

Helge Giese
Executive Editor
Health Psychology Bulletin

"Does Smokers' Self-Construal Moderate the Effect of (Self-)persuasion on Smoking?"

September 20th, 2018

Dear Mr. Giese,

We would like to thank you for the possibility to revise and resubmit our manuscript to *Health Psychology Bulletin*. We appreciated the helpful suggestions from both the editor and the reviewers very much, and we hope we have satisfactorily answered all questions and incorporated all suggestions for improvement.

To provide an outline of the changes in the manuscript, we first quote the comment of you or the reviewers (font: Courier New), and then describe what we did to improve the manuscript in respect to the specific point in question (font: Helvetica, in *italics* when citing new parts in the manuscript). The page numbers refer to the revised version of the manuscript.

We hope that the manuscript now meets the journal's requirements.

Yours sincerely, on behalf of all authors,

Shuang Li
Behavioural Science Institute
Radboud University
6500 HE Nijmegen, The Netherlands
E-mail: sli@psych.ru.nl
Phone: +31 684663515

Reviewers' comments:

Editor:

1. Mention the link to the replication package in text and not in supplements (as you do with the registered studies).

We now mention this at the end of the Introduction part (page 8):

“A full disclosure package of all the studies can be found at https://osf.io/u76hb/?view_only=81179953ca634b7389a5d049b9c1d191.”

2. Please report minimal p and maximal effect size instead of writing $p > 0.05$; always report effect sizes along with p values.

In the revised manuscript, we changed the p-values as requested, and we added effect size to each p-value.

3. The manuscript would also benefit from another round of proofreading so that mistakes like (“spread in the similar way”; p.15) do not occur. Specifically for this paragraph, please be more specific how you sampled.

We have proofread the revised version of the manuscript and sent it to a professional proofreading company PRS to correct all language mistakes.

The study link in Study 2 was distributed in the exact same way as in Study 1 (Study 2 was programmed in Google Forms and disseminated via the experiment platform SONA, the experimenter's and colleagues' social networks (e.g., Facebook), and via flyers handed out to smokers on Radboud University (RU) campus). To cut down the word counts, we mentioned in the Method part of Study 2 (page 14):

“Study 2 was programmed in Google Forms and disseminated in the same way as in Study 1.”

If the editor wants to see it in a more detailed way in the manuscript, we would be happy to repeat it.

4. In Study 2, the model phrased as “ANCOVA” needs further explanation: Typically, no interactions with covariates are assumed. Thus, a statement of the exact used model would be helpful (as you have two control variables in addition to a variable for which you want to test interaction effects). In

the syntax, you do not use both covariates (in addition to the typo) as you indicate in the text (last paragraph p.16).

We changed ANCOVA into MANCOVA, as MANCOVA is more appropriate for analyzing the data which contains 3 correlational dependent variables.

We agree that it might have been misleading to look at the interaction between the independent variable and a covariate without clarifying this in the manuscript. Now, the model has been explained in the text (page 16):

“To test our hypotheses, a MANCOVA was conducted on attitude, risk perception, and intention to limit smoking, with persuasion as the factor, and the two aforementioned control variables and self-construal as covariates. Specifically, the interaction between persuasion and self-construal was included in the model to investigate the assumed moderation effect.”

The syntax is now corrected and corresponds to the analysis reported.

5. Likewise, in Study 4 & 5, again, more details on the models should be provided (how the controls were added). Why was a MANCOVA not performed/ reported?

Further explanations about the model were provided, and a MANCOVA was performed (page 22, page 27):

“To test whether the effect of persuasion on attitude, risk perception, and intention to limit smoking differed between the two cultures, a MANCOVA was conducted on these three dependent variables, with persuasion and culture as the independent variables and the four aforementioned control variables as the covariates.”

“To test whether primed self-construal moderates the effect of persuasion on attitude, risk perception, intention to limit smoking, and change of craving, a MANCOVA was conducted with persuasion and primed self-construal as the independent variables, and the three aforementioned control variables as the covariates.”

The syntax is now corrected and corresponds to the analysis reported.

6. In Study 4, as both US and Chinese samples are not representative and sampling procedures diverge, please be most cautious in your inferences about the population level. Differences may as well be explained by diverging sampling procedures, education levels etc. and not by culture. How do you interpret the result, that you could not find effects of persuasion vs no persuasion?

We fully agree with the editor that conclusions should be made cautiously due to the various differences between the samples. We mentioned the sampling procedures as a limitation in the Results & Discussion part of Study 4 (page 25):

“It is noteworthy to mention that divergent sampling procedures were used for recruiting U.S. and Chinese participants due to resource constraints. Although as expected in our sample, Chinese participants were more interdependent than the U.S. participants, there were many other differences between the two groups of participants. A stronger design would be to specifically identify individuals with high or low self-construal living in the same culture, or recruit Chinese/U.S. individuals living in America/China.”

We added the explanation for no difference between persuasion and control groups in the Results & Discussion part of Study 4 (page 25):

“One possible reason could be that nowadays smokers are exposed so frequently to anti-smoking persuasive messages that the messages can be easily derogated (e.g., Schüz & Ferguson, 2014). It could also be that smoking is highly addictive, and thus both attitude and behavior are hard to change based only on exposure to a persuasive message once. For example, Schüz and colleagues (2016) found that the plain cigarette packaging that only contains health warning labels does not change smokers’ intention and behavior more or less than the traditional cigarette packets.”

