

The effect of response options on gender categorization (provisional title)

Elli van Berlekom¹ & Coauthors^{1,2}

¹ Stockholm University

² Lund University

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Data & scripts are available at osf link

The authors made the following contributions. Elli van Berlekom:
Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing;
Coauthors: A lot of things, Author order TBD.

Correspondence concerning this article should be addressed to Elli van Berlekom,
Albanovägen 12. E-mail: elli.vanberlekom@psychology.su.se

Abstract

I'm using a premade template & leaving some of their guidelines in place to help me.

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: keywords

Word count: X

The effect of response options on gender categorization (provisional title)

The experience of transgender and gender diverse (TGD) individuals suggests that sex/gender is a fluid category which can vary along a wide spectrum. In contrast, social categorization and face perception research often treats gender as a binary consisting of women and men (for example Webster et al., 2004). This is problematic because it indirectly delegitimizes TGD individuals' experiences. Additionally, it may restrict participants' answer, similar to how ratings of age along an old/young binary would restrict and distort ratings of age variation (see Westbrook & Saperstein, 2015; Lindqvist et al., 2019). Furthermore, it may distort answers by communicating ideas about gender. In this study, we aimed to investigate how various gender categorization paradigms influence participants' categorizations of faces.

A cursory glance at the literature on gender categorization reveals that the vast majority explicitly or implicitly suggest to participants that gender consists of the categories woman and man only. The most common method to measure gender categorization is a force-choice task, where participants are presented with a face and the choices are "female" and "male" (see for example, Cloutier et al., 2005; Campanella et al., 2001; Webster et al., 2004; Zhao & Bentin, 2008). A slightly different task asks participants to rate the faces on gender as a quality, rather than a category, often with "feminine" and "masculine" as endpoints on a single scale (e.g. D'Ascenzo et al., 2015; others). Overall, despite some variations, this is a literature where gender is frequently presented as a binary.

Presenting gender as a binary communicates to participants that the researchers do not view non-binary genders as legitimate. For TGD individuals, this may contribute to a wider pattern of cisgenderism, the ideology that discards people's own conception of their gender identity. Researchers may raise the objection that binary response options may be the most suitable for the research question or the planned statistical analyses. This may be

the case, but it should be weighed against the real harm that is being done by these options.

Furthermore, it is worth questioning whether a binary forced choice is ever the most appropriate method to measure gender categorization. This position seems to be premised on the assumption that there is some fundamental basis to gender, a truth which can be distorted. According to this view, binary is the neutral way to measure sex/gender categorization and anything else is the result of agenda-driven or political motivations. If gender is instead viewed as a social construct, which arise as a result of repeated discourse, this suggests that there is no neutral way to measure gender categorization. Rather, there are multiple alternatives which come with their own limitations and restrictions or suggestions.

Indeed, gender can be measured in many different ways, with drastically varying results. For example, Bem (1974) constructed scales to measure femininity and masculinity as separate personality traits. She found that many people had a mixture of feminine and masculine traits. In another example, when Joel and colleagues (2014) asked ostensibly cisgender participants whether they ever experienced shifts in their gender identity, a sizable group had. Lastly, and Westbrook and Sperstein (2015) showed that there are many potential ways participants answer questions about their gender identities, including rating femininity and masculinity on separate dimensions. When offered these separate sliders, participants generally offered a high degree of androgyny. These results, which primarily regard people's self-categorization and not categorization of others, nevertheless suggest that when people are given the options to categorize gender beyond the binary, they frequently use them.

Furthermore, gender binaries can be created or enhanced through statistical practices. For example, Hyde and colleagues (2018) concluded that the statistical practice of examining mean differences between women and men exaggerates the difference and

downplay gender similarities (Hyde et al., 2005). Hester and colleagues (2020), showed both that perceived differences between the faces of men and women were pronounced when only means were examined, and when gender was measured as consisting of a single dimension with femininity and masculinity at opposing ends. These studies show that when experiments are constructed to take diversity of gender into account, the results often reveal a diversity of gender. This primarily suggests that studies which only measure binary gender are unnecessarily and artificially restrictive.

