R Notebook

Start by loading dependencies and the data

Background

So, the background of this study is there's a pretty big literature that looks at social categorization. This tends to assume that gender is binary and that asking people to categorize faces as "man" and "woman" is a consequence free action.

We thought that maybe it wasn't, and we were interested in whether various response options that situated gender as more binary also shaped people's perception of gender to be more binary.

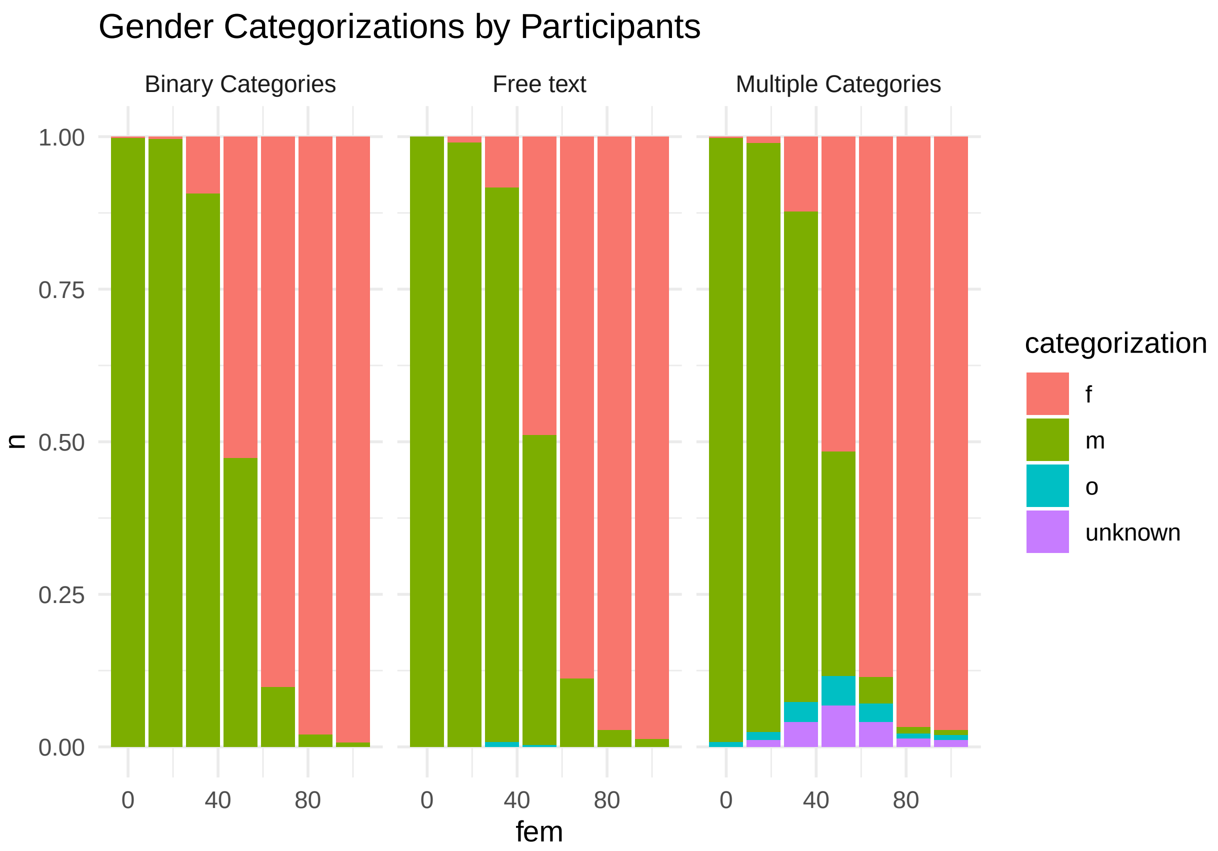
So for the experiment, we produced morphed faces of different levels of femininity and masculinity. There were 18 continua, where gender varied in seven increments, for a total of 126 faces.

There were five response options conditions:

1. binary categories - man/woman
2. multiple categories - man/woman/other/don't know
3. Freetext - a free text box
4. binary dimension - woman ——- man on a slider
5. multiple dimensions - woman / man on separate sliders.

