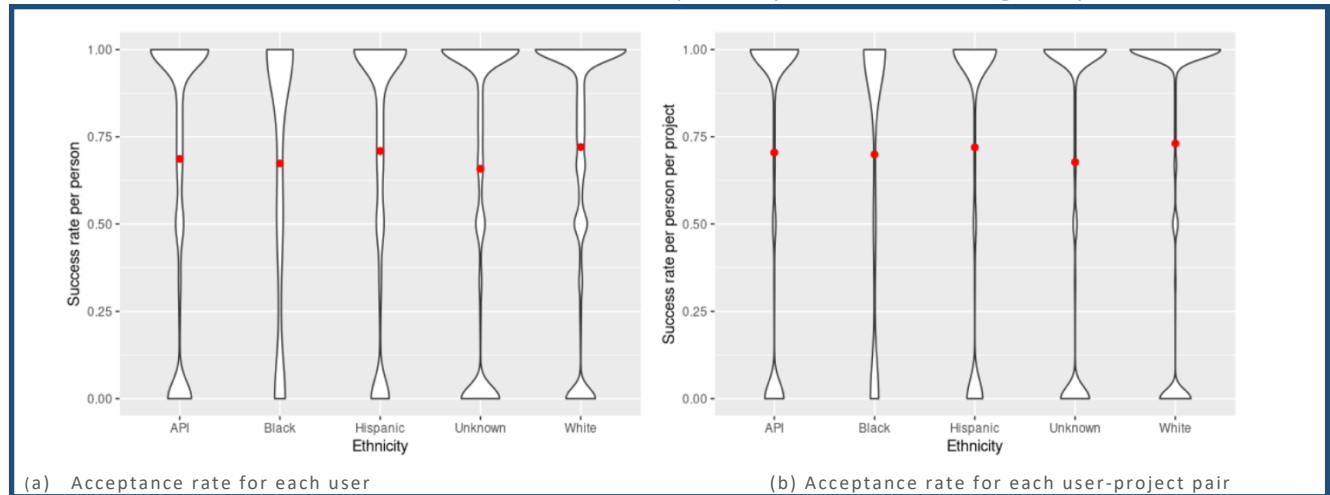
(b) Acceptance rate for each user-project pair

Figure Appendix.A.a: Acceptance plot. Vertical dashed lines show the averages for each group.

As we cannot use the results from Figure 5 to make sound conclusions because there are developers that can make a lot of pull requests and can have high acceptance rates, we removed the bias that these developers are introducing by calculating the acceptance rate for each developer in each perceptible ethnicity, individually. Figure Appendix.A.a shows the density of the number of developers in accordance with the acceptance rate for each perceptible ethnic group. The dashed vertical lines show the average acceptance rate. It is observable that developers with Unknown ethnicity have the least acceptance rate average (0.658). Developers perceptible as White have the highest acceptance rate average (0.720), followed by Hispanic (0.709), API (0.686) and Black (0.673).

We also removed any bias that might be introduced by developers in a project. For example, there might be cases of developers that are highly known and popular in a project, which might result in a high acceptance rate, but they are successful in other projects. Therefore, we analyzed the acceptance rate of each developer in each project, developer project pairs. Figure Appendix.A.b shows that developers perceptible as White have the highest average (0.73), followed by Hispanic (0.719), API (0.704), and Black (0.699). After removing the well-known developers within a project, we can see that the results are still consistent with the results shown in Figure 4a. Thus, developers without a perceptible ethnicity have the lowest acceptance rate average and developers perceptible as White have the highest acceptance rate average.

Finally, we run Kruskal-Wallis and Dunn tests to analyze whether the differences shown in Figure Appendix.1 are statistically significant. Kruskal-Wallis test results for the first dataset (acceptance rate per developer) ($H = 451.14$, $df = 4$, $P < 2.2e-16$) show that the means of the perceptible ethnic groups are not equal. Pairwise comparisons using Dunn's test indicated that the mean for perceptible API developers is statistically different than perceptible White developers ($P < 0.05$), perceptible Hispanic developers ($P < 0.05$), and perceptible Black developers ($P < 0.05$). The results for the second dataset (acceptance rate per developer-project pairs) are similar. In this case, we found that the mean for perceptible Hispanic developers is statistically different than perceptible White developers ($P < 0.05$). Thus, our pairwise results indicate that all perceptible ethnic groups are statistically significant different, except the perceptible Black group. Based on the results, the difference between averages is not the outcome of chance and developers perceptible as White have the highest average followed by Hispanic and API. When comparing Black to other groups we cannot draw any conclusion because of the lack of significant results.

