**Peer Review and Communication History**

**MS Title**: “They Are Such an Asshole”: Describing the Targets of a Common Insult among English-Speakers in the United States

**Author Names**: Brinkley M. Sharpe, Courtland S. Hyatt, Donald R. Lynam, and Joshua D. Miller

**Submitted:** Nov 7, 2021

**Editor First Decision**: Revise & Resubmit

Dec 20, 2021

Dear Ms. Sharpe,

Thank you for submitting your manuscript, ““They Are Such an Asshole”: Describing the Targets of a Common Insult among English-Speakers in the United States” for streamlined review at Collabra: Psychology. I want to begin by reiterating my sincere apology for the way that your manuscript was handled and the message you received from one of our editors during the initial screening phase. I am very sorry this happened and deeply regret it. As per our prior communication, I have handled your manuscript myself. I appreciate the changes you made in response to the reviews and editorial decision you received at a previous journal. I especially appreciate the change in title, and I think the added exploratory analyses examining correlates of perceived knowledge, caring, and ability to change are interesting. I did not feel the need to send it out for further review, but I did give it a very careful read (I apologize for the delay in getting this decision to you, but I wanted to give your manuscript the full attention it deserves). As you’ll see below, my overall assessment is that your manuscript has many strengths and is close to publishable in Collabra, but I would like to see one quite major change, and several other small-to-medium-sized changes. I think the major change will present an important decision point for you and your co-authors about whether you agree with my requested reframing. I see a lot of value in your manuscript and do not think the reframing would detract from its value, so I hope you will consider this request to be reasonable and resubmit your revised manuscript, but of course I understand that you may choose not to do so. If you do make the framing change I requested and address the rest of the points, I will be able to make a decision on your revised manuscript quickly (though I will be away from work during the period from Dec 30 to Jan 12, so there may be a delay if you resubmit during that window). I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

Major reframing request:

Throughout the manuscript, you present the study and its results as reflecting the personality of the nominated assholes.  Although I am generally optimistic about the accuracy of informant reports, I think this claim needs to be heavily qualified in this case. While I am sure the participants’ informant ratings of targets have a substantial amount of valid variance, because the participants are also the ones who nominated this target as the “biggest asshole” in their lives, there is more than a little bit of circularity in this design.  The fact that it’s the same informant making the nomination and providing the rating surely contributes to the magnitude of the effects observed here (i.e., a substantial proportion of the variance is surely unique to this informant/perceiver, in addition to the valid/shared variance with a hypothetical independent informant).  In other words, whereas you present your study as examining the actual personality of nominated assholes, I think a more cautious framing and interpretation is necessary: you are studying how people perceive the biggest asshole in their lives, not what that person is actually like.  I recognize that this is a major reframing (or it would have to be, to my mind, to meet the threshold for publication in Collabra).  That is, this cannot be addressed by, for example, expanding the discussion of this limitation of your study.  For publication in Collabra, I would want the new framing to drive how the research is presented and interpreted throughout the manuscript.  You would be welcome to include a paragraph or section in the discussion about the degree to which these perceptions likely reflect accurate estimates of targets’ personalities, but the overall framing and interpretation would not assume that they are largely accurate, or equivalent to an independent informant’s rating.  If you are willing to make this substantial change in framing, I would be happy to publish your manuscript, contingent on you also addressing the points below (which range from minor to moderate, but none are as big as this one, so I suspect this will be the major factor that determines whether you choose to submit a revision to Collabra).

Here are some examples of phrases you use that I think are too strong.  This is not an exhaustive list, and it would not be enough to just change these phrases, but I am including these to give you some concrete examples of claims that I don’t think your design can support, because they make a strong assumption that the participants’ perceptions of assholes capture almost exclusively the assholes’ actual personalities:“Results showed that the use of the insult “asshole” is associated with interpersonally relevant traits…” (abstract)“The “biggest assholes” in participants’ lives […] were chiefly characterized by a combination of antagonism…” (p. 18)“popular insults like “asshole” convey important information about an individual’s interpersonal functioning” (p. 19)
“we established that the term “asshole” conveys information regarding the target’s personality” (p. 20)

In other places in the manuscript, the framing is similar to what I am imagining.  For example, the heading on page 6 “What information do people communicate regarding a target’s personality when employing insults” is much more likely to make readers think that you’re studying what the person who uses the descriptor “asshole” is trying to communicate, rather than that you’re studying what the “asshole” is necessarily actually like (and I think this is a good thing). The same goes for the heading “What behaviors do people associate with “Assholes”?”.  As these examples hopefully illustrate, I think the question your design *can* address, namely what are the perceptions people have of the people they consider assholes, is a very interesting one, and changing this framing would not make your paper any less worthwhile or important, but I do think the framing needs to change so that it is better matched to the design of the study.

