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

Reviewer: Florian Pargent, LMU Munich

Summary

The stage 1 manuscript titled *Attraction depending on the level of abstraction of the character descriptions* (https://osf.io/7peyb) proposes an online experiment to test hypotheses derived from uncertainty reduction theory. This theory claims that people are more attracted to an individual if the perceived uncertainty about the behavior of the individual is low. The authors propose a pilot study to test stimuli which will be combined into character profiles that are rated by a sample of 500 Japanese subjects in the main study. The proposed main study shall investigate two hypotheses: H1) Subjects are less attracted to individuals that are described in online profiles with abstract sentences and more attracted to individuals described with concrete sentences. H2) The degree of confidence in how the described individual would behave mediates the effect of abstractedness on attraction.

I find the research question investigated in the proposed study scientifically valid and I think it would be feasible to design a meaningful study to provide evidence for H1 and to a lesser extend also for H2, if the study focuses on a clearly specified context: How the abstractedness of *profiles on social media platforms* influences how attracted people feel towards the individual described in the profile. However, in its current form I think the proposed study is not strong enough to provide enough evidence to answer the hypotheses of the researchers. I outline my comments on what I think would be the most important aspects on how to improve the proposed study below. For each issue I have referenced the relevant **Stage 1 submission criteria** outlined on the PCI RR homepage (https://rr.peercommunityin.org/help/guide_for_authors).

Major Issues

(1) The Methods section is not detailed enough

Submission criteria: **1D**

My biggest issue which is also related to most of my other comments is that the manuscript is not transparent enough and does not include enough information on the proposed study for a Stage 1 registered report. Most importantly, the Methods section is not detailed enough to replicate the proposed study based on the given information.

All measures used in the pilot and main study should be described in more detail: The authors mention that they measure *attributional confidence* based on Clatterbuck (1979) and *attraction* based on Montoya & Horton (2004). However, this does not precisely specify the applied scales. Clatterbuck (1979) includes both a long and a short scale. I suppose the authors refer to the 7-item scale, but when reading the items in the original publication I am not sure if all items can be used without adaptation in the context of the study. The same applies for the scale in Montoya & Horton (2004), in which case the authors mention that they will use an adaptation. All items which will be actually used should be included in the Stage 1 manuscript (or in supplemental materials uploaded to the OSF repository). The authors should also make sure that the intended item response format is described in enough detail. Currently it is not clear, whether all response options will be labeled or only the endpoints of the 7-point scale. In addition to the two scales, all other questions included in both the pilot and the main study should be described with the exact wording. This includes the questions on rating the abstractness and the favorability of acts in the pilot study, all questions on demographic variables, and the concrete DQS questions. I would suggest to add a dedicated measures section to the manuscript where all this information is given and to add all item wordings and response scales to the OSF repository.

Moreover, it must be better described how the final profiles will be visually presented in the main study. I would suggest to describe the appearance of the final profiles in a dedicated section and upload all materials to the OSF. Currently, it is not clear at all how the final profiles will look like, whether they include more stimuli than the manipulated sentences described in the manuscript, whether additional information will be displayed (e.g., does the person described in the profile have a name or will be a generic picture included). The concrete instructions for the participants that will rate the profiles in the main study should also be described.

(2) Important assumptions of causal mediation analysis are not discussed

Submission criteria: **1A**, **1B**, **1C**

Recent methods literature criticize how causal inference (e.g. mediation analysis) is performed in psychological research (e.g., Rohrer et al. 2022; Rohrer 2018), and emphasize the untestable assumptions these analyses rely on. I think it is very important that the authors discuss the causal assumptions necessary to test the hypothesized mediation effect in the proposed main study with as little bias as possible. To my understanding, the most crucial assumption in a design such as this (where the independent variable is randomized but the mediator is not) is the absence of post-treatment confounders that have a causal effect on both the mediator and

the outcome. I have visualized the causal DAG of this setting in Figure 1. In the current study, the mediator and the outcome are scores where different raters rate different profiles. Thus, both rater and profile characteristics that have a causal effect on both attributional confidence and attraction would bias the estimation of the mediation effect. I think it might be possible to rule out most profile effects by experimental control of the stimuli. However, I think the absence of confounding rater characteristics is unrealistic: Without being an expert on attraction research, I could imagine a number of vaguely defined personality traits that might result in raters perceiving both less confidence in the behaviors of others and less attraction to others in general. When such rater characteristics (which cannot be ruled out and are not affected by the randomization of the independent variable) are present, the analysis might find a mediation effect even if there is no causal effect of attributional confidence on attraction. I will give a short example based on a simple simulation in the Appendix. To summarize, I think that a better description of the theoretical estimand (i.e., the investigated total and mediation effects) of the study and a discussion on how to estimate those effects most precisely would greatly strengthen the theoretical and methodological foundation of the study.