7. In Study 5, do you have manipulation checks?

No, as in previous research (e.g., Flinkenflogel, N., Novin, S., Huizinga, M., & Krabbendam, L., 2017: Study 1 & 2; Holland, Roeder, Van Baaren, Brandt, & Hannover, 2004; Utz, S., 2004; Van Baaren, Maddux, Chartrand, de Bouter, & Van Knippenberg, 2003), there was no manipulation check in Study 5. The typical manipulation check, the Twenty Statement Test (a task in which participants are asked to write down 20 sentences starting with “I am”) (Trafimow, Triandis, & Goto, 1991) is time and effort consuming, and we were concerned that after this check, the priming effect might fade away.

Some researchers used self-construal scale as the manipulation check (e.g., Zampetakis, L. A., Kafetsios, K., Lerakis, M., & Moustakis, V., 2015), however, we thought it has the potential problem of confounding the chronic and the situational self-construal.

Because we had no manipulation check, the ‘Similarities and Differences between Families and Friends (SSDF)’ priming task was chosen to increase the chance that priming is successful (Trafimow et al., 1991). This task is one of the most widely-used self-construal priming tasks and proven to be highly effective. In a comprehensive review about self-construal (Oyserman & Lee, 2008), a meta-analysis of the consequences of various priming manipulations showed that the strongest effects were found using the SSDF task ($d = .44$). In addition, we handled strict criteria for excluding participants who failed in writing about their similarities and differences with/from social others:

“Participants in independent self-construal priming condition are required to write down their differences and uniqueness with others in their community and society. If they deny their uniqueness and respond like “I don’t think I’m unique, I think I’m similar to other people.”, then their data will be excluded from the dataset. Participants in interdependent self-construal priming condition are required to write down their similarities with others in their community and society. If they deny their common with others and respond like “I don’t see my similarities with others, I think I’m unique and different.”, then their data will be excluded from the dataset.” (for more information about excluding criteria, please go to pre-registration page at https://osf.io/mfvkb/?view_only=894e9d18e5a24c4db854e97d29070a8b).

8. Please report the reliability of the self-construal scale where appropriate.

We added this important information. The reliability of the self-construal is now reported for each study (page 10, page 15, page 18, page 21).

Reviewer A:

Section 1a, Full Disclosure: Replication Package: Note that for any of these elements, an explanation by the authors as to why this element is omitted, also serves, that is, if the explanation is acceptable. After all, one of HPB’s goals is to help emptying the file drawers, and older studies may no longer have all these resources available.:

- Includes protocols for recruitment of participants and guiding participants through the study

- Includes templates for communications with participants (if applicable)

- Includes operationalisations of measurements (e.g. questionnaires, source code)

- Includes operationalisations of manipulations (e.g. intervention manuals, stimuli, etc)

- Includes proposal for ethical approval

- Includes confirmation of ethics approval/Institutional Review Board (IRB) approval

If applicable, please indicate here to the authors what exactly was omitted, or which resources were not provided in an open format::

Some of the studies here are registered. Analysis plans for those are included as part of the registration.

Section 1b, Full Disclosure: Analysis Package: Note that for any of these elements, an explanation by the authors as to why this element is omitted, also serves, that is, if the explanation is acceptable. After all, one of HPB’s goals is to help emptying the file drawers, and older studies may no

longer have all these resources available.:

- Includes raw data
- Includes analysis scripts used to clean/preprocess/wrangle/munge the data
- Includes cleaned data
- Includes analysis scripts used to conduct statistical analyses
- Includes textual output
- Includes graphical output

If applicable, please indicate here to the authors what exactly was omitted, or which resources were not provided in an open format.:

We rearranged the Full Disclosure Packages according to the guideline. Documents/syntax/data files required were included and categorized in different folders with names clearly indicating what were in. In addition, we added separate word files to explain what exactly was omitted and why. The Full Disclosure Package can be seen at https://osf.io/u76hb/?view_only=81179953ca634b7389a5d049b9c1d191.

Section 2a, Reviewer comments: Here, please type or paste your review. Please make sure to adhere to the editorial policy for reviewers.:

This was an interesting piece of research, looking at many options on how the predictions in the literature could have panned out in empirical settings. I have a bunch of recommendations, which the authors can follow or ignore. In a nutshell, I suggest adding a significant amount of things to the limitations section.

On syntax:

Syntax files are somewhat inconsistent. The two files that have study 1 on their name, contain analyses, while the others seem to indicate how variables were calculated. There is no syntax with study 4 on its name.

On line 34 of Syntax-SUM IN INTER SC A I RP-study 2.sps, there's a typo in a variable name and thus it seems unlikely that the syntax has been used to produce that analysis.

On my Windows machine, there seems to be a syntax file named "Syntax-Æ«îþ«ùÕøáÔÿÚÇÅsps.sps". It's somewhat unclear as to which study this relates to.

Recommendation: for clarity, rename files to begin with study

number, e.g. "Study 2_Computing variables". Or align the file names with those in the folder "NHST analysis".

Thank the reviewer's suggestion. The data files, output, and the syntax were rearranged and placed in a good manner now (see the data package https://osf.io/u76hb/?view_only=81179953ca634b7389a5d049b9c1d191).

On paper:

1. Suggest adding confidence intervals in Figure 1-2, or explaining their absence due to e.g. resource constraints.

The confidence intervals as well as data points were plotted to provide more information. Independent and interdependent self-construal were marked on the x-axis (see file "Figure").

2. Suggest including randomisation method.

We used the randomization function of Google Form and of Qualtrics to randomize our sample. In the revised manuscript, this is mentioned in the procedure and materials in the method part of each study.

3. Suggest adding an explicit note on how the sampling plan came about, and if it was "sample as many as we can" (which is ok, just needs to be clearly communicated).