Other points:

A separate but related point is that, when comparing the personality ratings of participants vs. nominated assholes, there is a significant confound.  Putting aside the fact that the ratings came from the same rater, creating shared method variance and potentially contrast effects, there is also the problem that one set of personality ratings were obtained via self-reports (the participants’) and another via informant reports (the assholes’).  This would be a confound even if the informants who rated the assholes were not the same people as the participants who provided self-ratings and nominated assholes.  Given that we know there are differences in the typical levels, structure, etc. of self- vs. informant ratings, these need to be taken into account when interpreting differences between the participants’ and assholes’ personality ratings.  To what extent do you think these differences are influenced by well-established differences in the psychometric properties of self- vs. informant reports?

Another substantial issue I would like to see incorporated into a revision is a more nuanced discussion of your research hypotheses and what would count as support for those hypotheses.  In this context, I don’t think the null hypothesis is a useful alternative – surely you would not claim that your hypotheses were supported if the associations/differences you observed were small (even if they were statistically significant), right?  Especially given the difficulty of ruling out very plausible artifacts such as halo effects, contrast effects, etc., it seems important to benchmark your findings against a more realistic alternative.  Even if your hypotheses were false, we would not expect most of these associations/differences to be zero, so a discussion of the magnitude of effects that would support vs. go against your predictions is important.

A smaller point is that, in some places, it is not clear or not emphasized enough that the part of the study/results about “behavior” was based on reports of behavior, not actual behavioral observation.  For the most part, the discussion of this aspect of the study is clear and well-calibrated, but there were a few places where I think a reader might think you measured actual behavior, and it would be good to be clearer.  I would suggest qualifying “behavior” wherever it appears, e.g., by referring to them as “nominated behaviors”. (Please also update the running head to better match the title/framing, and be less about behavior.  I believe it is a remnant from the previous title.  To be honest I have no idea if the running head matters or ends up anywhere, but just in case!)

I am probably missing something, but it seems to me that your prediction of a mean difference on agreeableness between participants’ self-reports and their perceptions of their asshole is redundant with the prediction that the FFM profiles of the asshole perceptions would be characterized by low Agreeableness (whereas profiles of participants’ self-reports would not).

The label “meta-perceptions” is used in a way that is not consistent with my understanding of its meaning.  I can see how ratings of knowledge are meta-perceptions, but I don’t think ratings of caring are, and I’m pretty sure ratings of ability to change are not.  I think it would be less confusing to use a different label for these three items.

Please add more explanation, and a more conceptual explanation, of the double-entry method?  What is the purpose, conceptually, of entering both rows for each pair of values?  What does it mean that they capture “absolute rather than relative similarity”? (I realize I should know this, but even if I did, most Collabra readers likely will not.). More importantly, please provide more description of what these profile similarity scores capture, conceptually.  What does a high score mean?  Could a pair of profiles have a higher score because they are both similar to the typical profile?  What are the various processes that might drive up a score? (I am used to the terms normative and distinctive accuracy, so explaining this method in those terms would help readers like me, but unfortunately that won’t be the case for most Collabra readers so again, you will need to find a way to explain what these scores mean that is accessible to people with no background in personality profile scores.)

Please provide more specific values for your power/precision analysis (e.g., can you change “over 90%” to a more specific value?).  In addition, where additional assumptions had to be made for the power analyses (e.g., the correlation between the paired values in the matched pairs t-test), please describe the values you used.

Throughout the results, many of the interpretations depend on the response scale used (e.g., interpreting the means).  Please remind readers of the response scale more every now and then, and include this information in all relevant table notes.

Please change the titles of tables and figures to make it clear who provided the ratings.  For example, Table 2 and Figure 1 could easily be misinterpreted by a reader as presenting ratings of informants and targets that were provided by different raters.  In general, I like to assume that some readers will skip the text and go directly to tables and figures, and we should make it as easy as possible for readers to interpret these on their own.  I realize it’s impossible to provide all the nuance/context, but the important features of the measures/design should be easy to grasp from the titles, headings, notes, etc.

I think the label “-ist” for one of the groups of nominated behaviors is much less clear than the others.  I would recommend using a label that is more directly related to the substance of the category (maybe “prejudiced” or “bigoted”?).