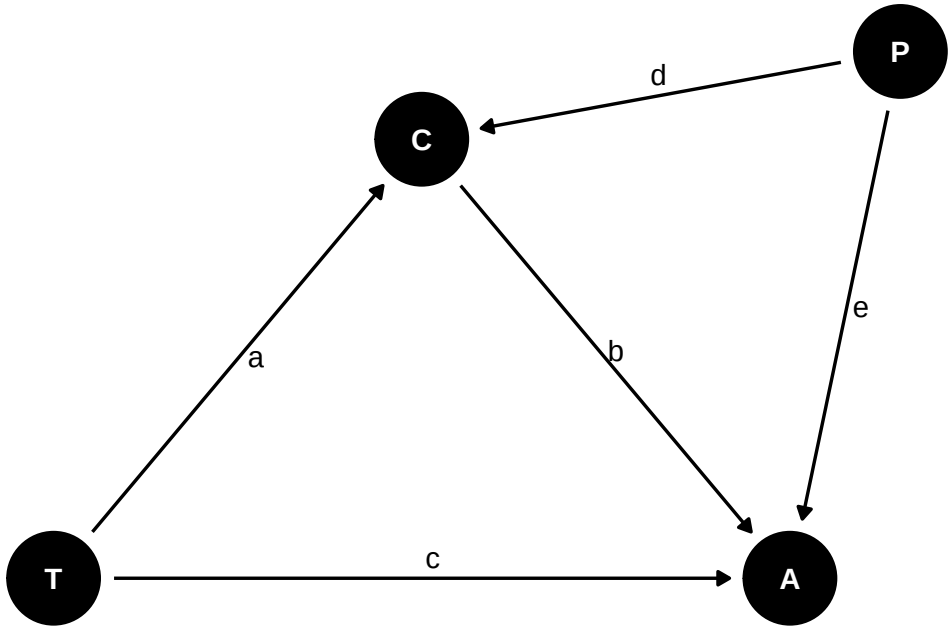

Figure 1: DAG of a simple mediation model with post-treatment confounder. T: treatment (concrete vs. abstract), C: mediator (attributional certainty), A: outcome (attraction), P: post-treatment confounder related to the rater (e.g., paranoia).

3) Aggregated data analysis might not be appropriate because of potential post-treatment confounders

Submission criteria: **1C**

The problem of post-treatment confounders described in the previous comment is related to the

question on whether the aggregated data analysis proposed in the manuscript is appropriate. Although the design consists of ratings made by several raters of several profiles, the authors plan to aggregate both attributional confidence and attraction by rater and perform data analysis on these aggregated scores. In contrast, it would be perhaps more appropriate to analyze the individual ratings by using multilevel models. I have put some thoughts into whether it matters which analysis to choose based on a simple simulation shown in the Appendix. I think the aggregated analysis can be used to estimate the total effect of the abstractedness manipulation on attraction. Thus, testing H1 should be possible to answer with the simple t-test proposed by the authors (although I would suspect higher power by using a multilevel model). To estimate the hypothesized mediation effect in H2, the situation is more difficult. I think the aggregated analysis should still be fine if we assume that there are no unobserved post-treatment confounders (as demonstrated in my simulated example). In contrast, if some personality traits of raters effect both the mediator and the outcome (which I think is highly plausible), the aggregated mediation analysis should always be biased. However, I think it might be possible to control for rater characteristics (at least in some simple scenarios) by using multilevel mediation analysis. As demonstrated in my simulated data example, it seems theoretically possible to estimate the correct mediation effect despite a post-treatment confounder. I think this might be a good argument to use multilevel mediation to investigate H2 (and in that case also using multilevel models for H1 would probably be more elegant).

