**Peer Review and Communication History**

**MS Title**: Does Going Green Feel Good in Russia: Implicit Measurements with Visual Stimuli

**Author Names**: Danila Valko

**Submitted:** Sep 16, 2022

**Editor First Decision**: Revise & Resubmit

Nov 15, 2022

Dear Danila Valko,

I have now received two reviews of your manuscript, “Does going green feel good in Russia: implicit measurements with visual stimuli”. I also independently read the manuscript before consulting these reviews. The reviewers had mixed reactions to your manuscript. I agree that your manuscript has important strengths, such as your transparency regarding the materials. However, there are substantive issues that need to be addressed. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

The reviewers did an outstanding job in their reviews. I will highlight issues I think are particularly salient here. In your resubmission, please include a document with a point-by-point response to both the points I list here and the reviewers’ comments, outlining each change made in your manuscript or providing a suitable rebuttal.

Both reviewers point out that the interpretation of the data is questionable, and I concur. In your treatment of the data, a careful distinction between valence and emotions is needed, as is a more careful treatment of whether your data speaks to attitudes, behaviors or behavioral intentions. In a revision, these interpretations must be substantiated to warrant further consideration of the manuscript.

Futher, the results are reported with insufficient detail about the analyses methods and test values. This makes it impossible to properly assess the merit of these analyses for the reader. Please revise your results section to fully report the analyses. In addition, providing the link to your OSF preregistration and other materials in the body of the text would be a further helpful step.

In summary, I hope you find the comments of the reviewers a helpful guideline for your revision of the manuscript for further consideration at Collabra: Psychology. Please see the instructions below for submitting your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Rima-Maria Rahal

# Reviewer 1

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) | ✔ |  |  |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) | ✔ |  |  |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) | ✔ |  |  |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) | ✔ |  |  |  |  |

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The manuscript presents two studies (?), one pre-registered (Ntotal = 394). The authors propose to study the attitude of individuals (more precisely of a Russian population) towards pro-environmental behavior, with the originality of using implicit measures (i.e., AMP, AFP) instead of using explicit measures. The authors propose that individuals would have positive attitudes toward pro-environmental behaviors and their results are consistent with this hypothesis. The data are accessible and the main study was preregistered which is a strong feature of the manuscript. Furthermore, the starting point and objective of the article seems to me extremely interesting to study. However, after reading this manuscript, I have many reservations about the theory, the method and the statistical analyses performed.

# MAJOR POINTS

## Introduction

1 - First, the authors confuse emotions with affect (i.e., in this case, a positively or negatively valenced attitude). Emotions are traditionally defined as multicomponent objects (e.g., Scherer & Moors, 2019; Scarantino, 2019) that are dependent on the culture of the individuals (e.g., Barrett, 2017). The authors in this manuscript are interested in the implicit attitude of individuals towards pro-environmental behaviors. In other words, are pro-environmental behaviors perceived as positive or negative? It is a very different question whether pro-environmental behaviors elicit specific emotions or whether specific emotions promote (or not) pro-environmental behaviors. Thus, in view of the studies conducted in this manuscript, I do not think that the authors are referring to the right theoretical anchor. Furthermore, the authors begin the manuscript by stating, “Emotions are a fundamental component of human psychological functioning, yet the study of their contribution in explaining a number of behavioral facets (pro-environmental behavior included) has been traditionally neglected (Tapia-Fonllem et al., 2013).” (p. 1). This statement is at best clumsy, at worst completely wrong! Many authors have postulated that emotions have (among other things) the function of guiding behavior (e.g., Frijda, 2011; Zajonc, 1980; Zeelenberg et al., 2008) and this has been empirically tested on behaviors whether minimal (i.e., action tendencies; Chen & Bargh, 1999) or more elaborate. It has also been tested on pro-environmental behaviors (e.g., Graton, Ric, & Gonzalez, 2016). Similarly, the authors go on to say that little attention has been paid to the joint role of emotions and pro-environmental behaviors (pp. 1-2). Is this the point of this manuscript? My point is that the authors are interested here in the attitudes of individuals toward pro-environmental behaviors, not in the ability of emotions to produce pro-environmental behaviors. Both questions are interesting in terms of the psychology of emotions and environmental psychology, but they are different. Why do the authors refer to work on the links between emotions and behaviors when they are studying attitudes? The authors need to clarify this and clarify their theoretical anchors. As it stands, I do not find the introduction to be logically leading to the question and studies being conducted.
2 - The authors then discuss methods for collecting behaviors and attitudes. The authors highlight that self-reported measures are subject to social desirability. I agree with the authors on this point. That being said, I do not find the argument relevant to their studies. Indeed, after reading this paragraph, my conclusion is that it is better to do a real behavioral measure to see the influence of emotions on our pro-environmental behaviors. But the authors do not study this. They want to know if the implicit attitude associated with pro-environmental behaviors corresponds to the explicit attitude (i.e., a positive attitude). Furthermore, I think it is essential that the authors refer to work on attitudes such as the APE model (Gawronski & Bodenhausen, 2011). This is much more central than the work presented on the influence of emotions on behavior. The authors do not work on emotions or behavior, but on attitude. The literature cited here is therefore not relevant to the research objectives. In the same vein, page 4, the authors refer to IAT in a short paragraph. I think this is far more central than anything said before. The authors should emphasize the difference between explicit and implicit measures because that is the central point of the manuscript. Also, the authors say that images can decrease cross-cultural differences. I am not convinced of this (e.g., Phillips, 2019). This point should be further developed and argued if the authors find it central or simply not discussed as it is questionable. Far more centrally, I think that work and models on implicit and explicit attitudes should be the focus of the entire introduction to this manuscript, because again, the authors have not studied the emotion-behavior link as the introduction might suggest. The authors should also incorporate a discussion of current debates around the explicit/implicit, direct/indirect distinction (e.g., Corneille & Hutter, 2020).
3 - In pages 2 -3, I’m not sure the authors can say that one component of attitude is behavior. It is rather the behavioral intention. The attitude-behavior link is not always highlighted (e.g., Haddock, 1993). Furthermore, I find that the table for presenting the work is not at all appropriate. It is hard to understand the purpose of this table. It would be better to describe the work and relate it to the attitudes. I also find that a central point (i.e., cultural difference) becomes completely buried. Finally, by trying to say that Russian culture has not been studied, the authors do not defend the idea that it is different from Western cultures. Do the authors mean that there is a common background to these attitudes that might not depend on culture or are the authors looking for the specificity of Russian culture? This second interpretation seems to me to be taken with caution. The authors did not make a comparison between cultures, so the conclusion remains on the Russian culture without making a comparison with other cultures. Finally, is it important for the study of implicit attitudes that the authors conduct the study with a Russian population? Or, is it just important to adapt the tools to this population? This needs to be clarified to make the manuscript more readable.
4 - Page 5, the authors present hypotheses about specific behaviors. Why were they studying attitudes toward these particular behaviors? The authors make no mention of this before in the theory. It is difficult to get a sense of the value of studying these behaviors. The same is true for the second hypothesis. There is no mention of these effects or variables prior to the presentation of the assumptions. Overall, this is consistent with my previous comments. There is no explicit link between the theory mentioned and the hypotheses/studies.