Please clearly flag any deviations from your pre-registration.  I did not notice any deviations from my quick comparison, but if there are any, please make sure they are very clear to readers.  In addition, please clearly identify any decisions about data collection, cleaning, or analysis that were not specified in the pre-registration (I assume there are some, as it is extremely rare to anticipate every decision that will have to be made).

Small typos/errors:

p. 8 “informant consent” should be “informed consent”
p. 12 “and broadly disagreeable (M = 1.95…” specify that the mean is on agreeableness, not disagreeableness, to avoid confusion
p. 13 “between informant’s self-reports and” should be the apostrophe should be after the s in “informants”
p. 17 you write that concordance was lowest for high Agreeableness, but then do not report any behaviors that were categorized as high Agreeableness (if I understood correctly).
p. 20 “Although participants provided…” “Although” seems like the wrong word in this sentence.

In summary, I think this is a strong manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving 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 is 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,

Simine Vazire
Editor in Chief
Collabra: Psychology

**Author Response**
Jan 5, 2022

Editor (Dr. Vazire):

1. *Throughout the manuscript, you present the study and its results as reflecting the personality of the nominated assholes.  Although I am generally optimistic about the accuracy of informant reports, I think this claim needs to be heavily qualified in this case. While I am sure the participants' informant ratings of targets have a substantial amount of valid variance, because the participants are also the ones who nominated this target as the "biggest asshole" in their lives, there is more than a little bit of circularity in this design.  The fact that it's the same informant making the nomination and providing the rating surely contributes to the magnitude of the effects observed here (i.e., a substantial proportion of the variance is surely unique to this informant/perceiver, in addition to the valid/shared variance with a hypothetical independent informant).  In other words, whereas you present your study as examining the actual personality of nominated assholes, I think a more cautious framing and interpretation is necessary: you are studying how people perceive the biggest asshole in their lives, not what that person is actually like.  I recognize that this is a major reframing (or it would have to be, to my mind, to meet the threshold for publication in Collabra).  That is, this cannot be addressed by, for example, expanding the discussion of this limitation of your study.  For publication in Collabra, I would want the new framing to drive how the research is presented and interpreted throughout the manuscript.  You would be welcome to include a paragraph or section in the discussion about the degree to which these perceptions likely reflect accurate estimates of targets' personalities, but the overall framing and interpretation would not assume that they are largely accurate, or equivalent to an independent informant's rating.*

In response to Dr. Vazire’s request, we have made numerous edits throughout the manuscript to clarify the nature of the obtained rating and to avoid implying that these perceptions of personality are “accurate.” We believe that our results speak to individuals’ use of the insult “asshole” to communicate information about their *perception* of a target’s personality.

1. *A separate but related point is that, when comparing the personality ratings of participants vs. nominated assholes, there is a significant confound.  Putting aside the fact that the ratings came from the same rater, creating shared method variance and potentially contrast effects, there is also the problem that one set of personality ratings were obtained via self-reports (the participants') and another via informant reports (the assholes').  This would be a confound even if the informants who rated the assholes were not the same people as the participants who provided self-ratings and nominated assholes.  Given that we know there are differences in the typical levels, structure, etc. of self- vs. informant ratings, these need to be taken into account when interpreting differences between the participants' and assholes' personality ratings.  To what extent do you think these differences are influenced by well-established differences in the psychometric properties of self- vs. informant reports?*

Dr. Vazire raises excellent points regarding the comparison between participant and target personality ratings. We do not believe that sufficient justification exists for inclusion of these analyses in the main text. We have chosen to move them to Supplemental Material (Appendix A) to balance these concerns with our values of transparency (and adherence to preregistration) in reporting.

1. *Another substantial issue I would like to see incorporated into a revision is a more nuanced discussion of your research hypotheses and what would count as support for those hypotheses.  In this context, I don't think the null hypothesis is a useful alternative — surely you would not claim that your hypotheses were supported if the associations/differences you observed were small (even if they were statistically significant), right?  Especially given the difficulty of ruling out very plausible artifacts such as halo effects, contrast effects, etc., it seems important to benchmark your findings against a more realistic alternative.  Even if your hypotheses were false, we would not expect most of these associations/differences to be zero, so a discussion of the magnitude of effects that would support vs. go against your predictions is important.*

This criticism is well warranted. While it is reasonable to infer from our hypotheses that a negative *rICC* value (i.e., dissimilarity) would not be interpreted as support and from our power analyses that effect sizes smaller than |.20| would be treated as unreliable, Dr. Vazire is correct to point out that we provide no explicit criteria for these or any other interpretations of our results. When formulating the present study, we should have described planned procedures and determined ahead of time what evidence would be required to support our hypotheses. However, our pre-registered hypotheses (<https://osf.io/9pafc>) are lacking in this level of detail. For example, when hypothesizing “similarity” between FFM profiles, we provided no minimum effect size of interest. To mitigate this oversight, we have revised our presentation of results to explicitly draw upon existing generalized recommendations for interpretation of effect sizes.