In study 1, we sampled as many participants as we can, and this information is now added (page 9):

"As many daily smokers aged over 18 as possible were sampled, and in the end, 137 participants were recruited."

In the other studies, sampling was based on a-priori power analyses, the sampling plan was written in the participants and design section of each method part (page 14, 17, 20, 26).

4. Suggest clarifying the use of "explicate" in this sentence: "However, it is hard to explicate our finding that independent smokers were more vulnerable to the direct persuasive labels." Perhaps the authors meant "explain"?

Thanks for pointing our attention to this mistake. The word "explicate" has been changed to "explain" (page 14):

"However, it is hard to explain why independent smokers were more vulnerable to the direct persuasive labels."

5. Problematic sentence: "Important to mention, Study 1 was slightly under power (achieved power = 70.8%) due to the lack of a prior power analysis as prior effect sizes were not available." Stating power this way, assumes that the observed effect size exactly corresponds to the "true" effect size. As one does not know this "true" effect size, one should show a curve depicting hypothetical effect sizes on the x-axis and the achieved power against that effect size on the y-axis (see an example figure and R code under "Statistical Power" in this supplement:https://heinsonmatti.github.io/sms-persuasion/sms-persuasion-supplement.html#study_design).

The sentence continues: "Therefore, the obtained moderation effect could result from the Type-I error and needs further examination." But increased power only reduces type-II errors. Take the case of extreme power, say, .99. This means that the probability of claiming there is nothing *when there actually is something* (type-II error), is $\leq 1\%$ in the long run. But obviously, this doesn't say anything about claiming there is something *when there actually is nothing* in the long run. There's an interesting, short paper on a Bayesian's perspective to power (<http://journals.sagepub.com/doi/pdf/10.1177/1745691614551642>), and the aforementioned supplement provides R code to do the presented analysis for the full range of unknown effect sizes in the case of simple group comparison. If the author is curious, (s)he may look into that, but it's not necessary.

Recommendation: delete the aforementioned sentences, perhaps move discussion of power to the discussion. Could state, that motivation for further studies was the exploratory nature of the study (which stems from lack of pre-registered hypotheses), relatively low strength of evidence due to low number of participants and high uncertainty in the size of the expected effects (and absence of correction for multiple testing), if this is what the author meant. Of course, also possible to state, that motivation for further studies was a misconception about the effect of sample size on error rates. Either discuss the effects in line with this paper: <https://academic.oup.com/psychsocgerontology/advance-article-abstract/doi/10.1093/geronb/gby065/5033832>, or simply refer to that paper and state as a limitation. Also, I'd fix the part stating underpoweredness of Study 1 in discussion section, accordingly. Just to be clear: whether Study 1 is underpowered or not depends on the effect one wishes to detect, and 80% is no more meaningful than an alpha of 0.05.

We thank the reviewer for this detailed explanation and recommendation. We followed the advice of deleting the problematic sentence, and moved the

discussion about power to the general discussion (page 31). There we referred to the recommended articles to explain the limitations of the current research:

“Furthermore, several statistical drawbacks need to be addressed. Firstly, as Albers and Lakens (2018) stated, power analyses based on pilot data are often biased, and follow-up bias could lead to underpowered studies. In the current research, power analyses of both Study 2 and Study 3 were conducted based on the effect size observed in former studies. In addition, for these two studies, no smallest effect size of interest (SESOI) was determined (see Lakens, 2014), due to the lack of knowledge of which effect sizes are meaningful and which are not. Taken as a whole, it is still pending whether the null results obtained reflected a real nonexistent effect or whether the actual effect size is too small to detect with the potentially biased sample size. Moreover, the evidence of null hypothesis by Bayesian statistics was mostly anecdotal and moderate – not strong enough to give an assertive conclusion. Therefore, we recommend the readers to interpret the current results with caution, and it could also be helpful to calculate the equivalence bounds of the meta-analytical effect size with a mini-meta-analysis (Lakens, 2017).”

In addition, we stated that our motivation to do Study 2 was the relatively low strength of evidence due to a low number of participants and high uncertainty in the size of the expected effects. This can be found in the last paragraph of the discussion of Study 1 (page 14):

“It is important to mention that the high uncertainty about the expected effect size made us unsure of the appropriate sample size, and due to resource constraints, we had to stop data collection at a certain point. These two points make the evidence of the obtained results relatively weak. To clarify the unexpected direction of the interactions and further investigate the moderation effect of self-construal, a second study was conducted in which a different paradigm was used to induce direct persuasion and self-persuasion.”

Thanks also to the reviewer for providing the resource of R code, we will look into it for future studies.

6. On that note, this is a problematic sentence: “For participants who were exposed to the direct-persuasive health-related labels, self-construal significantly predicted the attitude towardss smoking, $R^2 = 0.16$, $F(1, 25) = 4.88$, $p = 0.037$. However, for participants who were exposed to the self-persuasive health-related labels, self-construal was not predictive, $R^2 = 0.10$, $F(1, 27) = 2.93$, $p = 0.099$ ”. The p-values (which are random variables differ by about 0.06, as does R^2 . Even if one feels entitled to use the term “predictive” in the first place, it is simply not true that something which passes the 0.05 threshold predicts and something which doesn’t, does not predict -- it’s a matter of degree. See [e.g. https://www.ics.uci.edu/~sternh/courses/210/cohen94_pval.pdf](https://www.ics.uci.edu/~sternh/courses/210/cohen94_pval.pdf).

