Task 1

Researchers were interested in knowing whether the modality in which a recommendation is delivered affects people's choices. In order to answer the question, participants were randomly assigned into one of the three groups according to the modality of recommendations and asked to make choices in every different product category after the recommendation. The choice of whether the participants followed the recommendation out of two options are measured as the dependent variable (encoded as 0 and 1). The three modalities of recommendations are an auditory recommendation, a permanent-visual recommendation which appeared word-by-word but each word remained on the screen until the choice is made, and a disappearing-visual condition which appeared individually by word and disappeared after the next word appeared. Product categories included brownie recipe, hotel, shampoo, and toaster oven. This mixed-design approach aimed to explore whether the preference for auditory recommendations over written recommendations could be replicated. Additionally, the study sought to investigate whether this preference could be further elucidated by considering other variables such as the detailed form of presentation of visual recommendation, or if the preference is specific to a particular product category.

Each modality is distributed with around 700 participants. Variables such as gender or age are also controlled or counterbalanced through different modalities. We then analyzed the dependent variable using two-way mixed design ANOVA, with the independent variables being the modalities of recommendations and the categories of product. The modalities of recommendations is a between-subject factor with three levels, while the categories of product is a within-subject factor with four levels. The results are also shown in graphical form in Figure 1.

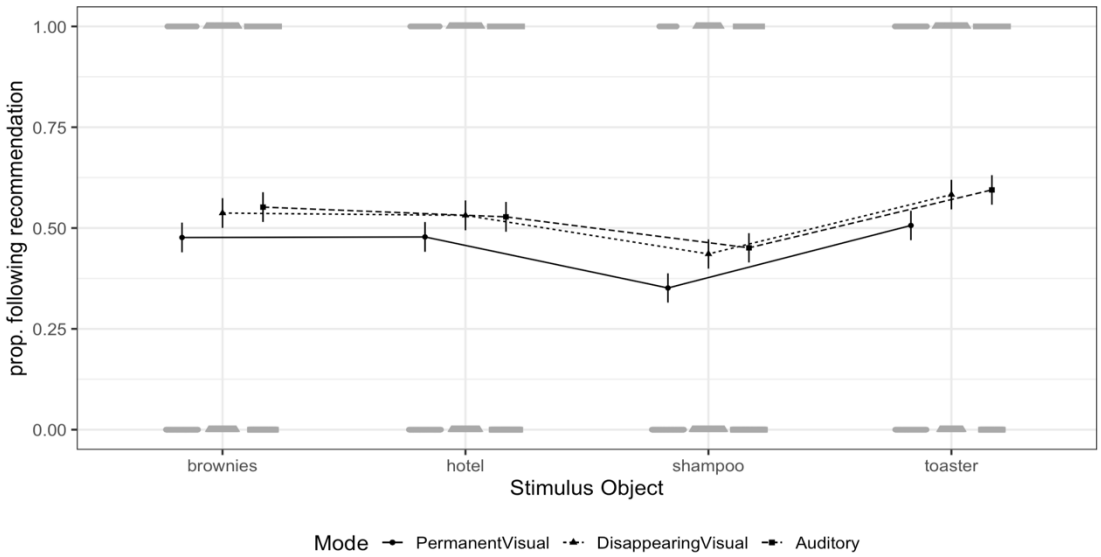

Figure 1: Probability of participants following recommendations in different modalities across different categories of stimulus object. Data points are shown as grey dots in the background with jitter. Black dots show the means of each group and the error bars show the associated 95% confidence intervals.

The ANOVA analysis for the dependent variable showed a significant effect of the modalities of recommendation, $F(2, 2103) = 11.50$, $p < .001$, as well as the categories of product, $F(2.98, 6261.05) = 46.88$, $p < .001$. The significant effect of the modalities of recommendation is within our expectation due to previous report of the preference of auditory recommendations over visual recommendations. On the other hand, the significant effect of the categories of product demonstrated that different products do have an effect on the probability of taking the recommendation or not. However, the modalities of recommendation by categories of product interaction was not significant, $F(5.95, 6261.05) = 0.48$, $p = .822$. This means the effect of the modalities of recommendation is general, instead of being dependent on particular categories of product.

As shown in Figure 1, the probability of following recommendation in the permanent-visual modality appears to be lower than the other two modalities. On top of that, the probability of following recommendation seems to fluctuate across different categories of product, with the shampoo category being visibly lower than other categories. To identify the specific recommendations and product categories that exhibited significant differences from others, we conducted follow-up contrasts for both independent variables.