## Method

4 - The authors start by mentioning two methods: AMP and AFP. First, it would be nice to reference the authors who developed these tasks… Also, I am not sure that the authors have performed both AMP and AFP tasks here. At least, they used extremely modified versions of them. For example, AMP (Payne et al., 2005) involves categorizing a neutral target after the presentation of a positive or negative stimulus. Thus, given the assumptions, I infer that the authors will present positive, negative, and behavior-related stimuli whose associated attitudes they want to know. However, the authors have modified the measure to transform the categorization into a form of evaluation. This does not refer to the same types of tasks or processes. This argument is the same as for AFP where the DV is the reaction time for compatible and incompatible trials. This is even more problematic for this task. Indeed, I am not sure that they can be interpreted as the authors intend. The DV should become the difference between the evaluation of the target stimuli after presentation of a neutral prime and the evaluation of the target stimuli after presentation of the other types of stimuli. My point is that DV should be a difference in evaluation and not an absolute evaluation. As it stands, I am not convinced that it is possible to make sense of this measure. Furthermore, these tasks are done with multiple trials. However, here, the authors have one and only one trial for each of the measures. So, the measure does not seem to me to be extremely reliable. I would be more convinced if the authors performed the tasks in a “classical” way and found results consistent with their hypotheses.
5 - Also, the authors should go back to the power calculations. I do not know what effect size the authors used as the average size, but taking d = 0.36 (average effect size in psychology, Lovakov & Agadullina, 2021), with MorePower (Campbell & Thompson, 2012) with a 6-modality within-subject response design for a power of .90, I find a minimum sample size of 104 subjects, not 38! In addition, the authors do a second analysis with 15 predictors. What are these 15 predictors? This is not mentioned before! I don’t understand why this analysis is done and what it refers to. The analyses should be redone and more details should be given to understand how the calculation was made and with which indicators.
6 - In addition, there is a great deal of detail missing from the procedure. Were the trials fixed or randomized? This should be included in the body of the text. When I read the procedure of a study, I should be able to reproduce the body of the study if I want to replicate observed results, for example. With the information in the manuscript, I am currently unable to do so.
7 - It would be better for the method and the results to separate the two studies. This would make the reading much easier and clearer.

## Results

8 - The authors have only reported the p-value. Besides the fact that this does not respect APA standards, this indicator is really criticized so presenting only this one give almost no information except the significance. What is the statistic? What are the CIs? What is the effect size? This makes it difficult to read the results and get an idea of the strength of the effects observed by the authors. In addition, the authors do not detail the analyses they performed. This more than laconic presentation of the results does not give us enough information to know if the statistics are really adapted and if they seem correct when read. This does not seem acceptable to me. So, I will critique the analyses and results with what I think was done and what I think would be more appropriate.
9 - Regarding AFP, I do not think the analyses should be done on reaction times considering the way the task was modified as I mentioned before. Also, if I understand correctly, the authors compared RT(Eco-positive) vs. RT(Eco-negative). I do not think these answers their question. It seems to me that Helmert contrast analysis (e.g., Walsh, 1991) is more appropriate to the question. The authors postulate that behaviors will be associated with a positive attitude so they will only differ with negative stimuli. It would therefore be necessary to compare whether participants responded differently to (positive - eco) vs. negative and at the same time whether positive - eco. This would directly answer the authors’ questions. That being said I am not sure that with an evaluative question this is interpretable as the original task. Also, the authors used reaction times. Did they use a transformation (e.g., Ratcliff, 1993)? Did they use cut-offs to exclude extreme data? They lack a lot of detail to be confident in the results presented here. In addition, one might question the role of the rating scale rather than a simple positive/negative categorization (e.g., Fazio et al., 1995). These previous remarks are also valid for the AMP analysis.
10 - On the other hand, concerning the tables, I do not think that saying that the hypothesis is accepted is the correct terminology. The authors seem to have performed inferential (not Bayesian) analyses, more precisely robust (as they refer to medians and not to the mean). In fact, it seems important to remember that we draw our conclusions on H0 and not on H1! Either H0 is rejected or H0 is not rejected (e.g., Cohen, 1988).
11 - In addition, the authors present two studies (although they say they only did one study in the abstract…). The authors present a “full sample” column in their results tables. As I understand it, the authors put the two data sets together to do the calculation (?). This is not something that can be done. Either the authors have one study, in which case I am not clear why they separated the two studies, or they have two studies, in which case, if they want to “put the two together” they need to perform a mini-meta-analysis (e.g., Goh et al., 2016).
12 - Regarding the analyses of the second hypothesis, Table 3 does not seem to provide useful information. Furthermore, I think it is important to carry out an AFE or CFA in order to verify the structure of the questionnaire in this study. It is possible that some items do not saturate correctly or that items saturate on several dimensions making them not readable to create an aggregate score. Also, I am not sure I understand the value of SEMs for testing correlations. Why did not the authors run multiple regressions with their three factors of interest? Also, why did the authors include sociodemographic variables? Does the literature mention any relationships between these variables and the measures? (This is not mentioned in this manuscript. Why these and not others? Also, what is the “positive emotions” measure? The authors do not specify how this score is obtained/measured. I am therefore not convinced that these analyses reflect the reality of the relationships in their studies.

Discussion
13 - Thus, for all of these remarks, I am not convinced in the state of the authors’ conclusions. The authors do not seem to have taken into account all the existing literature on the measures they have used and the data analyses do not seem to me to correspond to what they want to verify. Therefore, the conclusions drawn do not seem to me to be relevant here. For example, the authors (p. 15) say that their studies show the importance of looking for more determinants of pro-environmental behavior. I do not think that is what they did here. They showed that the attitude associated with pro-environmental behavior was positive. That does not make it a determinant, though. The authors would have to manipulate valence or specific emotions to see if behaviors are changed by it. At the end of my reading of this manuscript, I am not sure a) that the studies answered the authors’ question or b) that they used the right theoretical arguments to answer their question. In any case, the conclusions do not seem to me to follow from the studies that have been conducted.