We thank the reviewer for pointing out this statistical misconception. We have modified this sentence into (page 12):

“The effect of self-construal on attitude towards smoking was significant for participants who were exposed to the direct-persuasive health-related labels, $R^2 = 0.16$, $F(1, 25) = 4.88$, $p = 0.037$. For participants who were exposed to the self-persuasive health-related labels, the effect showed a non-significant trend, $R^2 = 0.10$, $F(1, 27) = 2.93$, $p = 0.099$.”

We also checked the rest of the manuscript to make sure no such mistakes were made in other result sections.

7. Please include in the limitations-section, that the power analyses (e.g. from Study 2: “With an estimated interaction effect size $\eta^2 = 0.106$ (based on Study 1) between persuasion and self-construal, a priori power analysis indicated 69 participants for achieving 80% power when 5% was used as the alpha level.”) are biased and refer to this paper: <https://www.sciencedirect.com/science/article/pii/S002210311630230X>.

This limitation has been added (page 31). For an elaborate response to this comment, please see R1 Comment 5.

8. Please include a mention, that G*Power was used for power calculations.

We now mention it in the Method part of Study 2 (page 14):

*“For Study 2 and all following studies, we used G*Power to conduct the power analyses (Faul, Erdfelder, Lang, & Buchner, 2007).”*

9. “... the assumed moderation effect of self-construal was not significant. One possibility could be that this effect was changed due to the differences in paradigms...” Yes, that is one possibility. Another is lack of statistical power, i.e. the effect being smaller than what can be detected in the design. Recommend including in limitations a discussion on why no “smallest effect size of interest” (SESOI) was determined. For me personally, it’s because we know very little about what size effects are meaningful and which are not.

We thank the reviewer for this suggestion and have added the following part in the limitation section (page 31). For an elaborate response to this comment, please see R1 Comment 5.

10. Pre-registration of Study 4 puzzlingly states “No specific hypothesis”. Still, hypothesis testing was used. I don’t know

what to suggest regarding this, perhaps it can be left as it is.

We see the reviewer's concern that it is tricky to apply hypothesis testing when no specific hypothesis is held. However, we still decided to leave it in the paper given hypothesis testing is a common practice used in exploratory studies, and it kept consistency in the five studies of our research.

11. Please include in limitations: fallacies of testing for baseline homogeneity after randomisation, and selecting covariates accordingly. See e.g. <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780131703> and <https://cris.maastrichtuniversity.nl/ws/files/12731740/12486596.pdf>

We thank the reviewer for this helpful suggestion. We added the following paragraph in the limitation part (page 31,32):

“Secondly, baseline homogeneity was tested after randomization, and covariates were selected accordingly in the current research. However, it has been argued that controlling for the significant baseline differences in statistical analysis can be an inappropriate practice (e.g., de Boer, Waterlander, Kuijper, Steenhuis, & Twisk, 2015; Gruijters, 2016). Some researchers suggested using all prognostic variables as covariates irrespective of whether differences on these variables were statistically significant at baseline, since the precision of the effect estimate can be increased in this way. Others expressed the worry that study results would be then largely influenced by the choice of prognostic variables (Pocock, Assmann, Enos, & Kasten, 2002). To deal with this problem, de Boer and colleagues (2015) recommended to extract the appropriate covariates from previous studies, make the decision before starting the experiment, and pre-register the covariates that will be included in the analysis. However, in the current case, previous research in the relevant domain had no consensus on the prognostic variables. Lacking pre-knowledge about the covariates, we pre-registered and tested all the potential ones at baseline and used the ones which differed significantly between conditions as covariates. Nevertheless, it can be regarded as problematic, since no sensitivity checks for different inclusion criteria of the covariates were performed. Future studies should pay more attention when selecting covariates and try to apply structural equation modeling (SEM)-based statistical approaches, which can make the measurement error associated with each factor an explicit part of the full model (for more information, see Westfall & Yarkoni, 2016).”

12. “Taking together, our studies seem to suggest that it is quite unlikely that self-construal moderate the effect of persuasion on smoking-related outcomes.” I would say that it's 'quite unlikely' that the effect size is so large it can be reasonably studied with these sample sizes.

This is perhaps not a reasonable sentence: “In contrast, the current research provided sufficient power by using large sample sizes according to the prior power analyses, with altogether 1063 daily smokers who participated.” The evidence of was mostly anecdotal, not being able to detect differences

in one way or another. Consider adding to the limitations, that given the resources, these studies should/could have been included in a mini-meta-analysis to figure out the equivalence bounds of the meta-analytical effect size (<https://www.ncbi.nlm.nih.gov/pubmed/28736600>).

Considering the scope of the current research and the authors' limited statistical expertise in conducting meta-analyses, we did not conduct the mini-meta-analysis. However, we followed the reviewer's advice to add this to the limitation part (see R1 Comment 5):

Additionally, we deleted the sentence "Taking together, our studies seem to suggest that it is quite unlikely that self-construal moderate the effect of persuasion on smoking-related outcomes.", and the second sentence mentioned has been modified to emphasize the diverse population of the current research instead of the sufficient power (page 29, 30):

"The current research was conducted online with 1063 daily smokers in total, aged from 18 to 73. This diverse population helps to generalize the research outcomes to a larger group rather than college students only."

13. Suggest adding a citation or revising this sentence:
"First, the secret of self-persuasion is decreasing targets' resistance by making them unaware that they are under influence."

The sentence has been revised and several citations were added (page 30):

"First, the current research might fail in covering the influence attempt in the self-persuasion manipulation. Since the awareness of influence could undermine its impact (e.g., Strack, Schwarz, Bless, Kübler, & Wänke, 1993; Williams, Fitzsimons, & Block, 2004), research studying self-persuasion effect often manipulated it in a subtle way or under a cover story so that participants could experience the information as their own thoughts and rely on these thoughts in arriving at a judgment (Müller et al, 2009; Müller et al, 2016)."

