

Stage 1 Review, Revision 1: Attraction depending on the level of abstraction of the character descriptions

Reviewer: Florian Pargent, LMU Munich

Summary

The revised stage 1 manuscript titled *Attraction depending on the level of abstraction of the character descriptions* (https://osf.io/69byn/?view_only=dc1bb4d7647046ccae4d64ba44448921) is a great improvement on the initial version. The authors have greatly improved the study design, added some manipulation checks, adopted more suitable statistical methods based on equivalence tests and multilevel models, and provide sample size justification based on power analysis for all tests. The manuscript now describes the methods and materials in greater detail, and R code and stimulus samples are provided in the accompanying OSF repository. Although I still have some remaining concerns (and a few minor ones), these are very easy to fix. For these reasons I recommend to accept the stage 1 manuscript, after these issues have been addressed.

I outline my remaining concerns below.

Major Issues

1. An error in the power analysis R code

I think the authors made a small mistake in their code to determine the sample size for their multilevel models. Lines 32 and 33 of their script `power_analysis_H12.R` read:

```
r.y <- rnorm(Npart*Nprof, sqrt(1-ICC))  
r.med <- rnorm(Npart*Nprof, sqrt(1-ICC))
```

This is equivalent to

```
r.y <- rnorm(Npart*Nprof, mean = sqrt(1-ICC), sd = 1)
r.med <- rnorm(Npart*Nprof, mean = sqrt(1-ICC), sd = 1)
```

which is probably not what the authors intended. (Note that the same error is also present on lines 27 and 28 in `analysis_code.R`.)

Instead the code should probably read:

```
r.y <- rnorm(Npart*Nprof, 0, sqrt(1-ICC))
r.med <- rnorm(Npart*Nprof, 0, sqrt(1-ICC))
```

In the following, I have adapted the R code and reran the simulation. I have marked each line where I made some changes with the comment `# CHANGES`. In addition to this review document, I will also attach a revised standalone R script that includes the following code.

```
#modified version of Pan et al. (2018)

#Reference: Pan, H., Liu, S., Miao, D., & Yuan, Y. (2018).
#Sample size determination for mediation analysis
#of longitudinal data. BMC Medical Research Methodology, 18(1), 32.

rm(list=ls())

library(lme4)
```

Loading required package: Matrix

```
library(mediation)
```

Loading required package: MASS

Loading required package: mvtnorm

Loading required package: sandwich

mediation: Causal Mediation Analysis
Version: 4.5.0

```

set.seed(0216)

med.power<-function(m.eff,y.eff,Npart, Nprof, ICC, Nrep){

pow1 <- 0
pow2 <- 0

for(aa in 1:Nrep)
{
  participant <- rep(1:Npart, each=Nprof)
  treatment <- rep(sample(c(0,1),size=Npart,replace=TRUE),each=Nprof)

  #simulate errors
  u.y <- NULL
  u.med <- NULL
  for(k in 1:Npart){
    u.y <- c(u.y,rep(rnorm(1,0,sqrt(ICC)),Nprof))
    u.med <- c(u.med,rep(rnorm(1,0,sqrt(ICC)),Nprof))
  }

  r.y <- rnorm(Npart*Nprof, 0, sqrt(1-ICC)) # CHANGES
  r.med <- rnorm(Npart*Nprof, 0, sqrt(1-ICC)) # CHANGES

  #simulate uncertainty and attraction
  uncertainty <- m.eff * treatment + u.med + r.med
  attraction <- y.eff * uncertainty + x.eff * treatment + u.y + r.y

  #H1
  model1 <- lmer(attraction ~ treatment + (1|participant))
  t_stat <- summary(model1)$coefficients[2, "t value"]

  ##need to adjust critical t value depends on the sample size
  if(t_stat > 1.647){
    pow1 <- pow1 + (1/Nrep)
  }

  #H2
  m <- lmer(uncertainty ~ treatment + (1|participant))
  y <- lmer(attraction ~ treatment + uncertainty + (1|participant))
  modelh2 <- mediate(model.m = m, model.y = y, treat = "treatment", mediator = "uncertainty",
                    boot = FALSE, sims = 100, conf.level = .90)
  lowerCI <- summary(modelh2)$d0.ci[["5%"]]
}

```

```

    if(lowerCI > 0){
      pow2 <- pow2 + (1/Nrep)
    }
  }
  return(cat("Given",Npart,"participants and",Nprof,"repeated measures for each subject,\n"))
}

#####

x.eff<-0
ALPHA<-0.05
VARR<-1

med.power(m.eff=0.39, y.eff=0.39, Npart=880, Nprof=10, ICC=0.6, Nrep=500)

```

Given 880 participants and 10 repeated measures for each subject,
the power for H1 is 83.4 %
the power for H2 is 100 %

As can be seen above, although the power estimates change, the authors' goal to collect 1000 participants should still ensure a well-powered study (at least under the assumptions of their simulation).