# MINOR POINTS

* Figure 3 is for me incomprehensible without at least a detailed description. And if there is this description, I do not think it is useful.
* Figures 4 and 6 do not seem to me to be suitable. It would be better to present the separate columns in a bar plot for each condition. I think that the graphs are too complex for the presentation of averages and they are extremely complicated to read.
* Figures 5, 7 and 8 should simply be removed. They are not commented and do not provide any important information for understanding the results. Why did you do it by age? Does it really matter? Why not male/female? If there are differential assumptions related to age, they should be discussed.
* The tables should all be removed from the body of the text. They are not meant to replace textual explanations and in any case. They do not provide any information relevant to the studies.
* The APA standards should be adopted and the manuscript written, according to them. This will provide a guide for authors to make the paper more readable for readers.
* Please have the English language checked throughout the paper.

# Reviewer 2

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) | ✔ |  |  |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

This manuscript reports 2 studies using implicit measures to assess environmental attitudes in a Russian sample. To position myself, I accepted this review invitation because I am very familiar with the implicit social cognition literature, and am interested to see research in this area generalized to non-WEIRD contexts. The research appears to be well-conducted. The author reports the same pattern of results across two implicit measures (AMP, EPT), which increases my confidence in the reliability of the findings. The author also includes some manipulation checks to provide a baseline of evaluation, which adds to the construct validity of the work by demonstrating how participants evaluate stimuli that are clearly positive or negative. I also appreciate the openness with which the author shares materials. That said, I have some suggestions to improve the manuscript, which I discuss below in their order of importance.
My first, and biggest, concern is about the interpretation of the findings. In the AMP, participants viewed environment-related prime stimuli followed by random geometric shapes as target stimuli, then categorized the target stimuli as positive or negative. As the author notes, “The procedure is based on the premise that affective responses to the initial target object will be misattributed to, and thus will influence evaluations of the neutral stimulus, thereby providing insight into the evaluation of the target object.” Thus, the AMP in this case measures evaluations – that is, degree of positivity or negativity – associated with environmental stimuli. However, to my surprise, the author interprets AMP responses to reflect emotions. To be sure, emotions have valence (evaluation), but they’re different constructs. This point can easily be illustrated by the fact that different emotions can share the same valence. Anger, fear, and guilt are all negative emotions. Happiness, excitement, and contentment are all positive emotions. So the author didn’t measure emotion, but instead evaluation. The same criticism applies to the author’s interpretation of the EPT. In the EPT, participants first viewed a prime stimulus with known valence followed by a target environmental stimulus of unknown valence, then categorize the target stimulus as positive or negative. As the author notes, the EPT “identifies whether the evaluation of the first stimulus (the prime, that is to be ignored) affects the processing of subsequent stimuli.” Indeed, the author selected the prime stimuli based on pre-ratings of valence. Consequently, both the AMP and EPT measure evaluations of, rather than emotions towards, environmental stimuli, and the manuscript must be updated throughout to reflect the correct construct.
My second, and equally important concern is also about the interpretation of the findings. The manuscript is framed in terms of environmental “actions” and “behaviors”. Thus, I expected that the author had measured participants’ behaviors, maybe in the form of retrospective self-report or perhaps in the form of behavioral intention. Instead, I was surprised to see that the author’s actual operationalization of environmental actions/behaviors was the pictures they used as targets in the EPT and as primes in the AMP. I was even more surprised to see that many of the pictures do not depict behavior at all. For example, a picture of an apple with a BIO sticker on it doesn’t depict a behavior, nor does a picture of a bag that says “Go Green”. (On a related note, I’m not sure how a picture of plastic bottle caps arranged in the shape of a heart, or of a city inside of a light bulb, are related to the environment. But perhaps these pictures have more clear meaning in a Russian context?) In my opinion, these pictures largely share the common theme of “sustainability”, but I will defer to others who have more expertise in that domain if there is a better term to describe them. Nevertheless, framing the manuscript in terms of environmental “actions” and “behaviors” is inappropriate given the research design and stimuli.
My remaining concerns are relatively more minor then the first two. The author notes that the SEM is based on a median split. I am not an SEM expert, but the limitations of median splits are well known: they create artificial differences between things that are not different (e.g., 3 is equally close to 2 and 4, but median splitting at 3 forces 3 to be equal to 4 but different from 2). Unless median split is established as the appropriate way to prepare data in this context, I would prefer to see another analytic approach employed.
The manuscript would be improved by a thorough proofread, and also with more detail in the methods section. The writing did not impact my evaluation of the science, but many parts of the manuscript were difficult to understand the author’s intended meaning. The author also does not report important details about the EPT and AMP procedures. How many trials? How long were stimuli presented, and what was the latency between trials? To be honest, in my first reading of the manuscript I thought that the probe procedures were instead modified versions of the EPT and AMP. It took me a while to understand that these were separate from the implicit measures, and I’m still not confident that I understand exactly what either implicit measure consisted of.

**Author Response**
Dec 22, 2022

We wish to thank the reviewers for their insightful comments which we feel have substantially improved our manuscript. We believe we have addressed all of the major and minor comments that were raised by the reviewers and, in doing so, have crafted a paper that is clearer in presentation.