14. "According to our findings, both statements such as 'Smoking causes fatal lung cancer' and questions such as 'What are the effects of smoking to your lungs?' seem not to be sufficient for successful smoking preventions. This suggests that the warnings currently available on cigarette packs are not exploited for the maximal effectiveness, and more effective messages should be developed for smokers' welfare." Recommend revising. The sentences are logically inconsistent, as it may well be that the effect is maximal for the medium, but it's not enough for it to be large enough to be detected with these samples.

We agree that the effect is possibly not big enough to be detected with our samples. A small expected effect size ($\eta^2=0.03$ (Cohen et al, 2003)) was used to calculate the

sample size in Study 5. However, the actual effect η^2 could still be smaller than 0.03, while in that case then it is indeed too small to be interesting for us. Taken the reviewer's advice into account, we deleted the sentences and mentioned in the discussion part that the current results should be replicated in future studies with more statistical power (page 32):

“To conclude, the current research initiates the investigation of (self)-persuasion in a cultural framework and found inconclusive evidence about the moderating role of self-construal. We think our preliminary results need to be replicated in future studies with more statistical power.”

15. Consider adding in limitations: lack of sensitivity checks for different inclusion criteria, and controlling for covariates (which is surprisingly difficult; e.g. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152719>, <https://www.ncbi.nlm.nih.gov/pubmed/11261398>).

We thank the reviewer for the provided articles. After reading <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152719>, we realized that the inclusion of additional imperfectly measured covariates would generally make detection of incremental validity more difficult, and we did not include sensitivity checks for inclusion criteria of the covariates. The authors (Westfall & Yarkoni, 2016) suggested avoiding the high Type I error rate by using SEM, which requires much bigger sample size, especially when the unique contribution of the construct of interest is relatively small. Thus, we think SEM is hard to be conducted with the current sample size. However, we added this point as one of the limitations (for details, please see R1 Comment 11).

16. Important inclusion to limitations: (“Due to resource constraints”?) the authors could not comply with the WAMBS checklist (When to worry and how to Avoid the Misuse of Bayesian Statistics), <https://www.ncbi.nlm.nih.gov/pubmed/26690773>. If they wish, this is also an opportunity to discuss using default priors.

Thanks for this informative article, which helps us to understand Bayesian statistics more deeply. We used default prior mainly due to two reasons: one is the lack of knowledge about how to set up different prior and we wanted to avoid making big mistakes; the other is that we had no strong pre-data knowledge to anticipate the effect size, which made it difficult to use alternative subjective priors. According to Wagenmakers, “the default settings are meant to serve as an “objective” specification that can be used across a wide range of different scenarios (<http://forum.cogsci.nl/discussion/2642/default-priors-for-bayesian-model-comparison-are-they-ready-to-use-without-further-user-input>)” and “the default priors provide a good reference point and seem to work well enough for the applications that have been encountered, and provide a useful alternative solution to the “ $p < .05$ ”

summary (<https://forum.cogsci.nl/discussion/4092/reviewer-wants-justification-for-the-default-prior>)". We hope this is a satisfactory justification of using default prior.

Since Bayesian was used as an additional analysis to frequent testing and considering the length of the paper, we chose not to discuss this issue in the text. If the reviewer would like to see the discussion in the text we are pleased to add it.

17. In addition, the authors could perhaps state, that the paper would have benefited from abiding with other reporting standards, such as CONSORT and/or JARS (<http://www.apa.org/pubs/journals/releases/amp-amp0000191.pdf>) (But this is just for their consideration.)

We thank the reviewer very much for the advice. We believe that following these reporting standards, we can make our work even more transparent, and that can no doubt benefit both ourselves and the field. We particularly checked what we have already reported in the manuscript with the recommended standards, and found the main parts of the standards were covered. Considering the length of the current paper (and considering that reviewer 2 asked us to cut the lengths of the paper), we did not add even more information. If there are special points the reviewer would like to see added, we are happy to do so.

Minor issues:

Minor error in p-value: the test $F(1, 52) = 6.08$ is reported to produce $p = 0.016$, whereas calculating from the test statistics, one gets $p = 0.017$.

We reanalyzed the data and found the same result. Breaking down the significant three-way interaction obtained from Study 1, for participants who watched the health-related labels the interaction between persuasion and self-construal was significant, $F(1, 52) = 6.15$, $p = 0.016$. This can be found in the file "STUDY 1 Output" in the Data Package.

Typos:

- "direct persuasion would work more effective than self-persuasion"
- Capitalisation: "from western (chronic independent self-construal) and eastern cultures"
- "anecdotal evidences supporting null hypothesis from the model were found"
- "we did not find significant difference between self-persuasion and direct persuasion"

We corrected the mentioned typos.

If the authors wish to have a look, there's an excellent and easy-to-understand MOOC, which dispels many potential confusions regarding power, errors, null effects etc.: <https://www.coursera.org/learn/statistical-inferences>

We thank the reviewer very much for providing us with this excellent material. And again, we would like to express our sincere gratitude to the reviewer for his/her time and detailed suggestion on this manuscript. The comments have inspired us a lot.

