Reviews of original submission

Below you can find the reviews for the original submission

Editor

Dear Dr Tobias Dienlin,

Thank you for trusting your manuscript "Who Needs Privacy" for review by Collabra: Psychology. I sent the paper to two scholars who collectively have expertise concerning personality traits, cyber-psychology, and research methods. These individuals are extraordinarily well qualified to review this paper and I thank them for their service to this journal. I independently read the paper and then consulted the comments from the reviewers.

As you will read below, the reviewers thought the topic was interesting and worthy of study. They also noted that the paper is well written. However, the reviewers raised a number of significant questions and offered constructive suggestions for improvement, reconceptualization, and extension. The nature of the suggested improvements are substantial and seem to require additional samples. Indeed, I think a new and larger sample will be required to address the reviewer comments. Thus, I am sorry to say that my decision is to the submission for publication based on the peer review recommendations. I would, however, be open to a new manuscript that includes a second larger sample (at least) that addresses the concerns of the reviewers with respect to evidence of replication, consideration of alternative analytic frameworks, and enhanced attention to measurement. I also think a sample with a wider age range might prove to be a useful compliment to the sample in the sample. However, a large sample of college students might also work.

Regardless of this decision, I hope you find the reviews helpful as you decide on the next steps for this work. Below I outline some of the reasons that figured most prominently in my decision. I urge you to pay particular attention to reviewer feedback as I think their reactions are especially insightful.

1. I agree with concerns about re-labeling the NEO PI-R facets. I actually think it might prove useful to use a long facet-based measure and keep labels consistent with those commonly used in the literature. The machine learning approach suggested by Reviewer B is intriguing.

2. I agree with the concerns of the reviewers about the need for additional studies to make sure there is adequate power and precision as well as evidence that the measures have acceptable levels of validity for the research purposes. This is the major reason why I elected to reject the paper. I think the current evidence is too preliminary to meet the rigor bar set for work published in this outlet.

3. Reviewer A raises an interesting conundrum about the overlap between a need for privacy and sociability. This is something to consider when interpreting the results.

In the end, I am sorry to deliver unwelcome news. My decision boiled down to concerns about the strength of the current contribution in light of the expectations for rigor for papers at this outlet. I think the idea of identifying personality correlates for privacy concerns is interesting. The Reviewers and I simply concluded that additional study is needed to make a stronger case for the approach and ideas advanced in the paper. Regardless of this decision, I wish you well as you pursue these topics.

Sincerely, Brent Donnellan Senior Editor

Reviewer A:

General comments and summary of recommendation

Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:

- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?:

This manuscript reports a cross-sectional correlational survey study on the link between privacy needs and various personality traits. The topic is interesting, and the manuscript is written in an accessible, transparent way, for which the authors should be applauded. However, I also thought there were several weaknesses, both conceptually (which may be addressed) and with regards to the empirical work. To this end, it may be reasonable to conduct a follow-up study (but this is a decision that should be made by the editor handling this manuscript).

BACKGROUND

The introduction and theoretical background is overall solid. The authors acknowledge that that the available evidence in this area is a) limited and b) results are rather mixed. As such, they carefully develop hypotheses but make it quite clear that the work is generally quite exploratory. I had a few thoughts that the authors might find helpful should they be given the opportunity to revise their manuscript:

p. 3, l. 34: „We think that a better understanding of the relation between personality and privacy is crucial" Perhaps it is, but the authors do not really say why. Other than academic interest, why is a better understanding "crucial"? A paragraph laying out the relevance of research on the link between personality and privacy, either in the introduction or the discussion, would be helpful.

p. 8, l. 178: "Somewhat related, prior empirical research has shown that people who are more concerned about their privacy are also more likely to withdraw online, for example by deleting posts or untagging themselves from linked content (Dienlin & Metzger, 2016)." Maybe I'm just not understanding the connection, but maybe the authors could explain how this is even "somewhat" related to the level of general anxiety.

p. 8, l. 184 to p. 9, l. 194: I found this entire paragraph on the link between anxiety and consent to government surveillance somewhat unconvincing. It is plausible that anxiety correlates with being in favor of government surveillance; however, this does not necessarily extend to someone's own data. One can be in favor of government surveillance programmes and a high need for privacy. Maybe this paragraph just needs some slight editing to emphasize the point the authors are trying to make here (similar to p. 9, l. 207: one can desire privacy with regards to their own information or data and still be in favor of governmental surveillance measures to avert threats).