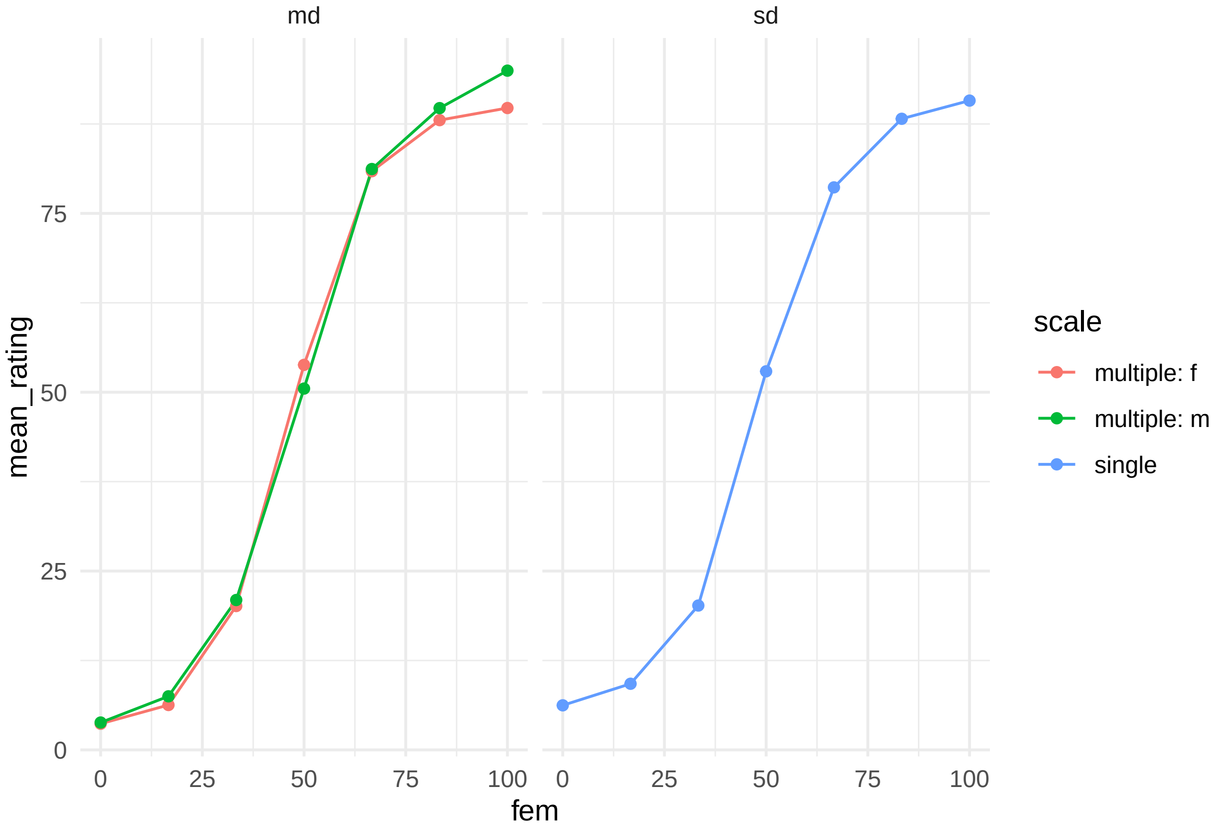
Visualizing the data

First I just want to get a sense of what the distribution of responses looks like. The following graph just shows the raw distribution of categorizations across the three "categorical" conditions. I.e. the conditions where participants respond with discrete categories. My first impression is that these all look quite similar.



Second, I did the same thing for the "dimensional" conditions, i.e. the conditions where participants respond using a dimensional slider. This image again shows the mean ratings at every level of masculinity at multiple dimensions (md) and single dimension (sd). I reverse-coded the femininity rating to make the two more comparable. Again, just a visual inspection of the curves suggest they are *quite* similar.

```
## 'summarise()' has grouped output by 'fem', 'scale'. You can override using the
## '.groups' argument.
```



Binomial models

I start by looking at the first three conditions. The same ones I called categorical in the earlier section. This is questionable at best, but I recoded the data so that I have the outcome based on the answer "woman" = 1, anything else = 0.

Having made this questionable choice, first I fitted a binomial logistic model with fixed effect of condition and facial masculinity and varying intercepts for faces, varying intercepts for subjects and varying intercepts for masculinity (i.e. allowing the effect of morph to vary for each subject).

$$
\begin{aligned}
\text{categorization}_i &\sim \text{Binomial}(1, p) \\
\text{logit}(p_i) &= \gamma_{cid[i]} + \alpha_{subject[i]} + \beta_{cid[i]} M + \gamma_{face[i],cid[i]} \\
\gamma_{cid} &\sim \text{Normal}(0, 3), \text{ for } cid = \text{ft, bc, mc} \\
\alpha_{subject} &\sim \text{Normal}(0, \sigma_{subject}) \\
\beta_{cid} M &\sim \text{Normal}(0, 3), \text{ for } cid = \text{ft, bc, mc} \\
\begin{bmatrix} \beta_{ft} \\ \beta_{bc} \\ \beta_{mc} \end{bmatrix} &\sim \text{MVNormal}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_{face} \right) \\
\Sigma_{face} &= \mathbf{S}_{\beta[cid]} \mathbf{R}_{\beta[cid]} \mathbf{S}_{\beta[cid]} \\
\sigma_{subject} &\sim \text{HalfCauchy}(3) \\
\sigma_{\gamma_{pronoun}} &\sim \text{HalfCauchy}(3) \\
\mathbf{R} &\sim \text{LKJcorr}(2)
\end{aligned}
$$