1. *A smaller point is that, in some places, it is not clear or not emphasized enough that the part of the study/results about "behavior" was based on reports of behavior, not actual behavioral observation.  For the most part, the discussion of this aspect of the study is clear and well-calibrated, but there were a few places where I think a reader might think you measured actual behavior, and it would be good to be clearer.  I would suggest qualifying "behavior" wherever it appears, e.g., by referring to them as "nominated behaviors". (Please also update the running head to better match the title/framing and be less about behavior.  I believe it is a remnant from the previous title.  To be honest I have no idea if the running head matters or ends up anywhere, but just in case!)*

We have made edits throughout the manuscript to clarify that “behaviors” were those the participants reported the target engaging in. We have also updated the running head to read **“PERCEIVED PERSONALITY OF INSULT TARGETS.”**

1. *I am probably missing something, but it seems to me that your prediction of a mean difference on agreeableness between participants' self-reports and their perceptions of their asshole is redundant with the prediction that the FFM profiles of the asshole perceptions would be characterized by low Agreeableness (whereas profiles of participants' self-reports would not).*

We hypothesized that the FFM profile of the prototypical asshole would be characterized by low Agreeableness in an absolute sense (i.e., below the midpoint of the Likert scale) and make no predictions regarding the absolute traits of our raters. Based on observed descriptive statistics for the IPIP-NEO-120 in other samples (e.g., *M* = 3.71, *SD* = 0.49, *n* = 1,194; Crowe et al., 2017), we believe it would have been extremely unlikely for our raters themselves to report low mean levels of Agreeableness. However, if we had inadvertently recruited a disagreeable sample, we would still have expected to see a significant difference between self-ratings and informant-ratings. As mentioned in Response #2, the comparison of participant and target FFM ratings has been moved to Supplemental Material (Appendix A).

Crowe, M. L., Lynam, D. R., & Miller, J. D. (2017). Uncovering the structure of agreeableness from self-report measures. *Journal of Personality, 86*, 771-787.

1. *The label "meta-perceptions" is used in a way that is not consistent with my understanding of its meaning.  I can see how ratings of knowledge are meta-perceptions, but I don't think ratings of caring are, and I'm pretty sure ratings of ability to change are not.  I think it would be less confusing to use a different label for these three items.*

We agree that this label is inaccurate and misleading and have replaced its use with direct descriptions of item content (e.g., “**For each target, scores were derived for perceptions of the target’s knowledge, care, and ability to change by taking the mean of the corresponding items across all three nominated “asshole-related” behaviors.”** p. 18).

1. *Please add more explanation, and a more conceptual explanation, of the double-entry method?  What is the purpose, conceptually, of entering both rows for each pair of values?  What does it mean that they capture "absolute rather than relative similarity"? (I realize I should know this, but even if I did, most Collabra readers likely will not.). More importantly, please provide more description of what these profile similarity scores capture, conceptually.  What does a high score mean?  Could a pair of profiles have a higher score because they are both similar to the typical profile?  What are the various processes that might drive up a score? (I am used to the terms normative and distinctive accuracy, so explaining this method in those terms would help readers like me, but unfortunately that won't be the case for most Collabra readers so again, you will need to find a way to explain what these scores mean that is accessible to people with no background in personality profile scores.)*

An explanation of absolute profile similarity as calculated using double-entry correlations was added to the main text (pp. 10-11): “**To measure absolute profile similarity, we employed the double-entry intraclass correlation (*rICC*) which is sensitive to both differences in score elevation (i.e., score means) and profile shape (i.e., pattern of scores; McCrae, 2008). A zero-order correlation (*r*) captures only similarity in profile shape (i.e., pattern of scores). In contrast, *rICC* captures similarity in both shape and elevation. Double-entry correlations are calculated by entering two rows of data for each pair of values (e.g., FFM ratings from “asshole” profile and expert-rated prototype) such that for a pair of values (X, Y), one row includes X in the first column and Y in the second, while the order is reversed in the second row (i.e., Y in the first column and X in the second). As an example, if column A had values of 1, 2, 3, 4, 5, 6, 7 and column B had values of 2, 4, 6, 8, 10, 12, and 14 – these two columns would have a perfect degree of relative similarity (*r* = 1.00) but only modest to moderate absolute similarity (*rICC* = .29) as the size of the elevations are reasonably different from one another.”**

We do not believe that similarity to a common normative (i.e., “typical’) profile is a relevant concern as we are comparing aggregate profiles (including the “normative” profile from participant self-reports) rather than those for individuals.

