

Peer Review Comments

Article: McHugh, C., et al. (2017). Searching for Moral Dumbfounding: Identifying Measurable Indicators of Moral Dumbfounding. *Collabra: Psychology*, 3(1): 23, DOI: <https://doi.org/10.1525/collabra.79>

Article type: Original Research Report

Editor: Yoel Inbar

Article submitted: 02 February 2017

Editor decision: Accept Submission

Revision submitted: 27 July 2017

Article accepted: 31 July 2017

Article published: 04 October 2017

Responses for Version 1

Reviewer: This reviewer has chosen to remain anonymous.

Affiliation: --

Competing Interests Statement: I have no competing interests to declare.

Review Completed: 07 March 2017

Recommendation: Revisions Required

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?

Overall, this is a well-written, straightforward, and timely paper. I greatly appreciate that the authors are helping to address our field's issues with replicability, in a simple and systematic manner.

One thing that I would like to see clarified:

p. 27: "participants were recruited from English speaking countries or from countries where residents generally have a high level of English (e.g., The Netherlands, Denmark, Sweden)."

For the purposes of generalizability, it would be very useful for the authors to provide a better indicator of sample composition (e.g., what percentage of participants were from the USA, what percentage from the Netherlands, etc.).

2) Figures/tables/data availability:

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

Underlying data and syntax for study 2 should easily allow others to check / reproduce the authors' results. However, when I open the data for study 1 in SPSS, the file for some reason appears to be blank. As such, I cannot yet fully review this part of the submission.

3) 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.

Details of appropriate ethics review were provided.

4) 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 paper is well-written and straightforward, and should be easily understood by people from a broad range of disciplines.

Reviewer: This reviewer has chosen to remain anonymous.

Affiliation: --

Competing Interests Statement: No competing interests declared.

Review Completed: 24 February 2017

Recommendation: Decline Submission

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?

In this paper, the authors present new evidence for the existence of the often-cited, but rarely-replicated, phenomenon of moral dumbfounding. I am a proponent of increased replication, so, while I suspect that some might criticize this paper for not "advancing theory" or being "novel enough", I will not. Replication is as valuable to the scientific enterprise as theory-building, because if we build our theories on unreplicable results, we haven't really accomplished much of anything. That being said, I have several concerns with the studies themselves, and I question how much they really tell us about the replicability of moral dumbfounding, as it is usually defined.

PRIMARY CONCERNS

I. I have serious concerns about taking "unsupported declarations" of wrongness and "tautological responses" to be instances of moral dumbfounding. In Haidt and colleagues' original definition, moral dumbfounding is "the stubborn and puzzled maintenance of a moral judgment without supporting reasons". The word "puzzled" and the absence of reasons seem very important here. It is one thing to say "I don't know why I think this is wrong" (i.e., to admit a lack of reasons, as in Studies 3a and 3b) - this would have all of the canonical features of dumbfounding. But to say that something is "just wrong" does not imply puzzlement. Rather, it is expressing a propositional belief, which could be considered a kind of lay deontological judgment (Royzman et al., 2015 make this point). By way of analogy, if I hold the propositional belief that lemons are yellow, and someone asked me why I believe this, I would not be dumbfounded, I would respond with an "unsupported declaration" and say, "because it's true that lemons are yellow" (probably followed up with "...and what is wrong with you, even asking this question? Everyone knows lemons are yellow!"). Yellowness is just a property that lemons have, and to many people, moral wrongness is just a property that incest has. "Unsupported declarations" do not imply a lack of reasoning or "puzzled" maintenance of a judgment; they are at least as likely to be sincere reasons that are appealed to in response to a baffling question. All of

this is to say that Study 2 probably *greatly* overestimates the prevalence of moral dumbfounding by confounding it with propositional beliefs about the nature of different kinds of acts (see Goodwin & Darley, 2008, for evidence that people think of moral beliefs as being nearly as objectively true or false as scientific facts). A lot of research has shown that people are roughly deontological in many moral judgments, so by rejecting these kind of rule-based justifications for their responses, the moral dumbfounding paradigm holds participants to a very "utilitarian" standard of evaluation that may be alien to them.

