Preference and detectability of cubist artworks

Preregistration Report - Group 8

Antonia Becker, Michelle Görlitz, Janne Kaeder, Clara Schier, Marisa Wodrich

Background:

In this replication of the experiment “Give me Gestalt! Preference for cubist artworks revealing high detectability of objects” (2012) by Muth et al, we want to investigate whether the viewers’ appreciation of cubist artwork is closely linked to the viewers’ ability to identify an object, or Gestalt, from partial clues. Therefore we examine if participants without expertise in cubism like cubist artworks more if they are able to detect concealed objects in them.

It is interesting to know whether there is a positive correlation between detectability and preference. We want to do the replication to investigate whether or not our results are similar to the results of the original study.

Research question:

Is the viewers’ appreciation of cubist paintings linked to their ability to identify objects (in a cubist painting)?

We will also investigate whether there is a difference between monochrome and colored pictures (variable: “color”), and between monochrome, sepia and colored artworks (variable: “color2”) because this could be a possible confounding variable. In the original study, the distinction was only made between colored and monochrome, but we chose to additionally distinguish between colored and sepia as well. We defined the sepia pictures as the ones which only included colors from the brown-ish color plate.

Furthermore we examine if there is a significant difference in preference depending on the artists (Picasso, Gris, Braque).

- Hypothesis 1: Viewers prefer paintings in which they are able to detect more ‘objects’.
- Hypothesis 2: How much a viewer likes a painting is significantly affected by its color.
- Hypothesis 3: How much a viewer likes a painting is significantly affected by its artist (which stands for its style).

Sampling plan:

A minimum of 20 participants is required, but we plan to get approximately 50. The experiment will be an online experiment and thereby we plan to acquire participants through social media posts and by recruiting friends and family. The qualifications for the participants are normal vision (or corrected-to-normal), especially color-blind subjects are to be excluded as well as experts in cubist artwork and people who participated in the Experimental Psychology Lab.

Materials:
We will be using 120 cubist artwork pictures from the original study, provided by the authors of the original paper. In our final submission we will provide all details for the stimuli artworks.

The stimuli set includes 47 cubist artworks by Picasso, 40 by Gris and 33 artworks by Braque. All pictures have the format of 450x600 pixels.

For the visual testing of the participants, we will include a Standard Snellen's eye chart test as well as two Ishihara color vision test plates. These were found and downloaded from google pictures where they were available for free. Concrete sources for the test plates will also be included in the final submission.

Measured variables:
What we are going to measure is (A) the subjective preference of the cubist artwork pictures as well as (B) the subjectively perceived visual detectability of any form within each picture. The detectability is the independent variable, while the preference is the dependent variable.
For both variables, the data will be treated on an ordinal scale (values are scaled from low to high), trying to find a correlation.
We also have the independent variables color (monochrome / colored; pictures from the original study are tagged like this), color2 (monochrome / colored / sepia) and artist (Picasso / Braque / Gris). Those variables are not measured: they are (hard) coded in the experiment for each picture.

Procedure:
Participants are shown written instructions and general information about the upcoming tasks on the screen. Prior to the actual experiments the subjects take two vision tests: the Standard Snellen's eye chart test and a short version of the Ishihara color vision test.
Unlike the original study, we will do the experiment online, so the participants can take part from their own device from home.

During the first block of the experiments the participants are presented pictures of 120 cubist artworks in a randomized order. For each picture they are asked to rate how much they like the artwork.
For the second block the participants are shown the same 120 stimuli again in a randomized order, but this time they are asked to rate how well they can detect objects within the artwork.

In both blocks the participants proceeds directly to the next trial after giving their rating. All ratings are chosen from a 7-point-Likert-scale from 1 ('not at all') to 7 ('very').

After finishing both blocks the participants are asked for further information, including gender, age as well as their expertise in cubism rated on a 7-point scale from 1 ('none') to 7 ('a lot'). Additionally they are given room for optional further comments.