*** Difference between the different perceptible ethnic groups ***

Now, we would like to analyze the distribution of developers and pull request acceptance rates, when considering the integrator's perceptible ethnicity.

(a) Acceptance rate density plot for perceptible White submitters

(c) Acceptance rate density plot for perceptible API submitters

(b) Acceptance rate density plot for perceptible Hispanic submitters

(D) Acceptance rate density plot for perceptible Black submitters (No statistically significant difference was observed)
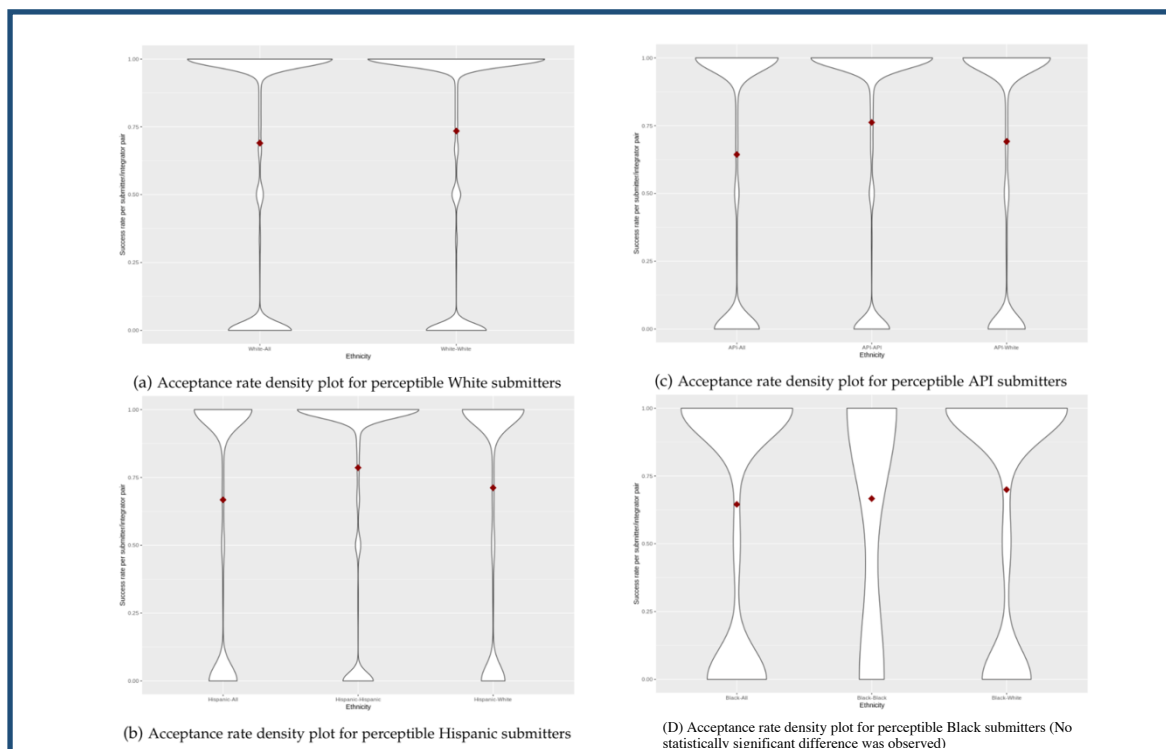
Figure Appendix.A.b: Acceptance plot. Vertical dashed lines show the averages for each group.

To assess the distribution of developers and pull request acceptance rates between different groups when considering the integrator's perceptible ethnicity, we analyzed the acceptance rate of each developer against each integrator. We plotted the distributions and the means for each pair, and we applied Kruskal-Wallis and Dunn test to measure the significance of the differences found between the distributions. We only considered pairs of developers without an Unknown perceptible ethnicity.

Figure Appendix.A.b shows the distribution of acceptance rate for each ethnicity pair. We can see in the Figure that submitters perceptible as White have an acceptance rate average of (0.734) when the integrator is also perceptible as White (but their acceptance rate average is (0.69) when measured against all integrators). The acceptance rate average of submitter-integrator pairs perceptible as API and Hispanic (when the submitter and integrator are perceptible to be in the same group) is 0.785 and 0.762, respectively. In addition, we found that this average is higher than the cases where the integrator is perceptible as a White developer. This average is 0.692 for API-White developer pairs and 0.711 for Hispanic-White developer pairs. All the comparisons are statistically significant, according to Kruskal-Wallis and Dunn test. Note that we found no statistically significant difference when making the comparisons for submitters with Black perceptible ethnicity because of the insufficient data points for perceived Black developers.

Thus, we found that acceptance rate average is statistically significant lower when the submitter is perceptible as Non-White (except Black developers) and the integrator is perceptible as White.