To run the simulation, I used the following R and package versions:

```
sessionInfo()
```

```

R version 4.4.0 (2024-04-24)
Platform: aarch64-apple-darwin20
Running under: macOS Sonoma 14.4.1

```

```
Matrix products: default
```

```

BLAS:   /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;

```

```
locale:
```

```
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
time zone: Europe/Berlin
```

```
tzcode source: internal
```

attached base packages:

```
[1] stats      graphics  grDevices  utils      datasets  methods   base
```

other attached packages:

```
[1] mediation_4.5.0 sandwich_3.1-0 mvtnorm_1.2-4  MASS_7.3-60.2  
[5] lme4_1.1-35.3  Matrix_1.7-0
```

loaded via a namespace (and not attached):

```
[1] utf8_1.2.4      generics_0.1.3  lpSolve_5.6.20  stringi_1.8.3  
[5] lattice_0.22-6  digest_0.6.35   magrittr_2.0.3  evaluate_0.23  
[9] grid_4.4.0      fastmap_1.1.1   jsonlite_1.8.8  backports_1.4.1  
[13] nnet_7.3-19     Formula_1.2-5   gridExtra_2.3    fansi_1.0.6  
[17] scales_1.3.0    cli_3.6.2       rlang_1.1.3     munsell_0.5.1  
[21] Hmisc_5.1-2     splines_4.4.0   base64enc_0.1-3  yaml_2.3.8  
[25] tools_4.4.0     checkmate_2.3.1 htmlTable_2.4.2  nloptr_2.0.3  
[29] minqa_1.2.6     dplyr_1.1.4     colorspace_2.1-0 ggplot2_3.5.1  
[33] boot_1.3-30     vctrs_0.6.5     R6_2.5.1        rpart_4.1.23  
[37] zoo_1.8-12      lifecycle_1.0.4 stringr_1.5.1    htmlwidgets_1.6.4  
[41] foreign_0.8-86  cluster_2.1.6   pkgconfig_2.0.3  pillar_1.9.0  
[45] gtable_0.3.5    data.table_1.15.4 glue_1.7.0       Rcpp_1.0.12  
[49] xfun_0.43       tibble_3.2.1    tidyselect_1.2.1 rstudioapi_0.16.0  
[53] knitr_1.46      htmltools_0.5.8.1 nlme_3.1-164     rmarkdown_2.26  
[57] compiler_4.4.0
```

I would suggest that the authors include their result of `sessionInfo()` as a comment in their R script, which would be a minimal solution to document their software versions.

2. Removing profiles that fail the manipulation check in the main experiment

The authors report the following with respect to the analyses of the main experiment (page 16):

Pairs of profiles whose perceived abstractness do not differ significantly by condition will be excluded from the analysis.

I do not think that this choice is a good idea, but I am not completely sure about that: The authors already use their preliminary survey (which uses a different sample of participants) to select appropriate profiles. If (in the main experiment) they again exclude profiles that fail the manipulation check (in the main experiment), I would worry that this limits the generalizability of their findings. The only plausible reasons apart from a type II error, why the profiles “work” in the preliminary survey but do not work again in the main experiment, would have to be attributed to sample characteristics of the participants in the main experiment. I totally agree

that having the manipulation check present in the main experiment is a good idea, and if they want, the authors can also add the analysis without the profiles that fail the manipulation check (in the main experiment) as an additional sensitivity analysis. However, I think that stronger claims (in the sense of a more severe test) can be made from the results of the main experiment if the preregistered analyses for H1 and H2 include all profiles that were chosen to be appropriate in the preliminary survey.

3. Decide whether testing H2 dependent on the result for H1

The authors report the following with respect to the analyses in the main experiment (page 16):

Only if H1 is supported, we will proceed to test H2 by conducting a multilevel mediation analysis on attraction with abstractness, random effects of participants as predictors, and attributional confidence as a mediator.

I do not think that it is a good idea to test mediation only if a total effect of abstractness on attraction is confirmed. Although the older literature on mediation claims the contrary, it is theoretically possible for the indirect effect to perfectly cancel out the direct effect, which could result in a pattern where the total effect is 0 but the indirect effect (and the direct effect) is unequal to 0. Additionally, because any decision against H1 can always be a type II error, I would strongly suggest to always report the result of the mediation analysis (H2), irrespective of the result for the total effect (H1).

Minor Issues

(1) Justification of effect sizes and variances for power analysis

Although I think the most important task is to make the assumptions for power calculations transparent (which is the case here, because the manuscript together with the provided R code includes all assumptions), the authors could give some short justification on how these values have been chosen. This includes the following parameters:

- The effect size, margins, standard deviation, and equivalence bounds for the equivalence tests of the preliminary survey.
- The ICC values (mediator and outcome model), and the effect sizes for the three path coefficients in the mediation model.