Table 1: Table 1. The effect of morph level on ratings of woman across three experimental conditions

	Slope	Est. Error	CI - L	CI - U
conditionft:fem	0.15	0.01	0.13	0.17
conditionmc:fem	0.16	0.01	0.14	0.18
conditionxb:fem	0.18	0.01	0.16	0.20

Note. CI-L = Lower credible interval, CI-U = Upper credible interval

```
fit_binary_index <- brm(f_cat ~ 0 + condition + fem:condition + (1 +fem|id) + (1|face), family = bernoul
          prior = c(prior(normal(0,3), class = "b", coef = "conditionmc"),
                    prior(normal(0,3), class ="b", coef= "conditionmc:fem"),
                    prior(normal(0,3), class = "b", coef = "conditionxb"),
                    prior(normal(0,3), class = "b", coef = "conditionxb:fem"),
                    prior(normal(0,3), class = "b", coef = "conditionft"),
                    prior(normal(0,3), class ="b", coef= "conditionft:fem")
                    ),
          data = tmp,
          iter = 4000, warmup = 1000,
          chains = 4,
          cores = 4,
          sample_prior = TRUE,
          file = "models/fit_binary_stair_index3"
          )
```

Cool, so we fit this data and what do we find? If we just start by getting a summary.

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'
```

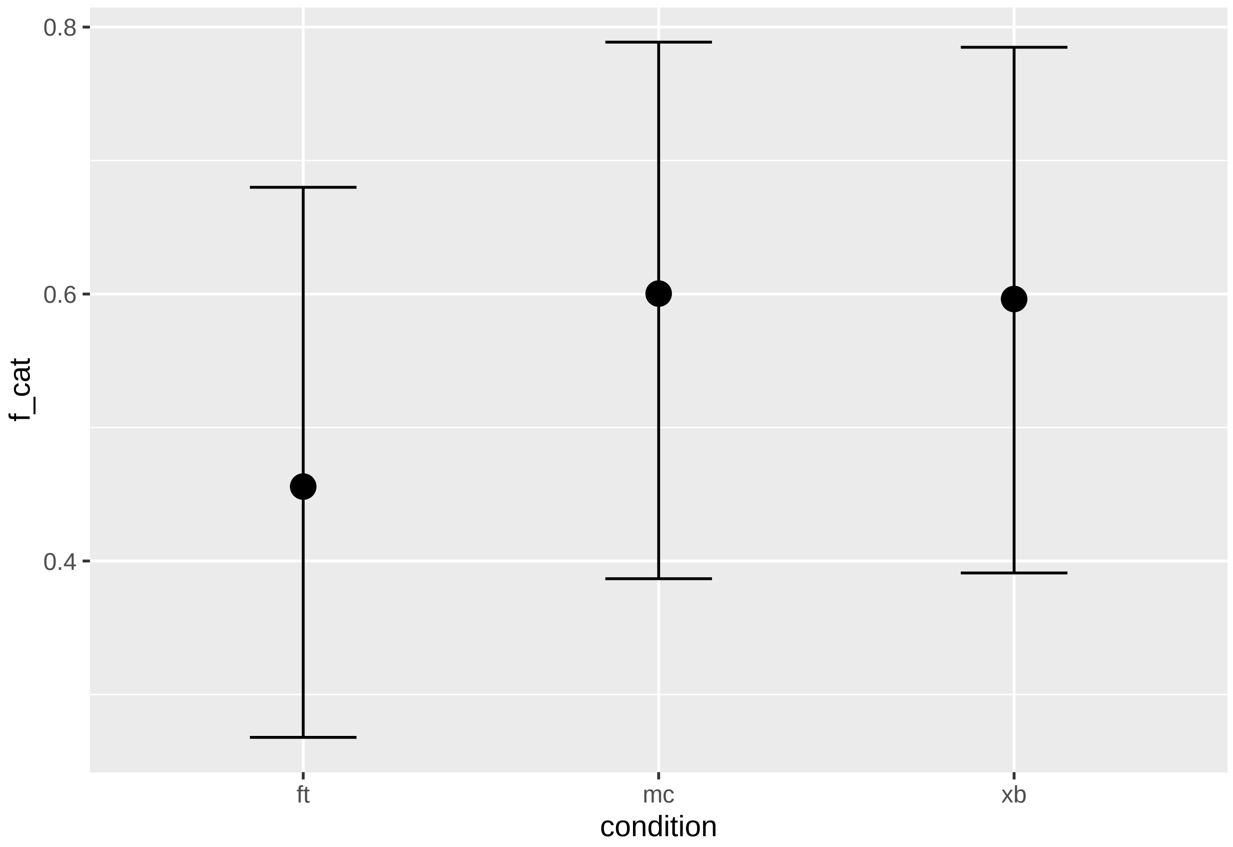
```
##
## Attaching package: 'kableExtra'
```

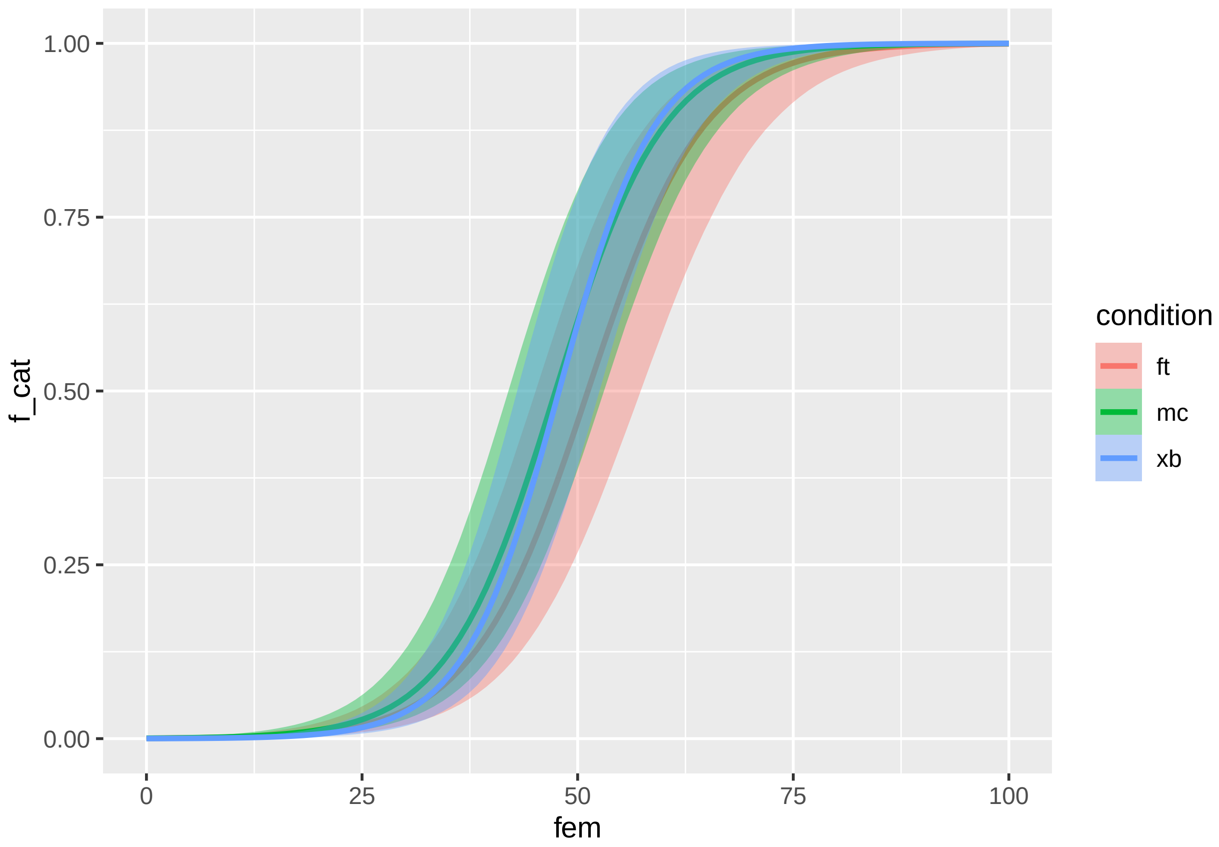
```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