(4) The number of different profiles might be too small

Submission criteria: **1C**

The possibility to control for post-treatment confounding with multilevel mediation seems to rely on a sufficient number of rated profiles per person. As can be shown with my simulated example in the Appendix, when confounding is present the mediation effect can only be estimated without bias if each rater rates a large enough number of profiles. I think, the proposed number of only three different profiles will not be enough to profit from the possibility to control for confounding with multilevel models. For this reason, I suggest that the authors think about increasing the number of profiles in their main study.

(5) Generalizability and ecological validity is questionable

Submission criteria: **1A**, **1B**, **1C**

The optimal number of profiles is also related to the question on the expected generalizability of the proposed study. Recent work has criticized that many psychological studies are not designed and analyzed in a way that justifies the intended generalizable claims (Yarkoni 2022). The current study hypothesizes that the abstractedness of profiles on social media platforms has a causal effect on how attracted people are towards the person described in the profile. However, the proposed study design relies on only three examples of such profile. Yarkoni (2022) calls this *"the stimulus-as-fixed-effect fallacy"*. I think it is highly plausible that even if abstractedness has on average a causal effect on attraction, this effect might vary between profiles (for some profiles the effect might even be positive). If only a small number of profiles is

included in the study, there is a high chance that the average effect across the small sample lies far away from the true average effect in the population of all relevant profiles. For this reason, I think including a larger sample of profiles would increase the generalizability of the proposed study in addition to the additional advantages of controlling for post-treatment confounding mentioned in the previous section. Note that multilevel models could be used to explicitly model profiles with varying effects, when the number of profiles is not too low.

Another important issue I have with the generalizability of the proposed study in its current form is the ecological validity of the profiles presented in the main study. In their introduction, the authors state that they want to investigate the hypothesized effect of abstractedness on attraction in the context of online social media platforms. However, they do not describe the intended setting any further, which greatly limits the potential generalizability of their findings. It is not clear to me whether the virtual profiles intended to be included in the main study are representative for actual profiles on social media platforms. Because the authors currently only give a small number of examples of act descriptions and they do not describe how the complete profile will be presented, the reader cannot asses whether the study findings might be transferable to actual data from social media. Based on the current examples included in the manuscript (e.g., "I belonged to a basketball club in my school days."; "He helped an elderly person."), I am skeptical that such descriptions are representative for real online profiles on social media platforms. I think to make the study stronger, the authors should pick a concrete real live setting (e.g., a social media platform like Instagram or a dating platform like Tinder) and aim to construct their profiles as representative for the chosen setting as possible (without disregarding important aspects of experimental control). I want to emphasize again that the concrete materials and a visual example of how the final profiles will be presented should be included in the OSF repository to strengthen the Stage 1 manuscript and increase both the generalizability and the reproducibility of the research.

Another relevant aspect of generalizability might be cultural differences in the psychological variables of interest (Deffner, Rohrer, and McElreath 2022). Because my knowledge on both attraction research and Japanese culture is limited, I was wondering whether Japanese people differ from people in North America or Europe with respect to how they describe themselves in social media platforms, how favorable they assess certain behaviors, or how they perceive or describe attraction towards people they meet online. A discussion on documented cultural differences on these topics might enable readers to better assess how well the findings of this study can be generalized to populations from different countries.

(6) Inadequate randomization to experimental conditions

Submission criteria: **1C**

The authors propose to ask the participants in the main study about their month of birth and randomize the experimental condition based on whether the month of birth is an even or uneven number. This randomization is not appropriate: First, birth months are not completely random because pregnancies can be planned (at least to a certain degree). Second, relying on

such imperfect randomization is totally unnecessary because all software packages for online surveys have implemented functionality for true randomization.

(7) Inadequate use of two-sided hypothesis tests

Submission criteria: **1C**

The authors propose two-sided hypothesis tests throughout their manuscript although they have directed theoretical hypotheses (e.g., lower abstractedness is supposed to lead to higher attraction). I think the authors should always perform one-sided tests in a preregistered study with directed hypotheses.