II. This point dovetails with the fact that many "unsupported declarations" and "tautological responses" were provided by participants who said that they did have a reason to support their judgments. This is *prima facie* evidence that these participants did consider these responses to be valid reasons. I think it is therefore inappropriate to code these responses as instances of dumbfounding. The participants offering them are not puzzled, and they do have reasons to support their judgments (the authors rightly point out that we cannot know if these reasons *caused* the judgments or not, but that is a separate question). The only true instances of what Haidt et al. call moral dumbfounding are the cases in Study 1 where participants admit to not having reasons (though the authors do not report how common this was, because they combine these instances with "unsupported declarations" in their coding scheme) and in Studies 3a and 3b, when participants selected "it's wrong, but I can't think of a reason").

III. Of course, even these apparently canonical instances of moral dumbfounding are subject to the critiques raised by Royzman et al. (2015). I won't reiterate the entirety of that paper here, but their point about credulity and conversational norms seems especially relevant. Some participants may just follow the experimenter's lead (even when it is just presented on a computer screen) and say that they have no reason to support their judgment. But this does not necessarily mean that they really believe this. They could just be following conversational norms and trying to be "good participants" despite being incredulous that no harm will be done in the scenarios. The authors say that addressing these criticisms is "beyond the scope of the current research" (p. 36), but is it? If Royzman et al.'s critique holds water, then this paper is just replicating a bad measure of moral dumbfounding, instead of trying to find new evidence for it that avoids the issues associated with this bad measure. Replicating an uninformative result is not itself informative. I'm not committing strongly to the position that Royzman et al. are correct in their criticisms, but if they are, then even the instances of participants admitting to having no reasons for their judgments are problematic. I do not think the argument that this kind of social pressure is lessened in a non-interview paradigm is sufficient; participants know that the experimenter is the one presenting them with the counter-arguments on the computer screen, so there is still pressure to accept the information presented by the authority figure.

IV. This leads to one final conceptual issue. This is not a criticism of this paper in particular, but rather an issue with how dumbfounding is defined across the literature (including here). The problem is that you can probably elicit all of the symptoms of "moral dumbfounding" even for classically immoral, harmful acts. Suppose that participants read a scenario about murder, instead of incest. They would surely say that it is wrong. When pressed as to why, they might respond with an unsupported declaration, "because murder is wrong." If asked why murder is wrong, they might then show all of the classic signs of dumbfounding, because there is not another "deeper" level of explanation than this. Indeed, this may be what is happening with the Trolley and Heinz scenarios, which typically are not thought to elicit dumbfounding, but do to some degree in these studies. The moral principles to support participants' judgments are "killing is wrong" and "saving lives is more important than making money", respectively. The reasoning goes no deeper than this. This raises questions about what, exactly, we are studying when we study "moral dumbfounding."

OTHER COMMENTS

-The authors frequently refer to "dumbfounding behaviors", but their DVs are not behavioral. What they are studying are judgments. This should be changed.

-I find the claim that “the large man is not presented as a character with competing interests” (p. 10) to be implausible. It’s true that we do not know anything about this man, but we can probably infer that, at the very least, he has an interest in continuing to live, which conflicts with “Paul’s” interest in stopping the trolley.

-Why were the Meaning in Life Questionnaire and Centrality of Religiosity Scale included in these studies? No theoretical rationale is ever provided for this.

-There is a typo on page 12: the participant who was neutral toward Trolley in his/her revised judgment is grouped with the participants who rated Paul’s action as wrong.

-It is never mentioned who did the coding of open-ended responses. Was it one of the authors? A research assistant? Were they aware of the purpose of the study, or blind to it?

-On page 35, the authors refer to their studies as “experiments”, but since there was no experimental manipulation, this is not technically correct.

-The Cannibal scenario in Appendix A does not show the revisions that are described on page 10.

CONCLUSION

I think that making a serious attempt to replicate a widely cited finding that has only been directly demonstrated once in an unpublished study with a small sample is a worthwhile endeavor. However, the authors are, in my view, excessively liberal about what qualifies as an instance of moral dumbfounding (I and II above), and they do not address recent methodological criticisms of the general paradigm that they are working in (III), so they may well be replicating a bad measure of a phenomenon that is not well defined in the first place (IV). Could these issues be addressed in a revision? Maybe. But it would have to be a very substantial one.

2) Figures/tables/data availability:

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

The figures in this paper are awfully hard to look at. Can the authors remove that dark gray tinge over everything?

3) 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.

No comments

4) 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 text is clear and readable.

Reviewer: This reviewer has chosen to remain anonymous.