Well, sort of as expected, the conditions don't look very different from each other. The main parameters (i.e. outcomes) of interest here are the slopes. These are written above as `conditionft:fem` etc. `conditionft:fem` represents the effect of morph value on ratings of "woman" in the free text condition, `conditionxb:masc` represents the effect of morph value on ratings of "woman" in the binary choice condition etc. The thing to note here is that the numbers are all very similar, suggesting that there isn't really a difference in the steepness of the slope depending on condition. If there was a categorical perception effect in for example the binary choice condition but not in the multiple categories condition, we'd expect to see a much steeper slope there. That's not what we found. At least not if we look at just the numbers.

But my stats guru McElreath always cautions against making too much of tables so we're also going to plot the data.

```
conditional_effects(fit_binary_index)
```





In the first of these figures, we have the main effect of condition, with the y axis showing the expected proportion of faces categorized as women (0.5 meaning, well, half). So, the fact that the free text condition, the estimate is at 0.4, suggests the distribution a slight male bias. I don't think this bias is obvious when I was just looking at the raw data. So maybe it's a bug?

The second image shows the curves, and this confirms what the table told us, that they were quite similar. How to read these curves? The y axis shows the proportion of faces categorized as womenm, the x-axis shows the morph value, with higher being more feminine. As we might expect, as the faces become more feminine, a larger number of them are categorized as women. What we are mainly interested is, again, *the steepness* of the curves. Or a more or less enhanced s-shape. The various colors represent the different conditions, and they are all similarly s-shaped, at least visually.