Minor Issues

(8) Sample size justification and description of mediation analysis is not detailed enough

Submission criteria: **1C**

For H1, the authors provide a sample size justification based on power analysis. Although I could reproduce their result in R (see below), the R code should be included in the OSF repository to show how sample size was calculated.

```r
library(pwr)
pwr.t.test(d = 0.3, power = 0.8, sig.level = 0.05,
  type = "two.sample", alternative = "two.sided")
```

```
     Two-sample t test power calculation

              n = 175.3847
              d = 0.3
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

For H2, the authors cite another mediation study to justify their sample size:

> "For H2, as a previous study showed that a sample size of 462 was needed to detect a small mediation effect with 1 - β = .80 (Fritz & MacKinnon, 2007), we decided the number of participants as 500."

Although I might agree that a sample size of 500 raters seems enough (at least if the number of profiles is increased) I think this justification is too weak and should be revised (see for example, Lakens 2022).

The authors should describe exactly how they plan to run their analyses in more detailed. Especially for the mediation analyse, the `mediation` package has several options which were not specified in them manuscript. I suggest to include the R code that will be used to investigate H1 and H2 in the OSF repository.

(9) Discussion of manipulation tests of the pilot study

Submission criteria: **1B**

The authors propose to test in the pilot study whether i) the rated abstractedness differs, and ii) the rated favorability of acts does not differ between sentences describing persons. However, the authors do not really discuss why these checks are performed and how they are connected to the theoretical assumptions of their causal model. Aspect i) is straightforward and seems like a reasonable manipulation check to ensure that the abstractedness manipulation works and thus abstractedness has the potential to influence attributional confidence (i.e., increase the plausibility that path $a > 0$). Aspect ii) seems less clear to me. One might argue that favorability of acts is another mediator besides attributional confidence. By removing profiles that differ with respect to this potential other mediator, the study will be designed in a way to increase the ratio of the total effect which can be mediated by attributional confidence. I am also not sure how favorability of acts is related to attraction in the theoretical model. Both constructs seem somewhat similar and the authors should discuss the theoretical differences. The potential problem I see is that the favorability ratings and attraction ratings might measure more or less similar constructs. If this were true, including only sentences with comparable favorability would lead to a reduction of the total effect to (almost) 0, which would ruin the intention of the main study. I think a deeper discussion of these aspects might be useful to strengthen the manuscript by clarifying the rationale of the manipulation tests of the pilot study.

(10) Include manipulation check in main study

Submission criteria: **1E**

Although the authors plan to test their stimulus material in their pilot study, it might be a good idea to also include a short manipulation check at the end of the main study, to ensure that the abstractedness manipulation worked in the final setting. As I understand it, the complete profiles and how the sentences are displayed will differ between the pilot and the main study.

(11) Literature section is a bit short

Submission criteria: **1B**

The introduction section discussing the literature around the hypothesized effect seems a bit short. Although I am no expert on attraction research, I would suspect that there are also theories that postulate that more vague information about a person might also lead to increased attraction (e.g., if the described person is perceived as "mysterious"). A more detailed introduction will strengthen the manuscript if it discusses the exact setting in which the uncertainty reduction theory should apply and make sure that the proposed study is as representative for this setting as possible.

(12) Some sentences are hard to understand

The following sentences were hard for me to understand, or contain grammatical mistakes:

- "The authors concluded that the increase in information decreased attraction since it clarified unfamiliarity, although it dissolved dissimilarities."
- "Contrarily, Reis et al., (2011) pointed out that the selection of items by Norton et al. in their study, deviated from the existing situation and conducted an experiment depicting a real interaction situation."
- "In their experiment, participants were paired and asked questions which written on cards to each other, as in a daily conversation."
- "In the main experiment, we will examine the effect of the abstractness of expression in a sentence that describes the target person's act on their attraction."
- "In addition, the act referred to in an abstract sentence semantically connotes what the opposing concrete sentence indicates."
- "We will set a significant level (α) as .05 in the analyses."
- "As d = 0.2 is considered a small effect size in behavioral sciences (Cohen, 1988), we expect that the TOST with $\Delta = 0.1$, could reduce the error sufficiently."

Appendix: A small data simulation to demonstrate the effect of post-treatment confounding and the advantage of multilevel mediation analysis

> **Note**
>
> The following simulation is just a small toy example to demonstrate the problem of post-treatment confounding and the potential of multilevel modeling. The data generating model is overly simplistic and the assumed effect sizes are not representative for actual psychological effects.

I simulate a simple linear causal model visualized by the DAG in Figure 1. In the simulation, there is **a direct causal effect** of the abstractedness treatment T on attraction A ($c = 1$),

but **no mediation effect** because there is no causal effect of attributional certainty C on A ($b = 0$). However, there is a post-treatment confounder P. In this toy example, P stands for an imagined personality characteristic called *paranoia*, which is just a placeholder for any potential unobserved confounder. Here, people with high scores on P in general perceive less attributional certainty and less attraction with respect to other people ($d = -1$ and $e = -1$).