**to reviewer 1**

|  |  |
| --- | --- |
| “Emotions are a fundamental component of human psychological functioning, yet the study of their contribution in explaining a number of behavioral facets (pro-environmental behavior included) has been traditionally neglected (Tapia-Fonllem et al., 2013).” (p. 1). This statement is at best clumsy, at worst completely wrong!  | Accepted. Have been removed.  |
| My point is that the authors are interested here in the attitudes of individuals toward pro-environmental behaviors, not in the ability of emotions to produce pro-environmental behaviors. Why do the authors refer to work on the links between emotions and behaviors when they are studying attitudes? The authors need to clarify this and clarify their theoretical anchors. | Accepted. The narratives have been reframed and rewritten. Now the right constructs are used and the main hypotheses relate only to attitudes. |
| They want to know if the implicit attitude associated with pro-environmental behaviors corresponds to the explicit attitude (i.e., a positive attitude). Furthermore, I think it is essential that the authors refer to work on attitudes such as the APE model (Gawronski & Bodenhausen, 2011). This is much more central than the work presented on the influence of emotions on behavior. | Added. |
| The authors should emphasize the difference between explicit and implicit measures because that is the central point of the manuscript...Far more centrally, I think that work and models on implicit and explicit attitudes should be the focus of the entire introduction to this manuscript, because again, the authors have not studied the emotion-behavior link as the introduction might suggest. The authors should also incorporate a discussion of current debates around the explicit/implicit, direct/indirect distinction (e.g., Corneille & Hutter, 2020). | Accepted. This discussion was mentioned briefly. |
| Also, the authors say that images can decrease cross-cultural differences. I am not convinced of this (e.g., Phillips, 2019). This point should be further developed and argued if the authors find it central or simply not discussed as it is questionable.  | Accepted. This weak suggestion has been removed. |
| I’m not sure the authors can say that one component of attitude is behavior. It is rather the behavioral intention. The attitude-behavior link is not always highlighted (e.g., Haddock, 1993). Furthermore, I find that the table for presenting the work is not at all appropriate. It is hard to understand the purpose of this table. It would be better to describe the work and relate it to the attitudes. | Accepted. This weak suggestion and the table have been deleted. The narrative has been rewritten. |
| I also find that a central point (i.e., cultural difference) becomes completely buried. Finally, by trying to say that Russian culture has not been studied, the authors do not defend the idea that it is different from Western cultures. Do the authors mean that there is a common background to these attitudes that might not depend on culture or are the authors looking for the specificity of Russian culture? This second interpretation seems to me to be taken with caution. The authors did not make a comparison between cultures, so the conclusion remains on the Russian culture without making a comparison with other cultures.  | Accepted. We have added a brief description of Russia's environmental culture in Appendix F., as well as some other comments that relate to such differences. |
| Finally, is it important for the study of implicit attitudes that the authors conduct the study with a Russian population? Or, is it just important to adapt the tools to this population? This needs to be clarified to make the manuscript more readable. | Accepted. We have tried to clarify that research in the Russian context is important in itself, all the more so when using implicit methods. |
| the authors present hypotheses about specific behaviors. Why were they studying attitudes toward these particular behaviors? The authors make no mention of this before in the theory. It is difficult to get a sense of the value of studying these behaviors. The same is true for the second hypothesis. There is no mention of these effects or variables prior to the presentation of the assumptions. Overall, this is consistent with my previous comments. There is no explicit link between the theory mentioned and the hypotheses/studies. | Accepted. We have added this remark to the limitations section and eventually excluded (did not report) weak hypotheses about specific pro-environmental behaviors. The focus was on a common understanding of sustainability. |
| The authors start by mentioning two methods: AMP and AFP. First, it would be nice to reference the authors who developed these tasks… Also, I am not sure that the authors have performed both AMP and AFP tasks here. At least, they used extremely modified versions of them. For example, AMP (Payne et al., 2005) involves categorizing a neutral target after the presentation of a positive or negative stimulus.  | Accepted. We have provided a more detailed description of the methods, highlighted any differences from the original methods, and pointed out the authors. There are no critical modifications for AMP, and a Likert-type scale is also a common approach (Payne et al., 2014). We test for differences in reported valence between env.related prime+neutral target and positive prime+neutral target as well as between env.+neutral and negative+neutral.AFP was based on the general idea of affective priming and Stroop effect. We have put additional details. |
| Thus, given the assumptions, I infer that the authors will present positive, negative, and behavior-related stimuli whose associated attitudes they want to know. However, the authors have modified the measure to transform the categorization into a form of evaluation. This does not refer to the same types of tasks or processes. This argument is the same as for AFP where the DV is the reaction time for compatible and incompatible trials. This is even more problematic for this task. Indeed, I am not sure that they can be interpreted as the authors intend. | Accepted. We have added a separate paragraph (2.) describing the original methods and modern modifications. Also, categorisation is a very basic cognitive activity that is involved in the evaluation process (Harnad, 1987), and we believe that more than one button for opposing categories has little effect because we explicitly labeled the buttons as positive and negative and asked participants to respond as quickly as possible. The basic evaluative priming task allows for any rapid cognitive response: irrelevant naming to the task, unevaluated categorisation, classification. The priming effect has also been found with the rating task (e.g. Appel, Englich and Burghardt, 2021). We followed the logic as in (Damen et al., 2018). To avoid confusion with other procedures we have used the acronym AFP. |
| The DV should become the difference between the evaluation of the target stimuli after presentation of a neutral prime and the evaluation of the target stimuli after presentation of the other types of stimuli. My point is that DV should be a difference in evaluation and not an absolute evaluation. As it stands, I am not convinced that it is possible to make sense of this measure. Furthermore, these tasks are done with multiple trials. However, here, the authors have one and only one trial for each of the measures. So, the measure does not seem to me to be extremely reliable. I would be more convinced if the authors performed the tasks in a “classical” way and found results consistent with their hypotheses. | Partially accepted. In AFP we vary only the target item (pre-rated positive or negative) in order to assess "compatibility" with the prime item, which is env.-related and has an unknown valence. We check whether there is a significant difference in response time between env. + positive and env. + negative pairs (8 per subject in total). This is a common idea, consistent with (Fazio, Jackson, Dunton, & Williams, 1995). For example, we have shown that env.+positive pairs appear to be 'compatible' because they have lower response times than env.+negative pairs.Additional trials (8 per subject in total) of positive+positive, negative+negative (congruent/compatible) and positive+negative (incongruent) are given to test whether things work as predicted. In AFP we did not take into account the reported valence, but an additional and detailed comparison is given in Appendix E, an additional analysis was added to the results. We conducted not one but 4 trials for env.-related + positive and 4 trials for env. related + negative, 8 per subject in total (this is less than in a typical laboratory experiment, but enough to detect a difference due to sample size). |
| Regarding AFP, I do not think the analyses should be done on reaction times considering the way the task was modified as I mentioned before. Also, if I understand correctly, the authors compared RT(Eco-positive) vs. RT(Eco-negative). I do not think these answers their question. It seems to me that Helmert contrast analysis (e.g., Walsh, 1991) is more appropriate to the question. The authors postulate that behaviors will be associated with a positive attitude so they will only differ with negative stimuli. It would therefore be necessary to compare whether participants responded differently to (positive - eco) vs. negative and at the same time whether positive - eco. | Accepted. Valence was only considered in AMP, it was not considered in AFP. The AFP makes sense in terms of (in)compatible trials. There is no explainable relationship between the env.+pos./neg. and pos.+pos./neg.neg. trials, so we do not think that comparing RT between them makes sense.However, we added robust linear regression (RLM) and Helmert coding, to compare whether subjects responded differently (reported valence) to env.+pos. AFP trials compared to env.+neg., and at the same time, whether the actual response to env. stimuli was positive. Despite dealing with outliers such an analysis looks redundant with the idea of AFP, since the env.+positive pair was unambiguously assessed as positive and the env.+negative as negative:\*cumulative proportion |