Affiliation: --

Competing Interests Statement: None

Review Completed: 03 March 2017

Recommendation: Revisions Required

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'm so glad to see that someone is providing peer-reviewed evidence for moral dumbfounding, given the importance of the phenomena within the field I cannot believe it has taken 15 years. That said, across studies, I was unsure how to interpret the pattern of results because there were features of data which seemed largely consistent with some of Royzman's criticisms of the original work. I think for the paper to move the field forward it needs to be more theoretically precise and discuss ambiguities in the results in more detail.

Theoretical suggestions:

(1) Be more precise in your definition of moral dumbfounding. Firstly, I think it would be useful to explicitly specify if dumbfounding occurs when: (1) the participant fails to provide reasons; (2) the participant is unaware of the reasons motivating choice. The phrasing of the definition is ambiguous in this regard. Given the operationalizations, I currently interpret the definition to mean without providing the experimenter reasons to support their judgment. This definition would mean that if I have a reason but choose not to provide it, let say because of the presence of an injunctive authority, it would still be moral dumbfounding. That seems wrong...

Secondly, I worry the second phrase of the current definition ("even though they provide no logical reason to support their judgement") is a departure from Haidt's definition "they tend to "stubbornly" maintain their disapproval of the act without supporting reasons." The word logical seems curious and is different from Haidt's definition. For example, while perceived consensus would be a valid reason under Haidt's definition, it is not clear this reason would be considered logical under the present definition.

Also, the omission of the phrase "stubbornly maintain" is curious. Under the current definition is dumbfounding simply an unsupported declaration? Haidt uses the image of rummaging for an object in one's pockets and coming up empty-handed as a metaphor for moral dumbfounding. This illustrative image suggests that dumbfounding involves a search process not simply a type of response.

I know dumbfounding thus far has an "I know it when I see it" feel, but I think definitional specificity would force a tighter coupling between theoretical and empirical operationalizations.

(2) The results and operationalizations need to be placed in the context of the theoretical claims associated with moral dumbfounding. When we look at the broader context what intellectual work does dumbfounding do? Moral dumbfounding matters insofar as it provides evidence that: (1) moral judgements can be driven by emotions/intuitions; (2) "moral reasoning does not cause moral judgment; rather, moral reasoning is usually a post hoc construction, generated after a judgment has been reached" (p. 814, Haidt, 2001). Claim 1 (that moral judgments can be motivated by intuitions or emotions) has been shown exhaustively elsewhere. Whereas this question feels unique to the dumbfounding account is the post hoc rationalization. This suggests that the search process is what the paper needs to demonstrate. I understand how the direct replication provides evidence for this.

However, given the pattern of results across studies 2 and 3 I am not sure they support or even test this claim.

(3) Discuss conversational norms and other potential confounds. The high sensitivity of response to the method of elicitation seems more consistent with Royzman's conversational norm argument. IF the WAY of asking is substantially altering response (far more the content of what is being asked), I worry it has more to do with norms and less to do with the intuitive Kantian. Royzman discusses this far more compelling than I can, but given the centrality of this criticism I thin it should to be discussed in detail in the introduction.

Methodological suggestions:

(1) I really think conversation norms are a huge confound.

a. The structure of the response options seems to create different demand effects. I would run a follow-up study looking at the social desirability of the responses across studies 2 and 3.

b. Given the degree to which conversational norms might shape dumbfounding, I worry the "emotionally" motivated scenarios may involve different conversational norms as these scenarios (incest and cannibalism) seem to systematically differ in the degree to which they involve taboos. I could imagine if a person feels uncomfortable discussing an issue he or she might be more likely to say "I don't know" or make unsupported declarations to avoid further discussing it. This would also be true outside of morality. For example, in the domain of preferences I could imagine a participant asked to justify liking golden showers (taboo) would appear more "dumbfounded" than if he or she were asked to justify liking apples (non-taboo). This effect could occur simply because the golden showers condition is much more uncomfortable for people (baring sex workers and moral psychologist) to talk about.

(2) Again, I really think there needs to be evidence for the post-hoc nature of justifications. Without this I'm not convinced the paper is testing dumbfounding. Here are some thoughts which might help address this concern.

a. This argument, that people are intuitive moral prosecutors is reminiscent of Tetlock's work on the intuitive prosecutor which applies a flexible correction model punishment decisions. One could, as Tetlock did, use some of very clever manipulations from Wegener and Petty to test if the justifications are indeed posthoc.

b. Another possibility, would be to adapt some of Nisbett's early work on mental processes. If you found the right paradigm, I could imagine it being really pretty.