Let's think about which categories would be less harmful to non-binary individuals. The inclusion of some sort of third gender option would be preferable, simply to acknowledge the existence of TGD individuals. However, the existence of a third option is not enough. Many TGD individuals argue that the best thing is just for everyone to not abstain from categorizing altogether. We can think of two ways to encourage that. One, is just add the options of choosing a third category and an I don't know category. The other is to allow participants to categorize using an open text box.

Research question 1: Do people use beyond-binary options when they have them?

Research question 2: To what extent do beyond-binary responses affect the distribution of woman/man responses?

Categorical Perception & Gender Categorization

Another question about response options is the extent to which they influence participants view of gender. When gender is measured as only the categories "woman" and "man" the implication may be that gender/sex consists of two discrete mutually exclusive categories (ref). When gender is measured on a continuum with femininity and masculinity as mutually exclusive polar opposites, it still reproduced femininity and masculinity as opposites, but suggests that degrees are possible. Additionally, there is a certain conceptual ambiguity around the terms femininity and masculinity, where they are conflated with the properties of women and men, even though there is evidence that these

terms are applied differently to women and men (Hester et al., 2020).

One way to consider how response options shape the perception of gender is to look at categorical perception. Categorical perception is a perceptual effect where people tend to accentuate the differences of continuous stimuli. It has been observed for colors and for sounds. The existence of categorical perception suggests that people have a strong sense that categories exist. Importantly, categorical perception has been observed for gendered faces (Campanella et al., 2001). However, if participants respond to gender categorization with options that are less binary, maybe they will exhibit less categorical perception?

General Method

Overview

We carried out two experiments. Experiment 1 was carried out to test research questions 1 and 2, and consisted of various

Stimuli

Faces were produced using faces from the London Face Database (deBruine) and the Chicago Face Database (ref) morphed with on Webmorph (ref). For Black, Asian and White faces, the six most feminine faces of women and the six most masculine faces of men were selected, using the codebook provided by the researchers. The faces were matched, so that the most feminine face were morphed with the most masculine face and so on.

Procedure

The same procedure were used for experiments 1 and 2. Participants rated all faces in turn, then filled out answered the gender binary beliefs scale.

Experiment 1

Participants

Participants ($N = 50$) were speakers recruited through advertising online and on the university campus ($M_{\text{age}} = 36.67$, $SD_{\text{age}} = 12.54$). All participants were informed that participation was voluntary. In term of gender X women and Y men participated The participants were randomly allocated to conditions.

Design

Experiment 1 was a between-subjects design with three conditions. Participants were randomly allocated to the conditions.

Data analysis

We used R (Version 4.2.2; R Core Team, 2022) and the R-packages *bayesplot* (Version 1.10.0; Gabry et al., 2019), *brms* (Version 2.18.0; Bürkner, 2017, 2018, 2021), *dplyr* (Version 1.0.10; Wickham et al., 2022), *ggplot2* (Version 3.4.0; Wickham, 2016), *papaja* (Version 0.1.1; Aust & Barth, 2022), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.9; Eddelbuettel & François, 2011), *tidybayes* (Version 3.0.2; Kay, 2022), *tidyr* (Version 1.2.1; Wickham & Girlich, 2022), and *tinylabls* (Version 0.2.3; Barth, 2022) and tidyverse for all our analyses. Bayesian modelling were caried out using the package. Additional

Results

Discussion

References

- Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://github.com/crsh/papaja>
- Barth, M. (2022). *tinylabls: Lightweight variable labels*.
<https://cran.r-project.org/package=tinylabls>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, 100(5), 1–54. <https://doi.org/10.18637/jss.v100.i05>
- Eddelbuettel, D., & Balamuta, J. J. (2018). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *The American Statistician*, 72(1), 28–36.
<https://doi.org/10.1080/00031305.2017.1375990>
- Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8), 1–18. <https://doi.org/10.18637/jss.v040.i08>
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in bayesian workflow. *J. R. Stat. Soc. A*, 182, 389–402.
<https://doi.org/10.1111/rssa.12378>
- Kay, M. (2022). *tidybayes: Tidy data and geoms for Bayesian models*.
<https://doi.org/10.5281/zenodo.1308151>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of data manipulation*. <https://CRAN.R-project.org/package=dplyr>

Wickham, H., & Girlich, M. (2022). *Tidyr: Tidy messy data*.

<https://CRAN.R-project.org/package=tidyr>