So the million-dollar question is, then, are these slopes the same? We can test this using bayes factors. These show that no, the evidence suggests pretty overwhelmingly that the slopes are the same (all BF >25) Well, we can directly find the values for the differences in the slopes though using the built-in `hypothesis`function in brms. It spits out a number of figures, but the "evidence ratio" is the same as the bayes factor. In this case it's the BF01. In all cases it's higher than 30, which suggests there's fairly strong evidence that these curves are basically the same (see table 2). Okay! Good to know. So the conclusion looking at the effect of different types of response options on ratings of "woman" is that including more non-binary and open gender options doesn't reduce the tendency for binary thinking. I think we are allowed to be a little disappointed by this.

```r
h1 <- hypothesis(fit_binary_index, "conditionft:fem= conditionxb:fem" )
h2 <- hypothesis(fit_binary_index, "conditionft:fem= conditionmc:fem" )
h3 <- hypothesis(fit_binary_index, "conditionmc:fem= conditionxb:fem" )

tests <- rbind(h1$hypothesis, h2$hypothesis, h3$hypothesis)
```

Table 2: Table 2. Bayes tests of the difference between slopes in three experimental conditions

Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Evid.Ratio
(conditionft:fem)-(conditionxb:fem) = 0	-0.03	0.01	-0.06	0.00	25.45
(conditionft:fem)-(conditionmc:fem) = 0	-0.01	0.01	-0.04	0.02	228.29
(conditionmc:fem)-(conditionxb:fem) = 0	-0.02	0.01	-0.05	0.01	91.93

Note. CI-L = Lower credible interval, CI-U = Upper credible interval

```
kable(
  tests[,1:6] %>%
    mutate_if(is.numeric, ~round(.,2)),
  booktabs = "TRUE",
  align = c("l", "c", "c", "c"),
  caption = "Table 2. \nBayes tests of the difference between slopes in three experimental conditions"
  ) %>%
  kable_classic(full_width = F) %>%
  footnote(
    general_title = "Note.",
    general = "CI-L = Lower credible interval, CI-U = Upper credible interval",
    threeparttable = TRUE,
    footnote_as_chunk = TRUE
    )
```

Gaussian models

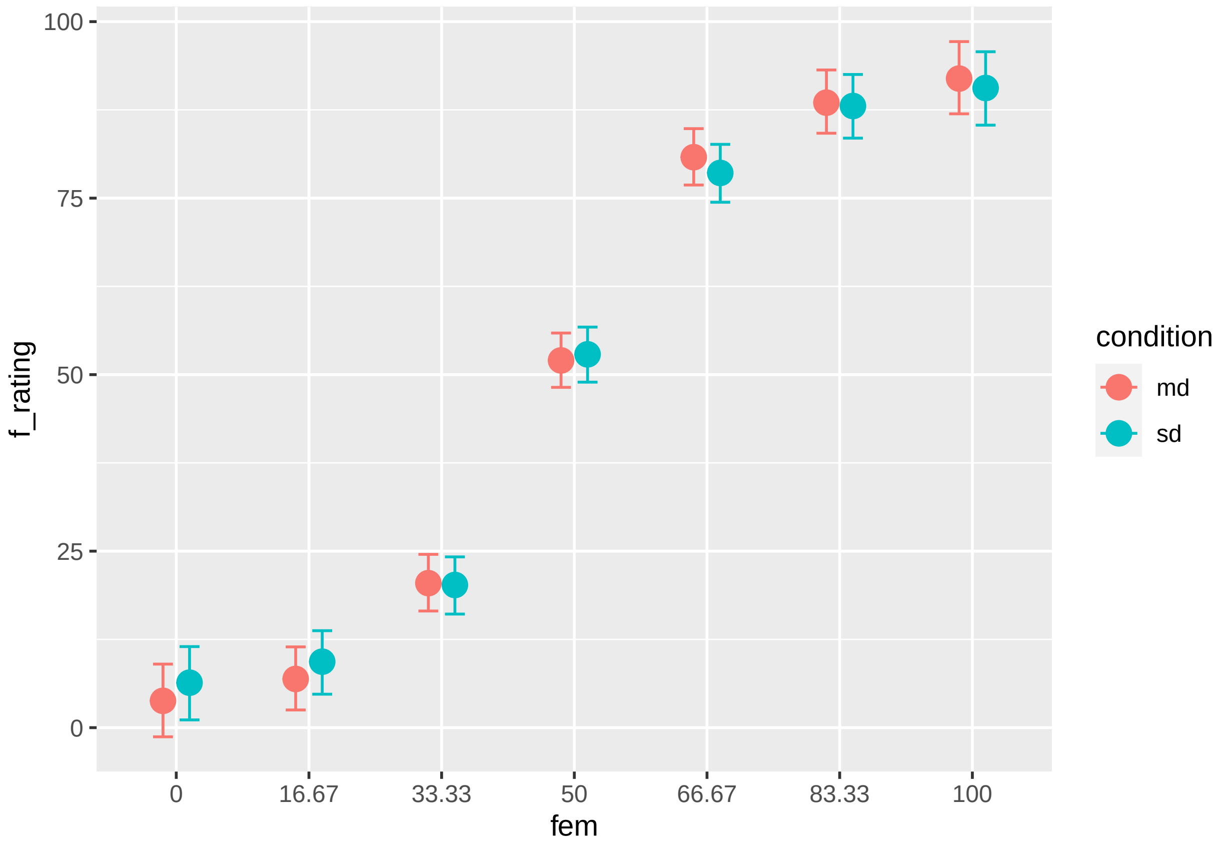
The next step would be to look at the last two conditions. In other words, the single dimension and the multiple dimensions. The visual inspection suggest they are very similar, but can we model that? This is where I'm a little bit uncertain about the best approach, and where talking to a curve-modelling expert would be useful. Of course, this is pretty challenging, and would possibly involve doing some non-linear curve fitting monstrosities. That's a little above me though! However, Somebody I talked to suggested skipping all the non-linear curve nonsense and just factorize all the points on the masc scale. The following just calculates a unique intercept at each level of masculinity and condition. Is this the best way to do it? Probably not, but it give another rough guesstimate.