Analysis Plan

Exclusion criteria:

We will exclude participants who gave incorrect answers on the vision tests or rated their expertise in cubist artwork as 6 or higher. In addition we will exclude participants who took less than 5 minutes for the whole experiment, as this would mean that they did not take the experiment seriously and just pressed as fast as they could.

Furthermore, we will exclude subjects which they participated in the Experimental Psychology Lab course as they might have prior knowledge about the experiment. In order to be able to do so, we have a question at the end of our experiment.

Confirmatory hypothesis testing:

Data transformations:

First of all, we only select the variables that are interesting for our further analysis (those are submissionID, trialName, response, pictureNumber, artist, date, age, gender, timeSpent, color and color2).

Then, we apply the exclusion criteria: We go through our data and look for every participant whether one of the exclusion criteria applies. If so, we will exclude all of this participants' data.

After calculating general things like how many participants we have (before and after excluding), mean age, age range and gender distribution, we will transform our data into the right form. Until now, every row in our data represents a response from a participant (those are towards different questions). But what we need is 120 rows per participant, each representing a different picture and the responses the participant gave towards the questions "How much do you like this picture?" / "How well can you detect objects within this artwork?". Therefore, we combine rows that have the same submissionID, pictureNumber, artist, date, color and color2, but a different trialName (ratingScaleLike vs. ratingScaleDetect). We do that with the spread() function and get rows that contain the responses towards both questions.

Statistical model:

We will use the ‘brms’ package to run Bayesian regression models regressing the dependent variable “liking” against the independent variable “detectability” (and later also “color” / “artist”). We do that with different models:

1. In the first approach, we treat the detectability scale as a factor and calculate the difference of the sample points: This model compares the sample points for each category of detectability (2 to 7) to the sample points from the category detectability = 1.
2. In a second model and all the further ones, we treat detectability as monotonic by adding mo(detectability). This model also includes random effects (random-by-participant intercepts by adding (1|submissionID)).
3. The third model is like M2, but without random effects.
4. The fourth model is like M2, but also includes the independent variables artist, color and color2. This model includes random effects (random-by-participant intercepts by adding (1|submissionID)).
5. The fifth model is like M4, but without random effects

Hypothesis testing:

1. H1: We will compute different models that fit this hypothesis and print them. Furthermore, we will plot them with the marginal_effects() function, one with categorical = T and once with categorical = F.
2. H2: We will compute different models that fit this hypothesis (M4, M5) and print them. Furthermore, we will plot them with marginal_effects() function, one with categorical = T and once with categorical = F.
3. H3: We will compute different models that fit this hypothesis (M4, M5) and print them. Furthermore, we will plot them with marginal_effects() function, one with categorical = T and once with categorical = F.

Reformulation of hypotheses:

1. If H1 is true, we expect that our model 1 only contains positive estimates for the effect of detectability which are higher the higher the detectability is. We also expect that the 95%-confidence intervals never contain the 0.
   Additionally, we expect a Pearson correlation to be similar to the one they found in the original paper. It should be positive and the 95%-confidence interval should not include the 0.
   We also expect that in all models, the estimate of mo(detectability) in the population-level effects is positive and that its 95%-confidence interval does not include the 0. The plots with categorical = F should look like liking increases when the detectability increases.
2. If H2 is true, we expect that the estimate of “colormonochrome” in the population-level effects is not 0 in M4 and M5 and that its 95%-confidence

interval does not include the 0. Additionally, we want to see a difference in the plots: colored / monochrome should have different effects on liking.

Furthermore, we expect that the estimates of the color2 variables in the population-level effects is not 0 in M4 and M5 and that its 95%-confidence interval does not include the 0. Additionally, we want to see a difference in the plots: colored / monochrome / sepia should have different effects on liking.

3. If H3 is true, we expect that the estimate of artistG and artistP in the population-level effects in M4 and M5 is not 0 and that their 95%-confidence interval do not include the 0. Additionally, we want to see a difference in the plots: P / B / G should have different effects on liking.
4. If all of our hypotheses are true, we expect that the loo() function returns that our Models M4 and M5 fit the data best (when we compare all models).