(2) Position of the manipulation check items in the main experiment

On page 14 of the manuscript, the authors write with respect to the main experiment:

Participants will answer one manipulation check item (“How abstract did you feel the profiles were?”; 7 points scale: “abstract” - “concrete”) for each profile.

Based on this description, it is not clear to me, whether this item is presented directly after each profile or at the end of the questionnaire together for all profiles. Both options probably have advantages and disadvantages. Personally, I think the stronger design would be to place all manipulation checks at the end of the questionnaire, to make sure that being asked explicitly about the abstractness of a profile does not change the response to the following profiles (because the abstractness of the profiles has been made more salient by the manipulation check of previous profiles).

(3) Document software versions

- As already mentioned earlier, I would encourage the authors to document the software versions they used when running their power analysis.
- The authors write in their manuscript that they will use R version 4.3.3 for their final analysis. Although preregistering the R version is of course fine, I think it is not absolutely necessary and I would not see it as a problem if the authors use a later R version for their final analysis, as long as they document this R version in their stage 2 manuscript.

(4) Preregister R code for equivalence tests

While the authors have provided R code for their final mediation analysis, they have not provided code on how to perform the equivalence tests in their preliminary survey, and how their power analysis for these equivalence tests has been conducted. Although not absolutely necessary, the authors could (for transparency reasons) also upload the code for running the planned equivalence tests, to ensure that there are no unspecified analysis settings that could be considered as unnecessary researcher degrees of freedom.

(5) Mention extended analysis options in a limitation section

Although I consider the preregistered analyses appropriate, I have thought about possible limitations of the current statistical approach and I want to briefly report my thoughts here:

- The authors use multilevel models with a random intercept for participants but no random intercept for profiles. I think this choice is acceptable because the *mediation* R package cannot handle multilevel model with multiple random intercepts. However, without this constraint I would consider a model with random intercepts for both participants and profiles even more appropriate. Although partly speculative, I would suspect that such an extended model might be able to control for unobserved confounding between the mediator and the outcome that is caused by attributes of the individual profiles. For this reason, it might be reasonable to include the lack of modeling random intercepts for profiles in the limitation section of the stage 2 manuscript. (*Side comment:* Running mediation analyses with more complicated models would theoretically be possible with the **brms** R package. But this would require computing the mediation effect “manually”, so I think this is not worth the effort for the current study.)

- The authors do not plan to run sensitivity analyses against potential unobserved confounding between the mediator and the outcome. I think this choice is acceptable because the `medsens()` function in the **mediation** cannot handle multilevel models. I think it might be reasonable to include the lack of sensitivity analyses in the limitation section of the stage 2 manuscript. (*Side comment:* Running sensitivity analyses with multilevel mediations models would theoretically be possible with the **brms** R package. But this would require computing the mediation effect “manually”, so I think this is not worth the effort for the current study.)

(6) Suggestions to further improve their R code

I have some minor comments on the R code in `power_analysis_H12.R` that do not affect the performance of the script:

- In line 66 and 67, the authors define the variables `VARR<-1` and `ALPHA<-0.05` that are not used anywhere else in the script.
- In line 7, the authors run `rm(list=ls())`. Although this is not problematic by itself, I just wanted to make the authors aware that [this practice is considered error-prone](#) by many R programmers, because it does not ensure a clean R session and can therefor affect reproducibility of results.

(7) Iterations for the quasi-Bayesian confidence interval

The authors report that they will run 10000 resamples for the quasi-Bayesian confidence interval of their final mediation analysis. Although more iterations are of course always better, this number sounds a bit excessive. It is so high that I could not run the analysis on my laptop with 16 GB memory. Perhaps 5000 or 2000 iterations might also be enough, but I leave this decision to the authors.

(8) Improve wording and fix small mistakes

- On page 16 of the manuscript, the authors write:

To test H1, we will include abstractness (dummy variable: abstract condition = 0, concrete condition = 1) and random effects of participants as predictors, with attraction as the dependent variable. [...] Only if H1 is supported, we will proceed to test H2 by conducting a multilevel mediation analysis on attraction with abstractness, random effects of 368 participants as predictors, and attributional confidence as a mediator.

I would suggest to use the more precise term “*random intercepts*” instead of “*random effects*” here.

- On page 14 of the manuscript, the authors write:

One item of DQS (Directed Questions Scale; Maniaci & Rogge, 2014) for each profile such as “Choose 1 in this question.” were operated in order to detect participation with paying attention (i.e., satisficers).

I think what the authors want to say is that these items are used to detect participants that do *NOT* pay attention.

- On page 11 of the manuscript, the authors write:

During stimulus selection, we will carefully consider their semantic proximity, as it may impact the perceived consistency of the target person.

I am not completely sure what the authors mean by this sentence, perhaps they could add another sentence to make clear how they plan to control for semantic proximity when selecting the stimuli.