Reviewer B:

Section 1a, Full Disclosure: Replication Package: Note that for any of these elements, an explanation by the authors as to why this element is omitted, also serves, that is, if the explanation is acceptable. After all, one of HPB's goals is to help emptying the file drawers, and older studies may no longer have all these resources available.:

- Includes protocols for recruitment of participants and guiding participants through the study

- Includes templates for communications with participants (if applicable)

- Includes operationalisations of measurements (e.g. questionnaires, source code)

- Includes operationalisations of manipulations (e.g. intervention manuals, stimuli, etc)

- Includes analysis plan

- Includes proposal for ethical approval

- Includes confirmation of ethics approval/Institutional Review Board (IRB) approval

If applicable, please indicate here to the authors what exactly was omitted, or which resources were not provided in an open format.:

Section 1b, Full Disclosure: Analysis Package: Note that for any of these elements, an explanation by the authors as to why this element is omitted, also serves, that is, if the explanation is acceptable. After all, one of HPB's goals is to help emptying the file drawers, and older studies may no longer have all these resources available.:

- Includes raw data

- Includes analysis scripts used to clean/preprocess/wrangle/munge the data

- Includes cleaned data

- Includes analysis scripts used to conduct statistical analyses

- Includes textual output

Includes graphical output

If applicable, please indicate here to the authors what exactly was omitted, or which resources were not provided in an open format::

We rearranged the Full Disclosure Packages according to the guideline. Documents/syntax/data files required were included and categorized in different folders with names clearly indicating what were in. In addition, we added separate word files to explain what exactly was omitted and why. The Full Disclosure Package can be seen at https://osf.io/u76hb/?view_only=81179953ca634b7389a5d049b9c1d191.

Section 2a, Reviewer comments: Here, please type or paste your review. Please make sure to adhere to the editorial policy for reviewers.:

This manuscript describes 5 studies examining how self-construal impacts responses to tobacco health warnings with direct- or self-persuasion messages.

1. The abstract highlights the importance of research on self-construal among participants from eastern countries. However, only Study 4 studies this population. The abstract also makes no mention of health warnings, which is the method by which direct-persuasion and self-persuasion messages were provided. The Introduction is well-written and comprehensively referenced. The authors set up the rationale for the studies well.

We follow the reviewer's suggestions and revised the abstract, deleting the country/culture part and emphasizing the role of self-construal instead. The health warnings and the argumentation task were briefly introduced.

2. The manuscript is long and although this is in part due to the four studies reported, I think the manuscript would benefit from a careful edit to cut out unnecessarily wordy sections and sentences. For example, the hypotheses are written twice (once in the Hypothesis section of the Introduction and a second time at the end of the Introduction). The second instance seems unnecessary, particularly given the length of the Introduction.

Following the review's suggestion, we deleted the description of hypothesis in the end of Introduction. Furthermore, we checked the manuscript again and cut out unnecessary sections to avoid repetition and to make the manuscript shorter and more focused.

3. Can the authors comment on the extent to which formulating warning labels as questions (i.e. self-persuasion messages) actually results in participants formulating arguments (in the 'Self-persuasion versus direct persuasion' section). Indeed,

the authors address this in Study 3 where they also measure cognitive effort, but paradoxically find that effort is higher among participants in the direct persuasion task. The extent to which forming messages as questions results in self-persuasion should be discussed in the limitations section.

We thank the reviewer for this comment and would like to add several clarifications regarding this point. We used open question warnings to induce self-persuasion in Study 1 and Study 5, since it is a common method in the self-persuasion literature (Bernritter et al. 2017; Glock, Müller, & Ritter, 2013; Müller et al., 2016). The assumption behind is that arguments can automatically be elicited when the receiver reads the question. It was supported by Loman and colleagues' research (2018) that people who had been exposure to the open-question indeed reported more related arguments when they were asked to list their thoughts while viewing the question than when presented with statements.

Furthermore, we would like to highlight that we examined another paradigm (i.e., reading or writing a paragraph regarding the negative points of smoking) in Study 3. No difference was found between the two paradigms but cognitive effort was counter-intuitively found higher in direct persuasion condition. However, we would like to clarify that whether question-warnings can elicit self-generated arguments is a different question than whether people put more cognitive effort into generating the arguments (though we intuitively expected this, too). For example, researchers found self-generating few arguments is more effective than generating many arguments in promoting helping behavior (Müller, van Someren, Gloudemans, van Leeuwen, & Greifeneder, 2017). It could be that in the situation where less cognitive effort is applied, people think of arguments with high quality. Fitzsimons and Williams (2000) also suggested that the effect of intent questions on subsequent behavior is primarily the result of automatic instead of effortful processing. Therefore, we think putting less cognitive effort does not necessarily mean that no arguments are generated.

4. Can the authors provide more information about Google forms and whether this allowed for question order to be randomised.

We provided the detailed information about the way we randomized participants using Google Forms (page 10, 11):

"We designed an item asking participants to click on the number which appears at the top of the option list. The numbers 1, 2, 3, 4 were given as four options and automatically raffled by Google Forms so that any number could be randomly present on the top. Each number was connected with one specific experimental condition in order to achieve randomization."