```
set.seed(1)

library(data.table)
library(lme4)
library(mediation)

# make assumptions for causal effects of interest
a <- 1
b <- 0
c <- 1
d <- -1
e <- -1

# simulate data
R <- 500 # number of raters
P <- 15 # number of profiles
N <- R * P # total number of ratings
rater <- rep(1:R, each = P)
profile <- rep(1:P, times = R)
# randomly assign raters to experimental conditions: 1 = concrete, 0 = abstract
treatment <- rep(sample(c(0, 1), size = R, replace = TRUE), each = P)

# simulate paranoia scores per rater
paranoia <- rep(rnorm(n = R), each = P)
# simulate certainty scores per rating
certainty <- rnorm(n = N, mean = a * treatment + d * paranoia)
# simulate attraction scores per rating
attraction <- rnorm(n = N, mean = b * certainty + c * treatment + e * paranoia)

# dataset in long format
dat_long <- data.table(treatment = treatment, rater = rater,
  profile = profile, certainty = certainty, attraction = attraction,
  paranoia = paranoia)
# dataset in wide format (aggregate certainty and attraction scores per rater)
dat_aggr <- dat_long[,.(treatment, paranoia,
  M_certainty = mean(certainty), M_attraction = mean(attraction)),
```

```
    by = rater]
```

H1) Estimate the total effect

The total causal effect of abstractedness on attraction can be tested with a simple **t-test** on the **aggregated data** as proposed in the manuscript.

```
  ## based on t-test
  t.test(M_attraction ~ treatment, data = dat_aggr, alternative = "less",
    paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
    Welch Two Sample t-test

data:  M_attraction by treatment
t = -39.036, df = 7417.1, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is less than 0
95 percent confidence interval:
       -Inf -0.9588193
sample estimates:
mean in group 0 mean in group 1
     0.01466689      1.01567026
```

Using **linear regression** on the **aggregated data** instead of the t-test directly shows that the model correctly estimates a total effect of about 1.

```
  ## based on linear regression
  summary(lm(M_attraction ~ treatment, data = dat_aggr))
```

```
Call:
lm(formula = M_attraction ~ treatment, data = dat_aggr)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9474 -0.7880  0.0261  0.6851  3.0315

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  0.01467    0.01789    0.82    0.412
treatment    1.00100    0.02561   39.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.109 on 7498 degrees of freedom
Multiple R-squared:  0.1693,    Adjusted R-squared:  0.1692
F-statistic:  1528 on 1 and 7498 DF,  p-value: < 2.2e-16
```

A more elegant way to estimate the total effect would be to use a **multilevel model** on the **unaggregated data**, which includes a random intercept for each rater. Such a model could also be extended to model more complicated random effects for both raters and profiles.

```r
## based on multilevel model
summary(lmer(attraction ~ treatment + (1 | rater), data = dat_long))
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: attraction ~ treatment + (1 | rater)
   Data: dat_long

REML criterion at convergence: 22553.1

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7729 -0.6502 -0.0130  0.6674  3.7100

Random effects:
 Groups   Name        Variance Std.Dev.
 rater    (Intercept) 1.1686   1.0810
 Residual             0.9725   0.9862
Number of obs: 7500, groups:  rater, 500

Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.01467    0.06941   0.211
treatment    1.00100    0.09937  10.074

Correlation of Fixed Effects:
          (Intr)
treatment -0.699
```

H2) Estimate the mediation effect

When not controlled for in the statistical mediation model, the post-treatment confounder P has the potential to bias the estimation of the mediation effect. If we could observe the confounder P in the study, it would be possible to control for it in the **mediation analysis** based on **aggregated data**.