| Also, the authors should go back to the power calculations. I do not know what effect size the authors used as the average size, but taking d = 0.36 (average effect size in psychology, Lovakov & Agadullina, 2021), with MorePower (Campbell & Thompson, 2012) with a 6-modality within-subject response design for a power of .90, I find a minimum sample size of 104 subjects, not 38! In addition, the authors do a second analysis with 15 predictors. What are these 15 predictors? This is not mentioned before! I don’t understand why this analysis is done and what it refers to. The analyses should be redone and more details should be given to understand how the calculation was made and with which indicators. | Accepted. |
| In addition, there is a great deal of detail missing from the procedure. Were the trials fixed or randomized? This should be included in the body of the text. When I read the procedure of a study, I should be able to reproduce the body of the study if I want to replicate observed results, for example. With the information in the manuscript, I am currently unable to do so. | Accepted. The methods section was rewritten. Trials and images were randomized but for a certain respondent trials of different categories were given in a fixed proportion.We hope that readers will be able to download the scripts from the repository and fully reproduce the procedure. |
| It would be better for the method and the results to separate the two studies. This would make the reading much easier and clearer. | Declined. Both studies had performed for a reason of repeatability and with the same algorithm. As there are no significant differences in the results or methods, we decided to put them together and not to duplicate the findings. |
| The authors have only reported the p-value. Besides the fact that this does not respect APA standards, this indicator is really criticized so presenting only this one give almost no information except the significance. What is the statistic? What are the CIs? What is the effect size? This makes it difficult to read the results and get an idea of the strength of the effects observed by the authors. In addition, the authors do not detail the analyses they performed. This more than laconic presentation of the results does not give us enough information to know if the statistics are really adapted and if they seem correct when read. This does not seem acceptable to me. So, I will critique the analyses and results with what I think was done and what I think would be more appropriate. | Accepted. Corrected with the APA standards. |
| Also, the authors used reaction times. Did they use a transformation (e.g., Ratcliff, 1993)? Did they use cut-offs to exclude extreme data? They lack a lot of detail to be confident in the results presented here. | Accepted. The methods section was rewritten. We did not use any transformations, the cut-off for responses longer than 4s was mentioned in the methods section. |
| In addition, one might question the role of the rating scale rather than a simple positive/negative categorization (e.g., Fazio et al., 1995). These previous remarks are also valid for the AMP analysis. | For AMP there are no critical modifications and Likert-type scale is also a conventional approach (see Payne et al., 2014).We have added a separate paragraph (2.) describing the original methods and modern modifications. Also, categorisation is a very basic cognitive activity that is involved in the evaluation process (Harnad, 1987), and we believe that more than one button for opposing categories has little effect because we explicitly labeled the buttons as positive and negative and asked participants to respond as quickly as possible. The basic evaluative priming task allows for any rapid cognitive response: irrelevant naming to the task, unevaluated categorisation, classification. The priming effect has also been found with the rating task (e.g. Appel, Englich and Burghardt, 2021). We followed the logic as in (Damen et al., 2018). To avoid confusion with other procedures we have used the acronym AFP. |
| On the other hand, concerning the tables, I do not think that saying that the hypothesis is accepted is the correct terminology. The authors seem to have performed inferential (not Bayesian) analyses, more precisely robust (as they refer to medians and not to the mean). In fact, it seems important to remember that we draw our conclusions on H0 and not on H1! Either H0 is rejected or H0 is not rejected (e.g., Cohen, 1988). | Accepted. It was removed. |
| In addition, the authors present two studies (although they say they only did one study in the abstract…). The authors present a “full sample” column in their results tables. As I understand it, the authors put the two data sets together to do the calculation (?). This is not something that can be done. Either the authors have one study, in which case I am not clear why they separated the two studies, or they have two studies, in which case, if they want to “put the two together” they need to perform a mini-meta-analysis (e.g., Goh et al., 2016). | Accepted. The calculations have been separated. Mixed data and analysis have been removed. Both studies have the same algorithm and are performed for repeatability purposes. |
| Regarding the analyses of the second hypothesis, Table 3 does not seem to provide useful information. Furthermore, I think it is important to carry out an AFE or CFA in order to verify the structure of the questionnaire in this study. It is possible that some items do not saturate correctly or that items saturate on several dimensions making them not readable to create an aggregate score. | Partially accepted. The table has been removed.The SEM analysis includes CFA and we decided to present all the statistics and factor loadings in an appendix. The Schultz’ questionnaire is widely used and has recently been tested in Russia (Sautkina et al., 2022), so explicit validation is out of the scope. We highlighted this point in limitations. |
| Also, I am not sure I understand the value of SEMs for testing correlations. Why did not the authors run multiple regressions with their three factors of interest?  | Declined. We were going to use the questionnaire responses as they are, without a separate CFA, and get bootstrapped estimations, thus SEM is more useful in this case. |
| Also, why did the authors include sociodemographic variables? Does the literature mention any relationships between these variables and the measures? (This is not mentioned in this manuscript. Why these and not others? Also, what is the “positive emotions” measure? The authors do not specify how this score is obtained/measured. I am therefore not convinced that these analyses reflect the reality of the relationships in their studies. | Partially accepted. We used socio-demographic variables as controls to draw more reliable conclusions, and did not test any hypotheses about them.The text relating to ‘emotions’ has been excluded. |
| Thus, for all of these remarks, I am not convinced in the state of the authors’ conclusions. The authors do not seem to have taken into account all the existing literature on the measures they have used and the data analyses do not seem to me to correspond to what they want to verify. Therefore, the conclusions drawn do not seem to me to be relevant here. For example, the authors (p. 15) say that their studies show the importance of looking for more determinants of pro-environmental behavior. I do not think that is what they did here. They showed that the attitude associated with pro-environmental behavior was positive. That does not make it a determinant, though. The authors would have to manipulate valence or specific emotions to see if behaviors are changed by it. At the end of my reading of this manuscript, I am not sure a) that the studies answered the authors’ question or b) that they used the right theoretical arguments to answer their question. In any case, the conclusions do not seem to me to follow from the studies that have been conducted. | Accepted. The narratives have been reframed and rewritten. Now the right constructs are used and the main hypotheses relate only to attitudes. |
| Figure 3 is for me incomprehensible without at least a detailed description. And if there is this description, I do not think it is useful.Figures 4 and 6 do not seem to me to be suitable. It would be better to present the separate columns in a bar plot for each condition. I think that the graphs are too complex for the presentation of averages and they are extremely complicated to read.Figures 5, 7 and 8 should simply be removed. They are not commented and do not provide any important information for understanding the results. Why did you do it by age? Does it really matter? Why not male/female? If there are differential assumptions related to age, they should be discussed.The tables should all be removed from the body of the text. They are not meant to replace textual explanations and in any case. They do not provide any information relevant to the studies. | Accepted. All these unrelated data have been deleted. |
| The APA standards should be adopted and the manuscript written, according to them. This will provide a guide for authors to make the paper more readable for readers.Please have the English language checked throughout the paper. | Accepted. |