For the plot below, I've reverse coded the man ratings to create one aggregate multiple dimensions (i.e. "md") scale. So the plot essentially shows how the average rating of femininity at each level of morph. For both conditions, again, the pattern of responses is suggestive of categorical perception. What am I basing this on? The critical levels are the levels between 0 and 50 and again between 50 and 100, particularly 33.33 and 66.67. The rated levels of femininity (or reverse coded masculinity) are lower and higher respectively, suggesting that there participant percieve more femininity than "there is" in the faces, which is to say, categorical perception.

```
#Wrangle data
tmp <- d %>%
  filter(condition == "sd" | condition == "md") %>%
  mutate(f_rating = as.numeric(categorization) %>%  ifelse(scale == "f", ., 100- .),
         scale_new = ifelse(scale == "f" | scale =="m", scale, "sd"),
         fem = as.factor(fem))
```

```
fit_dimensional_stair_factor <-
  brm(f_rating ~ 0 + fem:condition + (1 + masc|id) + (1|face), family = gaussian(link = 'identity'),
          prior = c(prior(normal(50,50), class = "b"),
                    #prior(normal(50,50), class = "Intercept"),
                    prior(exponential(1), class = "sd"),
                    prior(lkj(1), class = "cor"),
                    prior(exponential(1), class = sigma)),
          data = tmp,
          iter = 4000, warmup = 1000,
          cores = 4,
          sample_prior = TRUE,
          file = "models/fit_dimensional_stair_factor.3")

conditional_effects(fit_dimensional_stair_factor)
```



Adding pronouns to the mix

Okay, so why might we want to look at pronouns as an outcome? The upside of looking at pronouns are that they would be a more practically relevant outcome. Asking someone “which pronoun would you use to describe this person?” would get us a little closer to how they would treat a person in real life. What a lot of non-binary people are advocating for is this kind of habit, where most people refrain from using gendered pronouns without knowing them. Another upside of looking at pronouns would be that it would tie this study more neatly together with the other papers, as pronouns would be a more clearly occurring theme.

What are some downsides, or at least some challenges? There is a comparability problem. Is a condition where participants categorize using pronouns really the same as when participants are applying categories. This gets at the broader question of is a pronoun a category. And the difficulty is that it's kind of not. It's just a way to refer to someone. On the other hand, it is the case that using a pronoun is an indicator of having put someone into a category, and maybe more importantly, non-binary pronouns are specifically a way to not use place someone in a category. There's also the issue if we're talking about pronouns as an outcome, can we even talk about categorical perception? I mean, if people are using the pronoun she, is that based on them perceiving the person as a woman? It gets a little bit muddy there. So pronouns are related to categorization, are perhaps indicative of categorization, but can't be said to be a direct measure of categorization. Is that a problem? Maybe! It might mean that that directly comparing a pronoun condition to the ones above is misleading.

Another potential issue is that using pronouns would take us a couple of steps away from the original research question. Again, the original research question is how do various types of response options affect categorical perception. The sort of implicit in this question is that the response options are ones used by researchers in experiments. I don't think it ever happens to be the case that researchers in categorization studies are giving people the option to use certain pronouns. Fair enough, so we say that this study is about categorization on a broader, more naturalistic level. Okay, but even this, this type of design is a little awkward. It isn't very naturalistic for someone to pick a set of pronouns from a list. So a potential pronoun condition would sort of awkwardly straddle a middle ground of not being very similar to anything done by practicing researchers but also not being a very naturalistic situation.

Implications of pronouns

Okay, so given these strengths and challenges, where does that leave us? Does this suggest that we need to give up? Well, let's think about. First of all, something that I've noted is that we already have two types of outcomes: categorical and dimensional. I've already basically written them up here as two separate experiments. And if we already have an article with two experiments, why not add a third? So if we're in the mindset of a third experiment, then that creates a little bit of space for the pronoun experiment, however, it looks, to be a little different.