5. The Results section for Study 1 is difficult to follow.

We rewrote the result part of Study 1 to make it more readable (page 12, 13):

"To test our hypotheses, the variables of persuasion, label content, self-construal, and all the interactions among them were included in a general linear model. On attitude towards smoking, we found a significant three-way interaction between persuasion, label content, and self-construal, $F(1, 111) = 5.92, p = 0.017, \eta p^2 = 0.051$;

and a significant two-way interaction between persuasion and self-construal, $F(1, 111) = 4.18, p = 0.043, \eta p^2 = 0.036$. All the other main effects and interaction effects on attitude did not reach significance (minimal $p = 0.130$, maximal $\eta p^2 = 0.021$). Breaking down the three-way interaction, the interaction between persuasion and self-construal was only significant under the condition of health-related labels, $F(1, 52) = 8.03, p = 0.007, \eta p^2 = 0.134$. No significant interaction was found under the condition of POE-related labels, $p = 0.759, \eta p^2 = 0.002$ (Figure 1). Subsequently, post-hoc linear regression was conducted. The effect of self-construal on attitude towards smoking was significant for participants who were exposed to the direct-persuasive health-related labels, $R^2 = 0.16, F(1, 25) = 4.88, p = 0.037$. For participants who were exposed to the self-persuasive health-related labels, the effect showed a non-significant trend, $R^2 = 0.10, F(1, 27) = 2.93, p = 0.099$.

Similarly, we found a significant three-way interaction between persuasion, label content, and self-construal on risk perception, $F(1, 111) = 4.15, p = 0.044, \eta p^2 = 0.036$. All the other main effects and interaction effects did not reach significance (minimal $p = 0.067$, maximal $\eta p^2 = 0.030$). Breaking down the three-way interaction, again the result showed the interaction between persuasion and self-construal was only significant under the condition of health-related labels, $F(1, 52) = 6.15, p = 0.016, \eta p^2 = 0.106$ (Figure 2). Post-hoc linear regression demonstrated that the effect of self-construal on risk perception was significant for participants who were exposed to the self-persuasive health-related labels, $R^2 = 0.20, F(1, 27) = 6.91, p = 0.014$. For participants who were exposed to the direct-persuasive health-related labels, the effect was not significant, $R^2 = 0.05, F(1, 25) = 1.34, p = 0.258$.”

6. Figures 1 and 2 - these would be considerably more informative if the data points for the individual participants were also plotted. The x axis should also note which is independent and which is interdependent self-construal.

The data points as well as the confidence intervals were plotted to provide more information. Independent and interdependent self-construal were marked on the x-axis (see file “Figure”).

7. For Studies 1-3, data on participants' self-construal should be provided (at the least, mean, SD and range). This is particularly important for Study 3 where the authors argue that a lack of variation in self-construal may have led to the results.

We agree and added participants' self-construal score for Study 1-4 (see p.10, p.15, p.18, p.21).

8. Study 2 - Was any attention check performed for the participants in the self-persuasion task? If this was not possible this should be added as a potential limitation.

No attention check was performed in the self-persuasion condition as previous studies have not provided such a check either. We added this as a limitation in the general discussion (page 31):

“Lastly, no attention check was performed in the self-persuasion condition. Future research could consider creating such a task for checking participants’ attention during the argument-generation process.”

9. The authors conduct statistical analyses on the baseline characteristics of participants and then control for these differences by adding variables that differ in later analyses (Studies 2 and 3). This practice is generally considered inappropriate (see <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4310023/>)

For an elaborate response to this comment, please see R1 Comment 11.

10. In Study 4, the authors use culture (USA vs China) as a proxy for self-construal. While the Chinese participants were more interdependent than the participants from the USA, there were also many other differences between these groups of participants which is an important limitation of this work. A stronger design would have been to specifically identify individuals with high or low self-construal living in the same culture (and importantly with the same culture around smoking), or perhaps recruit Chinese individuals living in the USA. This should be discussed in more detail in the limitations section for this study.

We thank the reviewer for this very helpful comment and added this limitation at the end of Study 4 (page 25):

“It is noteworthy to mention that divergent sampling procedures were used for recruiting U.S. and Chinese participants due to resource constraints. Although as expected in our sample, Chinese participants were more interdependent than the U.S. participants, there were many other differences between the two groups of participants. A stronger design would be to specifically identify individuals with high or low self-construal living in the same culture, or recruit Chinese/U.S. individuals living in America/China.”

11. Self-construal is primed in Study 5 - can the authors provide evidence for the validity of this approach?

For an elaborate response to this comment, please see Editor Comment 7.

12. General Discussion - although the results sections are slightly difficult to follow in places, I believe that the conclusions are consistent with the results. The ‘unexpected interaction’ should be described, rather than alluded to. The

General Discussion would benefit from editing to improve grammar.

Study 1 was likely under power due to its exploratory nature and the resource constraint for recruiting sufficient participants. Additionally, considering the null results from the other studies, the observed unexpected interaction effect in Study 1 was likely found by chance. We explicitly discussed this issue in the first paragraph of Discussion (page 28):

“It should be mentioned here that Study 1 was likely underpowered due to its exploratory nature and the resource constraint of recruiting sufficient participants. We therefore conducted four follow-up studies to examine whether the observed effect was found by chance and to further investigate the proposed moderation effect.”

More about the power issue was discussed on page 31 (see R1 Comment 5).

The manuscript was sent to a professional company for proofreading before resubmission.

13. The limitations section should also address the extent to which the presentation format for the health messages (presented for 15 seconds via a video) is representative of how individuals view warnings in real life.

We agree with the reviewer and addressed this point in the limitation section (page 33):

“Third, health warnings were used as persuasive messages in Study 1, Study 3, and Study 5 in order to provide the smokers with real-life information. However, ecological validity stays low as several warnings were presented sequentially, and each one was presented for a fixed time period of 15 seconds, making it not representative of how individuals view warnings in real life.”

I would like to congratulate the authors on their open and transparent approach to research and wish them luck with their revision.

We thank the reviewer again for the helpful suggestions!