**to reviewer 2**

|  |  |
| --- | --- |
| My first, and biggest, concern is about the interpretation of the findings. In the AMP, participants viewed environment-related prime stimuli followed by random geometric shapes as target stimuli, then categorized the target stimuli as positive or negative. As the author notes, “The procedure is based on the premise that affective responses to the initial target object will be misattributed to, and thus will influence evaluations of the neutral stimulus, thereby providing insight into the evaluation of the target object.” Thus, the AMP in this case measures evaluations – that is, degree of positivity or negativity – associated with environmental stimuli. However, to my surprise, the author interprets AMP responses to reflect emotions. To be sure, emotions have valence (evaluation), but they’re different constructs. This point can easily be illustrated by the fact that different emotions can share the same valence. Anger, fear, and guilt are all negative emotions. Happiness, excitement, and contentment are all positive emotions. So the author didn’t measure emotion, but instead evaluation. The same criticism applies to the author’s interpretation of the EPT. In the EPT, participants first viewed a prime stimulus with known valence followed by a target environmental stimulus of unknown valence, then categorize the target stimulus as positive or negative. As the author notes, the EPT “identifies whether the evaluation of the first stimulus (the prime, that is to be ignored) affects the processing of subsequent stimuli.” Indeed, the author selected the prime stimuli based on pre-ratings of valence. Consequently, both the AMP and EPT measure evaluations of, rather than emotions towards, environmental stimuli, and the manuscript must be updated throughout to reflect the correct construct. | Accepted. The narratives have been reframed and rewritten. Now the right constructs are used and the main hypotheses relate only to attitudes. |
| My second, and equally important concern is also about the interpretation of the findings. The manuscript is framed in terms of environmental “actions” and “behaviors”. Thus, I expected that the author had measured participants’ behaviors, maybe in the form of retrospective self-report or perhaps in the form of behavioral intention. Instead, I was surprised to see that the author’s actual operationalization of environmental actions/behaviors was the pictures they used as targets in the EPT and as primes in the AMP. I was even more surprised to see that many of the pictures do not depict behavior at all. For example, a picture of an apple with a BIO sticker on it doesn’t depict a behavior, nor does a picture of a bag that says “Go Green”. (On a related note, I’m not sure how a picture of plastic bottle caps arranged in the shape of a heart, or of a city inside of a light bulb, are related to the environment. But perhaps these pictures have more clear meaning in a Russian context?) In my opinion, these pictures largely share the common theme of “sustainability”, but I will defer to others who have more expertise in that domain if there is a better term to describe them. Nevertheless, framing the manuscript in terms of environmental “actions” and “behaviors” is inappropriate given the research design and stimuli. | Accepted. Indeed, we asked the experts to rate images that have a relatively strong association to the sustainability domain and look related to pro-environmental behavior. So, we have added this remark to the limitations section and eventually excluded (did not report) weak hypotheses about specific pro-environmental behaviors.The narratives have been reframed and rewritten.  |
| My remaining concerns are relatively more minor then the first two. The author notes that the SEM is based on a median split. I am not an SEM expert, but the limitations of median splits are well known: they create artificial differences between things that are not different (e.g., 3 is equally close to 2 and 4, but median splitting at 3 forces 3 to be equal to 4 but different from 2). Unless median split is established as the appropriate way to prepare data in this context, I would prefer to see another analytic approach employed. | Accepted. This part of the analysis has been replaced by the full SEM analysis without splitting or any other manipulation. As there is nothing of interest, the SEM diagrams and statistics have been placed in the appendix E.  |
| The manuscript would be improved by a thorough proofread, and also with more detail in the methods section. The writing did not impact my evaluation of the science, but many parts of the manuscript were difficult to understand the author’s intended meaning. The author also does not report important details about the EPT and AMP procedures. How many trials? How long were stimuli presented, and what was the latency between trials? To be honest, in my first reading of the manuscript I thought that the probe procedures were instead modified versions of the EPT and AMP. It took me a while to understand that these were separate from the implicit measures, and I’m still not confident that I understand exactly what either implicit measure consisted of. | Accepted. We have made a more detailed description of the methods. The number of trials, delays, etc. have been mentioned not only in the figures, but also added in the text. |

**Editor First Decision**: Revise & Resubmit

Jan 31, 2023

Dear Dr. Valko,

I have now received two reviews of your manuscript, “Does going green feel good in Russia: implicit measurements with visual stimuli”, from researchers with special expertise in XXX. I also independently read the manuscript before consulting these reviews. The reviewers had mostly positive reactions to your revised manuscript. I agree that your manuscript has improved in many regards, but some issues remain that need to be addressed. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

The reviewers did an outstanding job in their reviews. I will highlight issues I think are particularly salient here. In your resubmission, please include a document with a point-by-point response to both the points I list here and the reviewers’ comments, outlining each change made in your manuscript or providing a suitable rebuttal. In your resubmission, please ensure that your response letter is easy to follow and explains in detail which changes you made in response to which prompt.

Both reviewers highlight issues with the statistical analyses, and I concur. Since the primary criteria for evaluating manuscripts submitted to Collabra: Psychology are scientific, methodological, and ethical rigor, providing rigorous analyses is an important aspect to warrant publication. In your revision, please ensure that the analyses reported are complete (i.e., all relevant test statistics are easily accessible, justification for the chosen method is stated, result is interpreted clearly) and comprehendible. In particular, please ensure that the relation of each reported analysis to the preregistered analyses is clarified.

In summary, I continue to think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision. Please see the instructions below for submitting your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Rima-Maria Rahal

# Reviewer 1

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  | ✔ |  |  |

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

I was Reviewer 2 on the original submission of this manuscript. The authors adequately addressed the concerns I raised in my initial review regarding the attitude and behavior measures. Thus, I will limit my comments here to the newly-added parts of the manuscript.
In their literature review, the authors note that “Attitudes are understood as the result of separate mental systems, one conscious and the other unconscious, and capable of influencing their behaviour beyond awareness (Wilson & Smith, 2017).” It is certainly true that some researchers conceptualize implicit versus explicit attitudes to correspond to conscious versus unconscious mental systems. However, this perspective does not reflect any sort of consensus within the field. See the recent issue of Psychological Inquiry featuring the target article from Gawronski, Ledgerwood, and Eastwick for a variety of perspectives on this topic. The sentence I’ve highlighted here is, in my view, not necessary to support the main claims the author makes in this manuscript, so I suggest it be removed.
In the next section, I was surprised to see the author characterize the IAT as a priming task (it is not) and the AMP as not a priming task (it is). This section should be reworked. In fact, I don’t think its necessary for the author to mention the IAT at all, given that they do not rely on the IAT in the present research.
And finally, I find myself flummoxed by the material contained in the section titled “Additional analysis”. As written, there is not enough information for me to understand the analyses performed, or what the variables listed in either of the tables refer to. Given that the findings reported here do not seem to provide any additional information beyond what is reported in the previous section of analyses, I would suggest that this section be cut entirely. Otherwise, it will require considerable work to make it comprehendible.