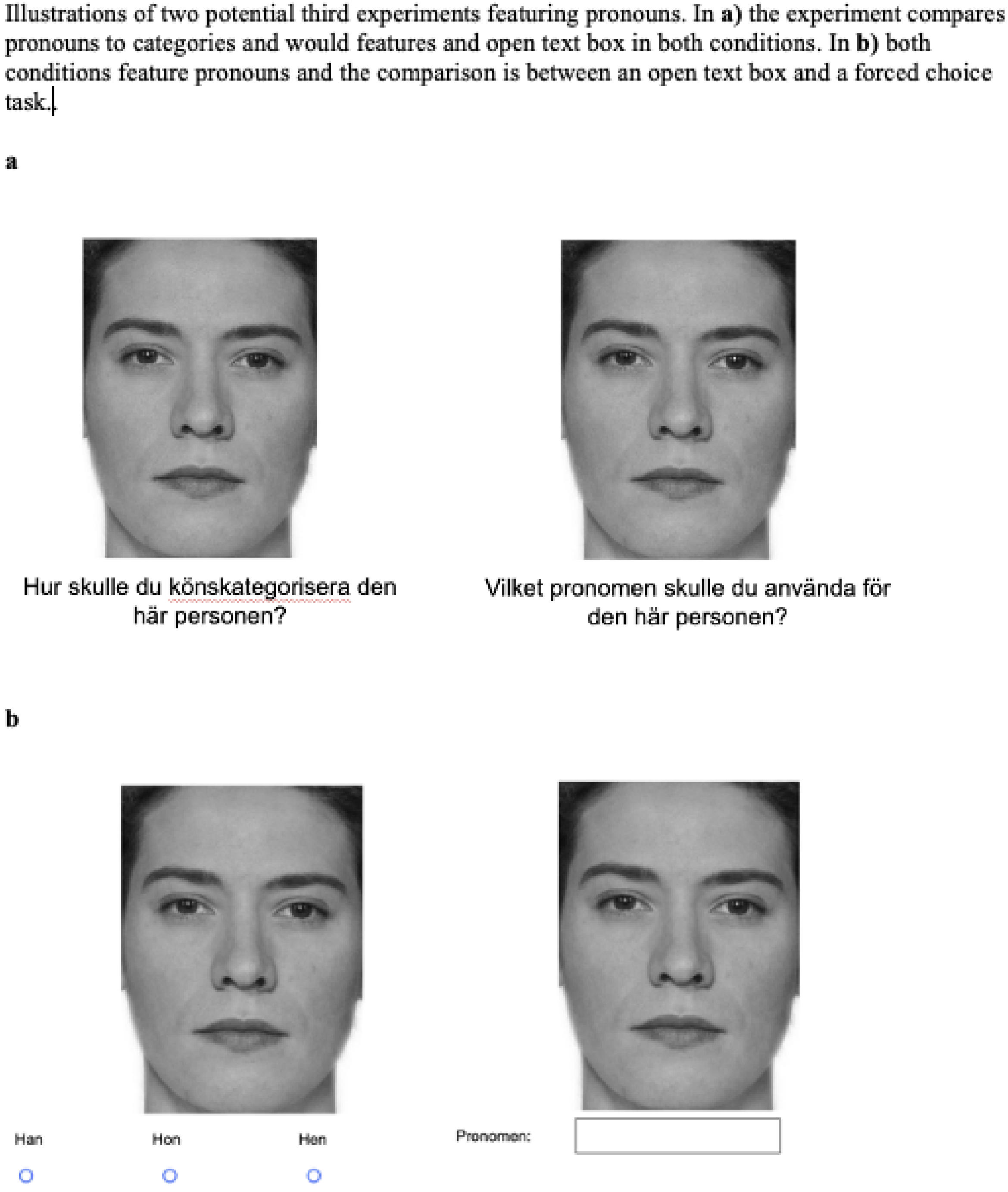
Adding a pronoun condition might mean having to broaden the research question. Rather than "specific response options" affecting "categorical perception", it could be something like "gender framing" affecting "binary treatment of gender" where binary treatment of gender is defined as categorical perception in experiment 1, 2 and 3 and maybe also just "beyond-binary responses" in experiment 3 (wow, return of a classic!)

So what could an additional experiment actually look like? One way is to compare a pronoun condition to the free-text condition (as in **a** in the picture below). When people spontaneously name gender categories, they think about of gender as something binary, but when people give pronouns, it might somehow be filtered through social interactions and they'd maybe be more open to going beyond the binary. That would be an interesting finding! Then the outcome of interest would be something more along the lines of "how often do you categorize beyond the binary"?