```
## based on linear regression
med.lm.conf <- lm(certainty ~ treatment + paranoia, data = dat_aggr)
out.lm.conf <- lm(attraction ~  treatment + certainty + paranoia, data = dat_aggr)
fit.lm.conf <- mediation::mediate(model.m = med.lm.conf, model.y = out.lm.conf,
  treat = "treatment", mediator = "certainty", sims = 2000)
summary(fit.lm.conf)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

               Estimate 95% CI Lower 95% CI Upper p-value
ACME            0.00750     -0.01416         0.03    0.49
ADE             0.98930      0.94021         1.04  <2e-16 ***
Total Effect    0.99679      0.95348         1.04  <2e-16 ***
Prop. Mediated  0.00761     -0.01442         0.03    0.49
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 7500


Simulations: 2000
```

However, when the confounder is unobserved (which is probably the case for many potential confounders in the current study), the **mediation analysis** based on **aggregated data** estimates a substantial mediation effect although the true indirect causal effect is actually 0 (because $b = 0$).

```
### based on linear regression
med.lm <- lm(certainty ~ treatment, data = dat_aggr)
out.lm <- lm(attraction ~  treatment + certainty, data = dat_aggr)
fit.lm <- mediation::mediate(model.m = med.lm, model.y = out.lm,
```

```
  treat = "treatment", mediator = "certainty", sims = 2000)
summary(fit.lm)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

               Estimate 95% CI Lower 95% CI Upper p-value
ACME              0.541        0.501         0.58  <2e-16 ***
ADE               0.460        0.401         0.52  <2e-16 ***
Total Effect      1.002        0.933         1.07  <2e-16 ***
Prop. Mediated    0.540        0.502         0.58  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 7500


Simulations: 2000
```

In contrast, **multilevel mediation analysis** on the **unaggregated data** would be able to correct for the confounding effect in this simple case, even if the confounder is unobserved and cannot be explicitly controlled for in the mediation model.

```
## based on multilevel model
med.lmer <- lmer(certainty ~ treatment + (1 | rater), data = dat_long)
out.lmer <- lmer(attraction ~  treatment + certainty + (1 | rater), data = dat_long)
fit.lmer <- mediation::mediate(model.m = med.lmer, model.y = out.lmer,
  treat = "treatment", mediator = "certainty", sims = 2000)
summary(fit.lmer)
```

```
Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

Mediator Groups: rater

Outcome Groups: rater
```

```
Output Based on Overall Averages Across Groups

               Estimate 95% CI Lower 95% CI Upper p-value
ACME             0.0771       0.0517         0.11  <2e-16 ***
ADE              0.9234       0.7374         1.10  <2e-16 ***
Total Effect     1.0004       0.8168         1.18  <2e-16 ***
Prop. Mediated   0.0766       0.0510         0.11  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 7500


Simulations: 2000
```

> **Note**
>
> Note that in the current simulation each person rates 15 profiles. If the number of profiles is too low, the estimate remains biased. You can see this if you repeat the example with $P = 5$.

References

Deffner, Dominik, Julia M. Rohrer, and Richard McElreath. 2022. "A Causal Framework for Cross-Cultural Generalizability." *Advances in Methods and Practices in Psychological Science* 5 (3): 251524592211063. https://doi.org/10.1177/25152459221106366.

Lakens, Daniël. 2022. "Sample Size Justification." *Collabra: Psychology* 8 (1): 33267. https://doi.org/10.1525/collabra.33267.

Rohrer, Julia M. 2018. "Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data." *Advances in Methods and Practices in Psychological Science* 1 (1): 27–42. https://doi.org/10.1177/2515245917745629.

Rohrer, Julia M., Paul Hünermund, Ruben C. Arslan, and Malte Elson. 2022. "That's a Lot to Process! Pitfalls of Popular Path Models." *Advances in Methods and Practices in Psychological Science* 5 (2): 251524592210958. https://doi.org/10.1177/25152459221095827.

Yarkoni, Tal. 2022. "The Generalizability Crisis." *Behavioral and Brain Sciences* 45: e1. https://doi.org/10.1017/S0140525X20001685.