# Reviewer 2

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) | ✔ |  |  |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  |  | ✔ |

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

# Reviews

## Does going green feel good in Russia: implicit measurements with visual stimuli

I thank the authors for their extensive revisions. In my opinion, the authors have decidedly improved the paper. A number of critical points have now been well addressed and I applaud the changes that were made. For instance, I think the new introduction about attitudes to pro-environmental behaviors is very informative. I also very much liked the added clarity throughout the manuscript. I think the inclusion of a section on the difference between explicit and implicit attitudes is particularly helpful in understanding what the authors are saying and the studies that have been done.

However, I think there are still some areas for improvement. In particular, I have several concerns about the statistical analyses that were chosen. The authors report using Wilcoxon tests. Why was this test chosen? This would need to be explained in the text. Moreover, from the results reported, the authors seem to have tested the effects in pairs (i.e., positive vs. negative; positive vs. environment; negative vs. environment). On the one hand, multiple two-to-two comparisons require the application of corrections, such as Bonferroni’s on the p-values. I don’t know if the authors have applied these corrections. I looked at the Jupyter Notebook (not using this data processing software, I am not sure) but I have the impression that the authors did not. On the other hand, I am not sure that these analyses best reflect the authors’ assumptions. The authors hypothesize that individuals will have a positive attitude toward pro-environmental behaviors. In other words, for the MPA, participants will have more positive evaluations of the target after exposure to positive or pro-environmental stimuli than after exposure to negative stimuli. This can be tested with contrast analyses (e.g., Judd, McClelland, & Ryan, 2017). Contrast analysis allows one to directly test a strong hypothesis by controlling for an alternative hypothesis (which here would be, do positive stimuli differ from environmental stimuli?). In concrete terms, this amounts to giving a weight to each condition, which here translates into the contrast codes (-1,-1,2) and (-1,1,0). The advantage here would be to present only one statistical test per study for each task, and thus gain statistical power. Moreover, for the affective priming task, the hypothesis is based on the interaction effect which is not directly tested. In fact, the authors only looked at simple effects here, whereas their hypothesis is about the interaction. Participants will be quicker to respond when the primes are positive or associated with the environment and the target is also positive rather than negative. This can also be reflected in contrast codes (-1,-1,2) X (-1, 1) vs. (-1,1,0) X (-1,1). Then, the authors could look at the simple effects by looking at the effect of the prime when the targets are positive (-1,-1, 2) X (1, 0), and when the targets are negative (-1, -1, 2) X (0, 1). This would allow us to be more precise and to answer more directly the hypotheses formulated by the authors.

In addition, in a more minor way, p. 9 “the combined sample size is N = 394,” the authors should change “the combined sample size” to “the total sample size of the two studies.” The term “combined” gives the impression that the authors have amalgamated the data from the two studies.

**Author Response**
Feb 6, 2023

|  |  |
| --- | --- |
| **Editor** |  |
| Please ensure that the analyses reported are complete (i.e., all relevant test statistics are easily accessible, justification for the chosen method is stated, result is interpreted clearly) and comprehendible. In particular, please ensure that the relation of each reported analysis to the preregistered analyses is clarified. | Checked. Made some corrections in section 5. |
| Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. | Checked. We have indicated the authors and sources. |
| **Reviewer 1 (previously 2)** |  |
| In their literature review, the authors note that “Attitudes are understood as the result of separate mental systems, one conscious and the other unconscious, and capable of influencing their behaviour beyond awareness (Wilson & Smith, 2017).” … The sentence I’ve highlighted here is, in my view, not necessary to support the main claims the author makes in this manuscript, so I suggest it be removed. | Corrected. |
| In the next section, I was surprised to see the author characterize the IAT as a priming task (it is not) and the AMP as not a priming task (it is). This section should be reworked. | Corrected. |
| In fact, I don’t think its necessary for the author to mention the IAT at all, given that they do not rely on the IAT in the present research. | Corrected. We have removed this text. |
| And finally, I find myself flummoxed by the material contained in the section titled “Additional analysis”. As written, there is not enough information for me to understand the analyses performed, or what the variables listed in either of the tables refer to. Given that the findings reported here do not seem to provide any additional information beyond what is reported in the previous section of analyses, I would suggest that this section be cut entirely. Otherwise, it will require considerable work to make it comprehendible. | Corrected. We have completely removed this part. |
| **Reviewer 2 (previously 1)** |  |
| The authors report using Wilcoxon tests. Why was this test chosen? This would need to be explained in the text. | Done. The data we collected are rank- or time-originated, which is likely determines that it is not normally distributed (we specially checked this on the AFP). For this reason, we choose the Wilcoxon test, which has sufficient power and robust enough. Also, initially we did not use any t- or F-based analysis and criteria. We have added this remark to the text (p. 5, footnote). |
| Moreover, from the results reported, the authors seem to have tested the effects in pairs (i.e., positive vs. negative; positive vs. environment; negative vs. environment). On the one hand, multiple two-to-two comparisons require the application of corrections, such as Bonferroni’s on the p-values. I don’t know if the authors have applied these corrections. | We have omitted redundant comparisons from the analysis. Technically it is enough to test one null hypothesis in each procedure with two opposite alternatives.Since p-values we obtained are mostly < .001, the Holm–Bonferroni correction (and other step-wise procedures) does not affect their significance. |
| On the other hand, I am not sure that these analyses best reflect the authors’ assumptions. The authors hypothesize that individuals will have a positive attitude toward pro-environmental behaviors. In other words, for the MPA, participants will have more positive evaluations of the target after exposure to positive or pro-environmental stimuli than after exposure to negative stimuli. This can be tested with contrast analyses (e.g., Judd, McClelland, & Ryan, 2017). Contrast analysis allows one to directly test a strong hypothesis by controlling for an alternative hypothesis (which here would be, do positive stimuli differ from environmental stimuli?). | Done. We decided to keep our analysis as it was preregistered, but we agree with these comments and have added the results of the contrast analysis in 4.1.Study 1Study 2Jupyter Notebook: <https://osf.io/download/63e011396946a005f07a4043>  |
| Moreover, for the affective priming task, the hypothesis is based on the interaction effect which is not directly tested. In fact, the authors only looked at simple effects here, whereas their hypothesis is about the interaction. Participants will be quicker to respond when the primes are positive or associated with the environment and the target is also positive rather than negative. | We have tried to answer a simple question which requires only a comparison between hypothetically compatible and incompatible pairs of prime and target (env+positive and env+negative). The comparison between explicitly compatible pairs (env+positive and env+negative) was introduced as a test of intervention validity, and we believe that it is not necessary to test interactions of any kind.Also, positive/negative contrast coding does not seem to be able to capture the idea of compatibility that participants also respond faster when primes and targets are negative. However, we very much appreciate the reviewer's help, so we have done this and included such an analysis in the text (4.1).Study 1Study 2Jupyter Notebook: <https://osf.io/download/63e011396946a005f07a4043> |
| In addition, in a more minor way, p. 9 “the combined sample size is N = 394,” the authors should change “the combined sample size” to “the total sample size of the two studies.” The term “combined” gives the impression that the authors have amalgamated the data from the two studies. | Corrected. |

**Editor Response: Revise & Resubmit**

Feb 17, 2023

Dear Dr. Valko,

I have received your revision of your manuscript, “Does going green feel good in Russia: implicit measurements with visual stimuli”.

