**Preregistration**

**1. Study information**

*Title (required)*

How do various constructions of gender through response options shape categorization of gender?

*Authors (required)*

*Description (optional)*

The act of ascribing a social category has consequences for how that person is understood. This means that the process of social categorization is an important topic to study. A challenge when conducting this type of research is that the researchers themselves cannot be completely separated from the processes which they study. In conducting experiments and surveys, researchers necessarily place boundaries and limitations on the type of categorizations. For example, gender can be conceptualized in multiple ways including as a broad spectrum, encompassing many and fluid categories or as an inflexible binary consisting of women and men only (Hyde et al., 2019). This implies that when researchers measure gender categorization of faces, the design of a study and how response option are used, re-create and communicate ideas about gender. The purpose of this study is to examine the specific effect of response options, investigating how various response options, allowing gender to be more or less fluid and open, affect the categorization of gender.

*Hypotheses (required)*

The study exploratory and descriptive. As such, no specific testable hypotheses are made. The aim of the study is explore how five different conceptions of gender, communicated through given response options to a gender categorization task, affect the distribution of gender categories.

In lieu of statistical tests of hypotheses, the data will be summarized and presented visually. Statistical modelling will be carried out, primarily in order to estimate the uncertainty of our measures. We expect to see a broad trend where participants given the option to make gender categorizations beyond the binary choose to do so. Furthermore, we expect to see a broad pattern of results consistent with categorical perception but only for the more binary response options. In practice this means that stimuli suggestive of women and men slightly tilted toward one category should be overwhelmingly categorized as belonging to that category, but only when participants categorize gender using binary response options.

## **Design Plan**

*Study design\**

Across two studies, two types of categorizations will be measured. In study 1, participants categorize faces which are morphed to appear more or less like women or men. In study 2, participants categorize descriptions which contain mixtures of traits stereotypically associated with women or men. In both studies, there are five response options conditions, where the response options are varied to be more fluid or less fluid (see variables for more precise details on the experimental conditions). Both studies are between-subjects experiments, where the subjects assigned a condition based on random chance.

*Procedure*

The procedures for both studies is very simple. Participants are presented with the stimuli and asked to categorize them according to gender. Although the number of stimuli vary across the studies (see next section) the order of stumulus presentation is randomized in both studies.

*Stimuli*

In study 1, participants are shown 140 faces. These are morphs, created using psychomorph software (Tiddeman et al., 2001). The morphs are created using pairs of women and men in the london face database (DeBruine et al., 2012). Twenty pairs are randomly selected from the database, and morphed in six increments, resulting in seven faces from each pair. To very roughly approximate the demographic breakdown of the country in which the experiment is conducted, half of the faces are white, a quarter are Asian and a quarter are Black (as self-identified by the photographed people). No morphing is done across race.

In study 2, participants are show 12 sets of descriptor. Each descriptor is a set of six adjectives stereotypically associated with either the communion or agency traits. The words are randomly selected from a pre-validated list of descriptors with the restrictions that each possible combination of number of communal and agentic traits is presented twice. In other words, two trials contain six communal words, two trials contain five communal and one agentic word and so on.

## **Sampling Plan**

**Check**Registration prior to creation of data

*Data collection procedures\**

**Study 1**

Study one will a laboratory experiment where participants are shown a series of faces and asked to rate them in terms of gender. Participants are randomly assigned one of three response options conditions. Based on available funding, 100 participants will be tested for study 1.

**Study 2**

Study one will an online experiment where participants are shown a series of descriptions and asked to rate them in terms of gender. Participants are randomly assigned one of three response options conditions. Based on available funding, XX participants will be tested for study 2.

## **Variables**

*Manipulated variables*

**Study 1**

The main manipulated variable will be the response options used to measure gender categorization/rating of faces. The following response options will be used.

-  *Binary categories*. Participants categorize gender as two discrete categories consisting of the options “woman” and “man”.

- *Multiple categories*. Participants categorize gender as four discrete categories, consisting of: “Woman”, “Man”, “Other” and “I don’t know”.

- *Binary scale*: Participants categorize gender using a slider, a single line where they indicate how they perceive the face on a continuous scale from “Clearly a woman” to “Clearly a man”.

- *Multiple scales*. Participants categorize gender using a slider, similar to the binary scale, with the sole difference being the scale ranges from “Not a woman/man” to “Clearly a woman/man”. Each stimulus is categorized twice, once for each gender identity.

- *Free text.* Participants are given an open text and are free to write in whichever categories they find most appropriate.

**Study 2**

Study 2 will contain identical conditions as study 1. However, the main difference is that the categorization/rating will be of descriptors rather than faces.

*Measured variables*

**Study 1**

The main measured variables are participants categorizations/ratings of faces. Because study 1 uses five different scales, there will be six outcomes, one for each condition, except for the *Multiple scales* condition, which has two outcomes.

**Study 2**

Again, study 2 uses an identical outcome measure, to a different stimuli.

**Analysis Plan**

*Statistical models\**

**Study 1**

Due to exploratory nature of the study, the primary purpose of modeling is to estimate the uncertainty of the measures, rather than to carry out inferential testing. Because the various measures consist of radically different outcome, separate models will be fit to each outcome measure. Two values are of interest. First, the overall tendency to which the various categories are used. Second, how often the categories are used at various morph levels. The modelling strategy reflects these goals, where the main variables of interestest are both overall intercepts and unique intercepts at each level of morph.

The goal of modeling is to estimate these parameters to quantify the uncertainty around them. Consequently, no inferential tests are specified, nor are any specific comparisons made. This modeling strategy consists of several multilevel models of increasing complexity. In the simplest model, (Model 1) unique intercepts are calculated for each morph level. Additionally unique intercepts are calculated for each subject as well for each unique face, with hierarchical priors on both. This allows us to estimate both the overall rate at which various categories are used as well the rate at which each morph level is categorized. Furthermore, the model incorporates two sources of uncertainty: individuals having different rating and the various face stimuli being differently perceived.

A somewhat more complicated model (Model 2) allows the effect of morph level to vary by race. A last, most complicated model (Model 3) allows the effect of morph level to vary by subject. These three models will be compared using leave-one-out cross validation. The purpose of fitting multiple models and carrying out model comparison is to account for the uncertainty in which model is correctly specified.

**Study 2**

The analyses for study 2 is identical with the exception that no faces are presented and consequently not modelled.

*Transformations*

*Inference criteria*

*Data exclusion*

Particpants that complete the experiment too fast are excluded. Furthermore, participants who answer the same category to all faces are assumed to not be trying very hard and are excluded unless that category is “I don’t know”. The reasoning behind this exception is that some participants may take the philosophical stance that gender cannot be visible from faces.

*Exploratory analysis*

The effect of race of face is not expected to have a large impact on the results. However, if the model 2 performs the best, this would suggest that there is an interaction between morph level and face. In that case we will further explore how various faces are categorized at different morph levels, broken down by race.