Reference

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, 74, 187-195. DOI: <https://doi.org/10.1016/j.jesp.2017.09.004>

- Bernritter, S. F., van Ooijen, I., & Müller, B. C. (2017). Self-persuasion as marketing technique: the role of consumers' involvement. *European Journal of Marketing*, 51(5/6), 1075-1090. DOI: <https://doi.org/10.1108/EJM-04-2015-0213>
- de Boer, M. R., Waterlander, W. E., Kuijper, L. D., Steenhuis, I. H., & Twisk, J. W. (2015). Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, 12, 4. DOI: <https://doi.org/10.1186/s12966-015-0162-z>
- Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191
- Flinkenflogel, N., Novin, S., Huizinga, M., & Krabbendam, L. (2017). Gender moderates the influence of self-construal priming on fairness considerations. *Frontiers in psychology*, 8, 503. DOI: 10.3389/fpsyg.2017.00503
- Fitzsimons, G. J., & Williams, P. (2000). Asking questions can change choice behavior: Does it do so automatically or effortfully?. *Journal of Experimental Psychology: Applied*, 6(3), 195-206. DOI: 10.1037//1076-898X.6.3.195
- Glock, S., Müller, B. C., & Ritter, S. M. (2013). Warning labels formulated as questions positively influence smoking-related risk perception. *Journal of health psychology*, 18(2): 252-262. DOI: 10.1177/1359105312439734

- Gruijters, S. L. (2016). Baseline comparisons and covariate fishing: Bad statistical habits we should have broken yesterday. *The European Health Psychologist*, 18(5), 205-209.
- Holland, R. W., Roeder, U., van Baaren, R. B., Brandt, A. C., & Hannover, B. (2004). Don't stand so close to me: The effects of self-construal on interpersonal closeness. *Psychological Science*, 15(4): 237-242. DOI:10.1111/j.0956-7976.2004.00658.x
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses: Sequential analyses. *European Journal of Social Psychology*, 44(7), 701–710. DOI: <https://doi.org/10.1002/ejsp.2023>
- Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355-362. DOI: 10.1177/1948550617697177.
- Loman, J. G., Müller, B. C., Oude Groote Beverborg, A., van Baaren, R. B., & Buijzen, M. (2018). Self-persuasion in media messages: Reducing alcohol consumption among students with open-ended questions. *Journal of experimental psychology: applied*, 24(1), 81-91. DOI: 10.1037/xap0000162
- Müller, B. C. N., Ritter, S. M., Glock, S., Dijksterhuis, A., Engels, R. C. M. E, & van Baaren, R. B. (2016). Smoking-related warning messages formulated as questions positively influence short-term smoking behavior. *Journal of Health Psychology*, 21(1): 60-68. DOI: 10.1177/1359105314522083
- Müller, B. C. N., van Baaren, R. B., Ritter, S. M., Woud, M. L., Bergmann, H., Harakeh, Z., Engels, R. C. M. E, & Dijksterhuis, A. (2009). Tell me why... The

- influence of self-involvement on short term smoking behavior. *Addictive Behaviors*, 34(5): 427–431. DOI: 10.1016/j.addbeh.2008.12.016
- Müller, B. C.N., van Someren, D. H., Gloudemans, R. T., van Leeuwen, M. L., & Greifeneder, R. (2017). Helping Made Easy. *Social Psychology*. 48: 113-121. DOI: <http://dx.doi.org/10.1027/1864-9335/a000293>
- Oyserman, D., & Lee, S. W. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological bulletin*, 134(2), 311-342. DOI: 10.1037/0033-2909.134.2.311
- Pocock, S. J., Assmann, S. E., Enos, L. E., & Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in medicine*, 21(19), 2917-2930. DOI: 10.1002/sim.1296
- Schüz, N., Eid, M., Schüz, B., & Ferguson, S. G. (2016). Immediate effects of plain packaging health warnings on quitting intention and potential mediators: Results from two ecological momentary assessment studies. *Psychology of Addictive Behaviors*, 30(2): 220-228. DOI: 10.1037/adb0000146
- Schüz, N., & Ferguson, S. G. (2014). Australian smokers' and nonsmokers' exposure to antismoking warnings in day-to-day life: a pilot study. *Nicotine & Tobacco Research*, 17(7), 876-881.
- Strack, F., Schwarz, N., Bless, H., Kübler, A., & Wänke, M. (1993). Awareness of

the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology*, 23, 53-62. DOI: <https://doi.org/10.1002/ejsp.2420230105>

Trafimow, D., Triandis, H., & Goto, S. (1991). Some tests of the distinction between the private self and the collective self. *Journal of Personality and Social Psychology*, 60(5): 649-655. DOI: 10.1037/0022-3514.60.5.649

Utz, S. (2004). Self-construal and cooperation: Is the interdependent self more cooperative than the independent self?. *Self and Identity*, 3(3), 177-190. DOI: <https://doi.org/10.1080/13576500444000001>

van Baaren, R. B., Maddux, W. W., Chartrand, T. L., de Bouter, C., & van Knippenberg, A. (2003). It takes two to mimic: Behavioral consequences of self-construals. *Journal of Personality and Social Psychology*, 84(5): 1093-1102. DOI:10.1037/0022-3514.84.5.1093

Williams, P., Fitzsimons, G.J., & Block, L. G. (2004). When consumers don't recognize "benign" intentions questions as persuasion attempts. *Journal of Consumer Research*, 31, 540-550.

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PloS one*, 11(3), e0152719. DOI:10.1371/journal.pone.0152719

Zampetakis, L. A., Kafetsios, K., Lerakis, M., & Moustakis, V. (2015). Investigating the role of self-construal in the formation of entrepreneurial intentions. *Frontiers in psychology*, 6, 1085. DOI: 10.3389/fpsyg.2015.01085