For the contrasts of the modalities of recommendations, we compared each modality

in pairs as well as auditory modality with the mean of two visual modalities. We applied Bonferroni-Holm correction to control for false positivity during multiple testing across the four contrasts. In the pairwise contrasts of each modality, the permanent-visual vs. disappearing-visual contrast difference is 0.07 with $t(2103) = -3.86$, and $p < .001$; the permanent-visual vs. auditory contrast difference is 0.08 with $t(2103) = -4.40$, and $p < .001$; the disappearing-visual vs. auditory contrast difference is -0.01 with $t(2103) = -0.53$, and $p = .594$. In the auditory modality vs. the mean of visual modalities, the contrast difference is -0.04 with $t(2103) = -2.85$, and $p = .009$. This follow-up analysis demonstrated that the preference for auditory recommendation over visual recommendation is limited to only permanent-visual recommendation since the effect of disappearing-visual recommendation is similar to auditory recommendation. Moreover, the auditory modality vs. the mean of visual modalities contrast illustrates that previous report that indicated the preference for auditory recommendation over visual recommendation could be due to the lack of grouping or definition in visual recommendation. The results implicated that instead of auditory or visual modality being the variable that effects participants' choice, other variables such as the disappearing characteristic of modalities are the factors that really make an impact.

For the contrasts of the categories of products, we compared each category in pairs. We again applied Bonferroni-Holm correction to control for false positivity during multiple testing across the six contrasts. Only the brownies recipe-hotel contrast didn't show significant difference, with a difference of 0.01, $t(2103) = 0.72$, and $p = .473$. The other contrasts all showed significant difference, with the brownies recipe-shampoo contrast showing a difference of 0.11, $t(2103) = 8.76$, $p < .001$, brownies recipe-toaster contrast with a difference of -0.04, $t(2103) = -2.92$, $p = .007$, hotel-shampoo contrast with a difference of 0.10, $t(2103) = 7.54$, $p < .001$, hotel-toaster contrast with a difference of -0.05, $t(2103) = -3.75$, $p < .001$, and shampoo-toaster contrast with a difference of -0.15, $t(2103) = -11.53$, $p < .001$. The results demonstrated that participants are more likely to follow recommendations in the toaster category, and less likely to follow recommendations in the shampoo category. One explanation is that people tend to follow recommendations in product categories that involves higher levels of function such as machinery since people are likely to have less knowledge about it. On the other hand, recommendations about daily necessities might appear to be more commercial, and thus reduced the probability of people following the recommendations. However, further research is required to examine this hypothesis.

Overall, the preference for auditory recommendations over written recommendations can only be replicated when the written recommendation is in the form of permanent visual recommendation in this data. The disappearing visual recommendation, on the other hand, has a similar effect to the auditory recommendation. Furthermore, the category of recommended product also independently effects the probability of whether participants follow the recommendation.

Task 2

Stereotype threat happens when people feel themselves to be at risk of conforming to stereotypes about their social groups, which could negatively affect the performance of the stigmatized groups. Researchers were interested in knowing whether the effect of stereotype threat can be perceived in math performance. In this study, participants were randomly assigned into a stereotype-activation condition or a control condition, and then asked to take a math test. The stereotype-activation condition was executed through displaying a video that discussed gender gaps in math performance, while in the control condition a video about human memory is displayed. Additionally, participants were also asked to take a verbal test besides the math test to see if the stereotype threat effect can be extended to verbal performance.

The scores of both tests are the measured dependent variables, and the condition of whether activating stereotype as well as gender are the two considered independent variables. Both the condition as well as gender are between-subject factors with two levels. The age of each group is controlled, but the participants numbers of each group is slightly unbalanced, with the female participants being twice the amount of male participants in the stereotype condition, and the female participants being three times the amount of male participants in the control condition. We then perform two-way ANOVA analysis on both tests to examine if there is an effect of stereotype threat on tests, and if the effect depends on the gender. The results are also shown in graphical form in Figure 1.