However, in your response letter, you do not supply sufficient details on what was changed in response to the comments you received from the reviewers and from me.

I pointed the importance of the response letter out in my last decision on this manuscript already, but perhaps I should have been clearer about my expectations for your resubmission. It is often helpful for the review process if the response letter supplies text examples of what was added, removed or rephrased in response to reviewer prompts. Using tracked changes in the document is another way to facilitate assessment of the changes made. If you would like to receive more guidance on how to write a response letter that can facilitate the review process, perhaps you could refer to the APA guidelines (<https://apastyle.apa.org/style-grammar-guidelines/research-publication/response-reviewers>) where you will also find a sample response letter. Please resubmit a version of your materials that makes it possible to understand in detail how you responded to the prompts of the reviewers.

Further, I believe I pointed out to you the need to identify for each of the analyses you report in the manuscript whether you followed a preregistered analysis or report an exploratory, post hoc analysis. In your revision of the manuscript, I cannot see that this has been done. Given that rigor is the central facet on which the manuscript is being evaluated on, it is necessary that the nature of the analyses reported be clarified.

In summary, I continue to think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision. Please see the instructions below for submitting your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,
Rima-Maria Rahal

**Author Response**

Feb 6, 2023

|  |  |
| --- | --- |
| **Editor** |  |
| Please ensure that the analyses reported are complete (i.e., all relevant test statistics are easily accessible, justification for the chosen method is stated, result is interpreted clearly) and comprehendible. In particular, please ensure that the relation of each reported analysis to the preregistered analyses is clarified. | Checked. We have made some corrections in section 5. The sub-headings were changed to highlight types of the analyses. |
| Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. | Checked. We have indicated the authors and sources and put link to an online repository. |
| …Using tracked changes in the document is another way to facilitate assessment of the changes made. | Tracked changes were attached. |
| **Reviewer 1 (previously 2)** |  |
| In their literature review, the authors note that “Attitudes are understood as the result of separate mental systems, one conscious and the other unconscious, and capable of influencing their behaviour beyond awareness (Wilson & Smith, 2017).” … The sentence I’ve highlighted here is, in my view, not necessary to support the main claims the author makes in this manuscript, so I suggest it be removed. | Corrected. |
| In the next section, I was surprised to see the author characterize the IAT as a priming task (it is not) and the AMP as not a priming task (it is). This section should be reworked. | Corrected. |
| In fact, I don’t think its necessary for the author to mention the IAT at all, given that they do not rely on the IAT in the present research. | Corrected. We have removed this text. |
| And finally, I find myself flummoxed by the material contained in the section titled “Additional analysis”. As written, there is not enough information for me to understand the analyses performed, or what the variables listed in either of the tables refer to. Given that the findings reported here do not seem to provide any additional information beyond what is reported in the previous section of analyses, I would suggest that this section be cut entirely. Otherwise, it will require considerable work to make it comprehendible. | Corrected. We have completely removed this part. |
| **Reviewer 2 (previously 1)** |  |
| The authors report using Wilcoxon tests. Why was this test chosen? This would need to be explained in the text. | Done. The data we collected are rank- or time-originated, which is likely determines that it is not normally distributed (we specially checked this on the AFP). For this reason, we choose and preregistered the Wilcoxon test, which has sufficient power and robust enough. Also, initially we did not use any t- or F-based analysis and criteria. We have added this remark to the text (p. 5, footnote). |
| Moreover, from the results reported, the authors seem to have tested the effects in pairs (i.e., positive vs. negative; positive vs. environment; negative vs. environment). On the one hand, multiple two-to-two comparisons require the application of corrections, such as Bonferroni’s on the p-values. I don’t know if the authors have applied these corrections. | We have omitted redundant comparisons from the analysis. Technically it is enough to test one null hypothesis in each procedure with two opposite alternatives.Since p-values we obtained are mostly < .001, the Holm–Bonferroni correction (and other step-wise procedures) does not affect their significance. |
| On the other hand, I am not sure that these analyses best reflect the authors’ assumptions. The authors hypothesize that individuals will have a positive attitude toward pro-environmental behaviors. In other words, for the MPA, participants will have more positive evaluations of the target after exposure to positive or pro-environmental stimuli than after exposure to negative stimuli. This can be tested with contrast analyses (e.g., Judd, McClelland, & Ryan, 2017). Contrast analysis allows one to directly test a strong hypothesis by controlling for an alternative hypothesis (which here would be, do positive stimuli differ from environmental stimuli?). | Done. We decided to keep our analysis as it was preregistered, but we agree with these comments and have added the results of the post hoc contrast analysis in 4.1. Briefly: Study 1Study 2 |
| Moreover, for the affective priming task, the hypothesis is based on the interaction effect which is not directly tested. In fact, the authors only looked at simple effects here, whereas their hypothesis is about the interaction. Participants will be quicker to respond when the primes are positive or associated with the environment and the target is also positive rather than negative. | We have tried to answer a simple question which requires only a comparison between hypothetically compatible and incompatible pairs of prime and target (env+positive and env+negative). The comparison between explicitly compatible pairs (env+positive and env+negative) was introduced as a test of intervention validity, and we believe that it is not necessary to test interactions of any kind.However, we very much appreciate the reviewer's help, so we have done this and included such a post hoc analysis in the text (4.1). Briefly:Study 1Study 2 |
| In addition, in a more minor way, p. 9 “the combined sample size is N = 394,” the authors should change “the combined sample size” to “the total sample size of the two studies.” The term “combined” gives the impression that the authors have amalgamated the data from the two studies. | Corrected. |

**Editor Final Decision:** Accept

Feb 21, 2023

Dear Danila Valko,

I have now had a chance to read over your manuscript “Does going green feel good in Russia: implicit measurements with visual stimuli”, along with the letter describing the changes you made. Thank you for your responsiveness to the concerns that the reviewers and I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point.

You will be receiving separate correspondence regarding any production and technical comments, data deposits, as well as publication charges. We work with the Copyright Clearance Center to process any applicable APC charges. Please note that your APC transaction must be completed before your article gets published.

You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Rima-Maria Rahal