1. *Please provide more specific values for your power/precision analysis (e.g., can you change "over 90%" to a more specific value?).  In addition, where additional assumptions had to be made for the power analyses (e.g., the correlation between the paired values in the matched pairs t-test), please describe the values you used.*

Exact values are now provided in the manuscript (p. 12): **“Our effective sample size of 397 exceeded the minimum recommendations for stable estimates of correlations (*n* = 250; Schönbrodt & Perugini, 2013) and, according to G\*Power (Faul et al., 2007), provided 91.93% power to detect correlations as small as .20 at α = .01.”**

As mentioned in Response #2, the comparison of participant and target FFM ratings has been moved to Supplemental Material (Appendix A). The power analysis section in this document reads: **“For our comparisons between self and target ratings on the IPIP-NEO-120, we conducted a power analysis for our matched samples t-test using an alpha of .01 (two-tailed) and an effect size of .207. The effect size was arrived at using the average of the SDs across facets as the group SDs (i.e., .90), a small correlation between groups (i.e., .01), and a difference of .25 on the five-point IPIP scale (i.e., 2.75 vs 3.0). Under these parameters, with 397 participants, we have 93.7% to detect an effect size of 2.07 or larger.**”

1. *Throughout the results, many of the interpretations depend on the response scale used (e.g., interpreting the means).  Please remind readers of the response scale more every now and then, and include this information in all relevant table notes.*

Both the text of the results as well as table notes now orient readers to the relevant response scales (e.g., “**On the IPIP-NEO-120 (Table 2), participants described targets as angry (facet *M =* 4.34, *SD* = 0.83; 1 - *Disagree Strongly* to 5 - *Agree Strongly*) and broadly disagreeable (i.e., low Agreeableness; domain *M* = 1.95, *SD* = 0.55).”**, p. 13).

1. *Please change the titles of tables and figures to make it clear who provided the ratings.  For example, Table 2 and Figure 1 could easily be misinterpreted by a reader as presenting ratings of informants and targets that were provided by different raters.  In general, I like to assume that some readers will skip the text and go directly to tables and figures, and we should make it as easy as possible for readers to interpret these on their own.  I realize it's impossible to provide all the nuance/context, but the important features of the measures/design should be easy to grasp from the titles, headings, notes, etc.*

Table headings and notes have been edited to clarify rater source.

1. *I think the label "-ist" for one of the groups of nominated behaivors is much less clear than the others.  I would recommend using a label that is more directly related to the substance of the category (maybe "prejudiced" or "bigoted"?).*

We now refer to this theme as “Bigotry.”

1. *Please clearly flag any deviations from your pre-registration.  I did not notice any deviations from my quick comparison, but if there are any, please make sure they are very clear to readers.  In addition, please clearly identify any decisions about data collection, cleaning, or analysis that were not specified in the pre-registration (I assume there are some, as it is extremely rare to anticipate every decision that will have to be made).*

We have added footnotes and clarifications throughout our methods section to clearly identify deviations from pre-registration.

1. *p. 17 you write that concordance was lowest for high Agreeableness, but then do not report any behaviors that were categorized as high Agreeableness (if I understood correctly).*

For comprehensiveness, raters were provided the option of using high Agreeableness to categorize the reported behaviors but given the low concordance in its use it never appeared as the modal rating for a behavior. This is sensible, given that we asked participants to report behaviors that the current data suggest are typical of the opposite pole of this trait domain. We now highlight this absence (p. 18): **“High Agreeableness did not appear as the modal rating for any behavior, further supporting the primary role of disagreeableness.”**

**Editor Final Decision: Accept**

Jan 15, 2022

Dear Brinkley ,

Thank you for submitting your revised manuscript ““They Are Such an Asshole”: Describing the Targets of a Common Insult among English-Speakers in the United States”, along with the letter describing the changes you made. I appreciate your responsiveness to the issues I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this 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 (after our major mistake upon your original submission) 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. Our managing editor will contact you in case there are any pre-prodution file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Simine Vazire
Editor in Chief
Collabra: Psychology