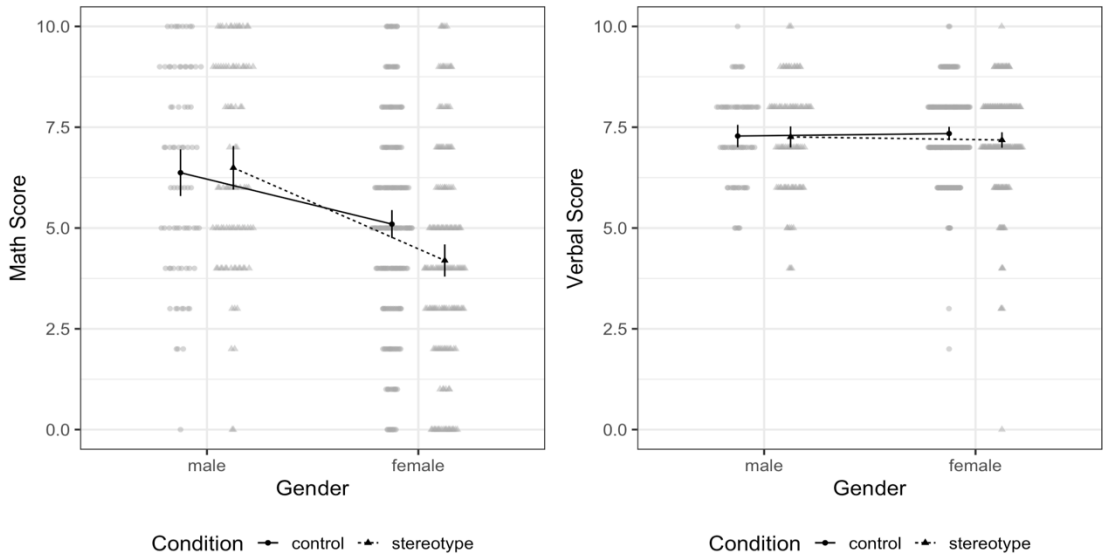

Figure 1: Participants' math test score and verbal test score in different conditions and genders. The left panel shows participants' performance for the math test and the right panel shows participants' performance for the verbal test. Data points are shown as grey dots in the background with jitter. Black dots show the means of each group and the error bars show the associated 95% confidence intervals.

The ANOVA analysis for the math test score showed no significant effect of stereotype activation, $F(1, 543) = 2.58$, $p = .109$. This is within our expectation since it is unlikely that stereotype threat activation alone will have an effect without being the stigmatized group. However, there was a significant effect of gender, $F(1, 543) = 54.31$, $p < .001$. This supports the stereotypical idea that there is a gender gap in math performance in general. Moreover, the conditions by gender interaction effect was also found to be significant, $F(1, 543) = 4.46$, $p = .035$. The analysis revealed that the effect of the stereotype activation is dependent on gender.

As shown in Figure 1, female scored lower than male in the math test across both conditions. Furthermore, the score seemed to be even lower under the stereotype-activation condition. Thus, we conducted follow-up analysis to examine our observation and the details of the interaction effect.

We then compared two contrasts, one being control condition vs. stereotype-activation condition in male, and the other being control condition vs. stereotype-activation condition in female. After controlling for multiple testing using the Bonferroni-Holm correction, we found a difference of -0.12 in the male contrast, $t(543) = -0.30$, $p = .761$, and a difference of 0.90 in the female contrast, $t(543) = 3.35$, $p = .002$. The results demonstrated that stereotype threat effect on math performance is only

significant when applying to female group. This aligns with our expectations, since female group instead of male group is the social group that is thought to be negatively affected by stereotype in math performance. Moreover, this illustrates that although the gender gap between male and female in math performance does exist, the activation of this stereotype makes math performance of female even worse.

The ANOVA analysis for the verbal test score showed no significant effect of stereotype activation, $F(1, 543) = 0.62$, $p = .433$. There was also no significant effect of gender, $F(1, 543) = 0.00$, $p = .949$. Condition by gender interaction didn't show significant effect either, $F(1, 543) = 0.34$, $p = .560$. In Figure 1, we can also see that there is not much difference between groups. This analysis revealed that gender gap does not exist in verbal performance. However, it is worthy to question if the stereotype threat effect does not exist in verbal performance. Although the effect is not seen in the present study, the stereotype activation is triggered with gender gap in math performance instead of verbal performance. Therefore it is hard to conclude that verbal performance is not affected by stereotype threat. The only conclusion that we could draw from this analysis is that the effect of stereotype threat in math performance does not extend to verbal performance.

Overall, the data provides evidence of stereotype threat on subsequent math performance. Female do have a worse math performance than male in general, but the stereotype threat further worsened the math performance of female. Furthermore, the data demonstrates the stereotype threat in math performance cannot be extended to verbal performance.