Prototypical Attendees

In this project, we generated a number of models to use to test our hypotheses. Our statistical analyses resulted in significant findings from our models that we discuss in our paper. However, we acknowledge the complexities involved in interpreting these analyses. Thus, we present examples of "prototypical attendees" in which we calculated the predicted likelihood of continued participation in computational activities given a specific set of model inputs.

In Figure 2 in the paper, looking at the effect of ethnicity, predicted curves are generated as follows:

1) The 10^{th} and 90^{th} percentiles are calculated for *ethnicity_pct* for the original dataset. The variable *ethnicity_pct* reflects the proportion of attendees that share the same identity as the index attendee (as opposed to the binary *notwhite* variable described later).
2) The values obtained above are used as endpoints for a sequence of 200 equally spaced points (resulting in our x-axis).
3) Prototypical attendees are given "group average" coding interest, represented by the variable *interest*, with the average calculated from the original data.
4) Half of the prototypical attendees are assigned to "White" and half to "Non White". This assignment is used to determine the value of their *notwhite* variable (0 = "White", 1 = "Not White").
5) Half of the "White" prototypical attendees are assigned to the "Presented" group. Similarly, half of the "Non White" prototypical attendees are assigned to the

"Presented" group. These assignments are used to determine the value of their *presented.new* variable.

6) All prototypical attendees are assigned to the same date/location event, represented by the *eventregion* variable.

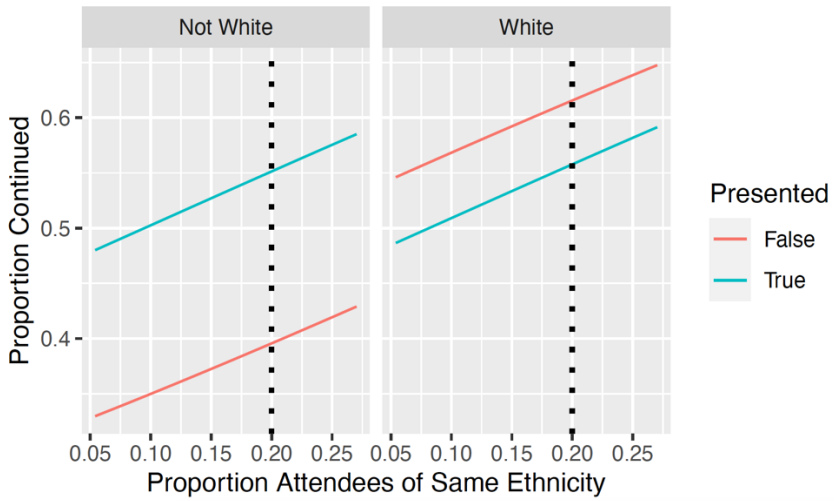
7) Curves are generated using our model for H3:

   *model.h3.eth <- glmer(continued ~ notwhite + presented.new + notwhite * presented.new + ethnicity_pct + interest + (1|eventregion), data = d, family=binomial("logit"))*

8) Our plot is faceted on Ethnicity, so 2 subplots are generated: subplot 1 (White), subplot 2 (Non White).

9) In each plot, 2 curves are displayed: blue curve (Presented), red curve (Did Not Present).

In this way, we are guaranteeing that the only differences between our prototypical attendees are the assigned inputs we have given them for *ethnicity_pct*, *notwhite* (either 0 ("White") or 1 ("Non White"), and *presented.new* (either 1 ("Presented") or 0 ("Did Not Present")).



10) The vertical line at x = 0.2 in each subplot looks at the case where the prototypical attendee attends an event where 20% of attendees share the same ethnicity as the index prototypical attendee – The “Non White” subplot encompasses all the options other than “White”.
11) When 20% of other attendees share their ethnicity, a Non White attendee who presents is predicted to have a close to 55% likelihood of continuing with coding activities after the event, as opposed to just under 40% had they not presented.