Another option would be to double down on pronouns, and compare two different conditions using pronouns (as in **b**). For example one condition where participants choose pronouns based on forced choice and another where participants select pronouns based on free text. This would get us a little bit closer to the original research question of how the way that options are presented affects the outcome. This kind of setup has the advantage of being more comparable.

A quick pilot

To explore whether this idea is even viable, I made a quick pilot where 20 people rated 32 faces on the basis of which pronoun they would use to refer to the person with that particular face. This study was carried out on an english-speaking samples so the options were they/them, she/her and he/him.



Illustrations of two potential third experiments featuring pronouns. In **a)** the experiment compares pronouns to categories and would features and open text box in both conditions. In **b)** both conditions feature pronouns and the comparison is between an open text box and a forced choice task.

a

Hur skulle du könskategorisera den här personen?

Vilket pronomen skulle du använda för den här personen?

b

Han ○ Hon ○ Hen ○

Pronomen: □

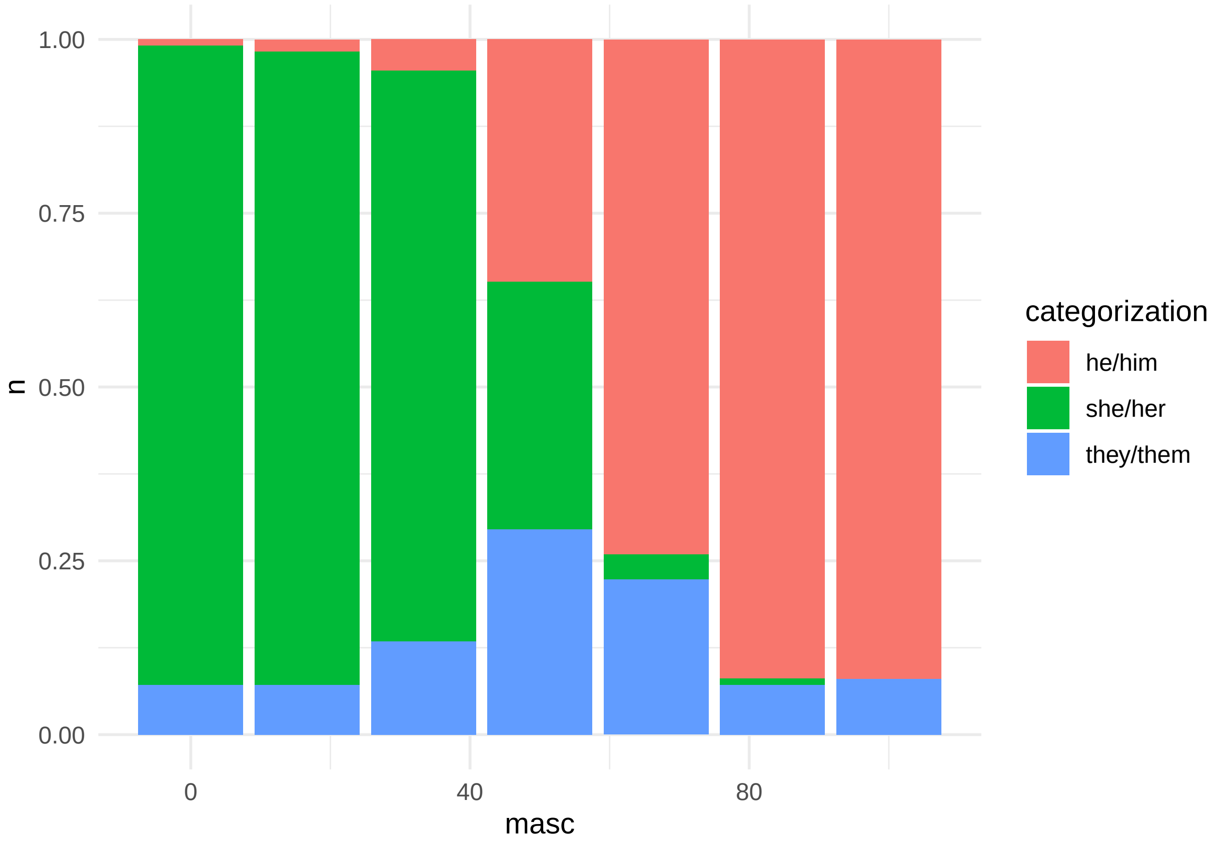
Figure 1: showing two possible experiments

Each person looked at 32 faces, both black and white.

Okay, so we've got the data, let's plot it ouuuut! Some takeaways: more people used they/them than would categorize another person the categories "other" and "I don't know" in the previous study. Looking at the distribution of picks, it also looks like there is still a categorical perception pattern. ´

```
d %>%
  group_by(masc, race) %>%
  count(categorization) %>%
  ggplot(aes(x=masc, y=n, fill=categorization)) +
  geom_bar(stat="identity", position = "fill")+
  theme_minimal()
```

```
## Warning: Removed 5 rows containing missing values (‘position_stack()‘).
```


n
masc
categorization
he/him
she/her
they/them