Maybe writing this will get me a ticket straight to personality psychology hell, but since the authors have already moved away from personality factors and instead consider facets (e.g. gregariousness instead of extraversion, p. 5), is it not conceivable that "need for privacy", too, is another of those facets? Or maybe why it's not! I'm not suggesting it actually is (I'm very much agnostic on that question) but maybe considering the idea would help the authors more clearly distinguish between the many constructs and also types of constructs they introduce and connect in their manuscript. In particular, I find it currently difficult to understand the meaningful difference between sociability ("whether people prefer to spend their time alone or with company") and privacy ("a voluntary withdrawal from society") as introduced in the manuscript.

DATA ANALYSIS

p. 11, l. 257: "Because all predictors were chosen on the basis of a-priori reasoning we did not control for alpha inflation." That is not a reason not to control for alpha inflation. If anything, in fact, one might even argue that it would be a good reason to do the opposite, i.e. apply some sort of inflation correction: Committing an alpha or Type I error means rejecting a true null hypothesis. When one is in "discovery mode", or largely exploring, there could be an argument to set alpha relatively high (.1 or maybe even .2) in order to balance it with Type II errors and not miss potentially important relationships (at the cost of a greater number of false positives). However, when one is relatively sure that observed variables are indeed related, there is no good reason to do so because p-values will tend to 0 and should not be "just below .05" anyway (I will return to this point later in my review). Here, it is more important that they survive "stricter" thresholds (as there should not be many false negatives). This is not to say I am asking the authors to definitely apply some correction method (although they reasonably might), but to reconsider the argument they present to have alpha fixed at .05.

MEASURES

Quite a few items from validated scales were dropped. I am not a big fan of this practice, and the authors note that it is certainly not ideal, but mention that it did not make a huge difference. Maybe they could expand a little what this means in practice (e.g. whether the bivariate correlations change meaningfully, etc).

I found it striking that the correlation between privacy need government and privacy need anonymity was > .8, which seems to suggest these two factors treated separately are massively overlapping. Maybe it would be good if the authors expanded a little on this.

RESULTS

Currently, I see the greatest issues with the results of the statistical analyses, or rather, their interpretation, in the context of the observed sample size and predefined SESOI. To me it seems that with the exception of the link between Sociability and Privacy need Interpersonal, all hypotheses should be rejected. This is greatly visualized in Figure 2, where a number of interesting observations can be made

1. all but one of the 95% CIs overlap with the equivalence region
2. all but the age CIs are actually wider than the equivalence region, which means that for most variables, the precision was too low to perform meaningful equivalence tests.

From what I understand, the authors did not actually conduct equivalence tests (otherwise they would have used 90% CIs instead of 95%), but use the thresholds descriptively (i.e. without testing "against" them). Still, taking into account the standard errors makes sense, but also implies that almost all of the coefficients are not larger than the authors' defined smallest effect size of interest. This is further corroborated by the observed p-values reported in Table 3 (in increasing order, < .05 only): < .001, .004, .008, .011, .025, .027, .031, .036, .037, .042. Under the assumption of true alternative hypotheses, this distribution of p-values should be quite rare, as they seem to tend away from rather than to 0. Even a very relaxed alpha inflation correction (e.g., fixed alpha/2, would have rendered 50% of these p-values nonsignificant). This, of course, would change the interpretation and discussion substantially, and I am happy to see counterarguments by the authors.

Notably, I am wholly unconvinced by the last paragraph in the results section (p. 19, l. 397 an onwards), purportedly presenting arguments why the results cannot be explained by chance alone. Although the authors seem to be very careful in their statistical inference (defining SESOIs, discussing power, etc), this paragraph seems to reintroduce several common misconceptions about p-values that they elsewhere carefully avoid, the most important being that P-values are conditional probabilities and as such cannot be used to infer that the results could not have been explained by chance alone (this is why we define a priori error rates). This entire paragraph is built on the comparison of the observations with a hypothetical "null hypotheses". However, as far as I can tell, the study tested a very large number of hypotheses (at least 24), and as such, speaking of one uniform null would be the most extreme case (24 out of 24 hypotheses are false). But just consider that, say 20 out or 24 tested hypotheses are actually false. Would observing 11 significant tests be so unusual? What about under 19 out of 24, or 17? These comparisons are meaningless as they do not tell us which of the 11 significant tests represent true hypotheses, and which are false positives.

DISCUSSION

I will limit my comments on the discussion for now as, depending on the revision of the manuscript, it might look very different from what is currently reported (see above).

CONCLUSION

Overall, the research topic is interesting, and the authors wrote up a transparent, accessible report. However, I wonder, given the concerns identified, the potentially insufficient precision, and the fact that the authors failed to meet their own sample target (300), whether it might not be the best route to conduct a follow-up study to substantiate many of these initial findings. This would also put a suggestion to the test the authors make on p. 19, l. 395: that a larger study, which yields smaller standard errors, would find additional effects. This is certainly possible, but it is also entirely conceivable that some of the initially observed effects would be smaller (regression to the mean).

Figures/tables/data availability:

Please comment on the author's use of tables, charts, figures, ifrelevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here).:

Figures and tables are used appropriately, and data are available.

Ethical approval:

If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared. If not, please detail where you think a further ethics approval/statement/follow-up is required.:

It is not reported whether consent was obtained (the study does have IRB approval though)

Language

Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:

The quality of English is fine; some slight editing might be good, but not necessary

Reviewer B:

1) General comments and summary of recommendation

Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:

- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?:

I am glad to have the opportunity to review this manuscript. By evaluating the individual characteristics that relate to the perceived need for privacy, it addresses a topic of increasing social relevance and I think there is a need for more research along these lines. I also applaud the authors' use of advanced analytic methods and their efforts to address issues important to open science.

That said, I feel the impact of this work is limited by their methodological approach. Specifically, I find myself unable to endorse their approach to measurement due to two concerns.

The first and more straight-forward concern relates to their measures of privacy. The standards for new measures that are derived through rational methods (the approach taken here) increasingly require more extensive procedures than those that are described in this manuscript. For example, it is becoming standard practice to attain expert input via qualitative data collection for concept elicitation (not mentioned here) in addition to an exhaustive literature review (conducted, at least in part, for this manuscript though it would be helpful to more directly explain how the prior research described in the introduction ties to procedures used to develop this measure of privacy). Ideally, when budgets allow, one would also co-administer extant measures of the construct before choosing content for preliminary/potential inclusion in the new measure, as this allows for more informed understanding of the overlapping and unique features of prior operationalizations. It is also important that the preliminary content is evaluated across a range of qualitative attributes (e.g., literacy levels, readability, avoidance of colloquialisms, etc.) prior to large scale data collection. I should note that the final items (in Table 2) look pretty good with respect to these qualitative attributes but little detail is given about the procedures used to write the new content. Then, the overly-inclusive pool of preliminary items would be administered to one or more large and representative sample(s) – EFAs and CFAs would not be used on the same data. Optimally, EFA results would be replicated before moving to confirmatory evaluations of fit. Finally, once a final set of items has been identified on the basis of these structural analyses, reports on their psychometric properties should be made. The authors did a nice job of addressing questions about dimensionality and reliability but validation evidence was lacking. It is most common for these procedures to be detailed in a stand-alone project/publication, and sometimes more than one. Of course, validation is an on-going endeavor and I think this paper would constitute a nice example of a validation project if the measure had previously been introduced in prior work (using something like the steps described above).

My second concern is more subtle than the first, but not less important. It relates to the authors' approach to measuring the other facets of personality. After reading the abstract, I was immediately curious to know which measurement framework the authors had used for personality assessment, as it was clear that they had deviated from using one of the mainstream assessment models – not necessarily a problem on its own merits, but this does introduce some challenges in terms of situating the findings relative to other personality research. As I proceeded through the introduction, I began to understand that they endorsed a Big Five approach but were measuring unfamiliar facet-level constructs. Only when reviewing the Measures section of the Methods did I understand that they were using facets of the NEO-PI-R with different names than the constructs being measured. While I am sympathetic to the notion that there are many plausible features of personality that *may* be covered by unnamed nooks and crannies of the NEO and various other measurement frameworks (more technically, "homogenous item clusters"), it is not valid to use – for example – the reversed scores of the Actions facet of Openness to Experiences as a measure of Traditionalism in the absence of empirical evidence confirming that this measures what the authors claim it measures. Though cold comfort for the authors, I understand the temptation to do this and I think stems from a real weakness for our field – the fact that we do not have a strong working knowledge of the nomological relations among empirically-identified personality traits that are less abstract than the Big Five.

I have a possible suggestion but it would require a substantial overhaul to the text. If the authors happened to collect data on all 30 facets of the NEO (not clear) in addition to the privacy items, an informative analysis would be to use machine learning methods capable of showing the NEO items that are most predictive of the privacy variables (whether scales or items). Some would argue that this constitutes baseless data mining. To counter that argument (a little), a nice addition might be pre-registration and analysis by someone who is not familiar with the data/findings and/or the use of a training and test sub-samples (given the size of this sample, I would recommend using it as the training sample and then collecting a new sample as the test set).

In addition to the comments provided here, I have attached a marked-up copy of the manuscript with several more minor comments and questions.

Figures/tables/data availability

Please comment on the author's use of tables, charts, figures, ifrelevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here).:

I have not attempted to reproduce these findings nor is it clear to me whether the data and code needed to do so are available. If so, I must have missed it. Given the comments provided in my review, reproduction of the results did not seem like an immediate concern.

Ethical approval:

If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared. If not, please detail where you think a further ethics approval/statement/follow-up is required.:

The study was subject to IRB approval.

Language:

Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:

Yes, I think it is well-written