(3) I don't think the authors stringently tested moral dumbfounding as a construct, but rather tested Haidt's paradigm. I find this worrisome because that work never underwent peer review.

I think this paper is testing an important idea in moral psychology and more support for this idea should be published. However, I don't think one can draw clear conclusions about it from the current set of studies. I worry publishing the paper in the present form would misrepresent support for moral dumbfounding. While I would prefer an additional more incisive study, at a minimum its important the discussion in the intro and abstract need to be significantly more qualified.

2) Figures/tables/data availability:

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

n/a

3) 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.

n/a

4) 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. Although there were some errors in the references.

Additional Comments:

There were some smaller methodological concerns I had with studies 2 and 3, however, given I did not think these studies were testing the construct, smaller concerns seemed somewhat irrelevant.

Also, I think that the anilingus example originates with Paul. He was looking for something in between scat and cunnilingus. I think him and Ed have a paper on this under review somewhere.

Editor Decision for Version 1

Editor: Yoel Inbar

Affiliation: University of Toronto, CA

Editor decision: Revisions Required

Decision date: 09 March 2017

Dear Mr Cillian McHugh,

Thank for your submission to Collabra: Psychology, "Searching for Moral Dumbfounding: Additional Evidence for the Existence of Moral Dumbfounding". I was lucky enough to obtain reviews from three highly qualified experts. Based on their comments, I am asking you to revise the manuscript.

The reviewers raised a number of excellent points, and I am not going to repeat them all here. The most important concern, it seems to me, is whether the operational definition of dumbfounding you use in your MS actually captures the concept of dumbfounding as described by Jon Haidt et al. Both Reviewer B and Reviewer C highlighted a key aspect of Haidt's definition: that moral dumbfounding consists of the "stubborn and puzzled" maintenance of a moral judgment, even without supporting reasons. (Reviewer C points out a nice metaphor used by Haidt--moral dumbfounding is akin to "rummaging for an object in one's pockets and coming up empty-handed.") The reviewers argue persuasively that simply stating that (for example) "incest is wrong" does not qualify as moral dumbfounding by this standard--it could simply be the endorsement of a deontological principle (and, as Reviewer B points out, we know that people are often intuitive deontologists). I think you need to deal with this issue, either with new data or by re-analyzing the data you have already.

I also think you need to engage more seriously with Royzman et al.'s "conversational norms" alternative explanation of moral dumbfounding effects, either with new data or expanded discussion. Their objections are persuasive enough that they deserve to be addressed in more than a single paragraph in the general discussion.

One minor point that struck me was that the first paragraph of your results sections often had a long list of descriptive statistics. These were a bit overwhelming and I think should be in tables.

Given that all three reviewers were excited by the challenge you've taken on with this paper, I do hope that you are willing to undertake the revisions necessary to deal with their critiques. If you do submit a revision, please outline in your cover letter how you have addressed each of the reviewers' points.

The full review information should be included at the bottom of this email. There may also be a copy of the manuscript file with reviewer comments available once you have accessed the submission account.

To access your submission account, follow the below instructions:

- 1) login to the journal webpage with username and password
- 2) click on the submission title
- 3) click 'Review' menu option
- 4) download Reviewed file and make revisions based on review feedback
- 5) upload the edited file
- 6) Click the 'notify editor' icon and email the confirmation of re-submission and any relevant comments to the journal.

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 do contact us.

I am asking you to submit your revisions to me within 60 days (that would be June 4). If you cannot make this deadline, please let me know as early as possible.

Kind regards,

Yoel Inbar

yoel.inbar@utoronto.ca

Author's Response to Review Comments for Version 1

Author: Cillian McHugh

Affiliation: Mary Immaculate College, University of Limerick, IE
Revision submitted: 27 July 2017

Editor Decision for Version 2

Editor: Yoel Inbar
Affiliation: University of Toronto, CA
Editor decision: Accept Submission
Decision date: 31 July 2017

Dear Mr Cillian McHugh,

After review, I have reached a decision regarding your submission to Collabra: Psychology, "Searching for Moral Dumbfounding: Additional Evidence for the Existence of Moral Dumbfounding", and are happy to accept your submission for publication, pending the completion of copyediting and formatting processes. Thank you for a very responsive revision that, in my opinion, really improved the paper.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. The accepted submission will now undergo final copyediting. You will be contacted once this is complete to answer any queries that may have arisen during copyediting and to allow a final chance to edit the files prior to typesetting. If you wish to view your submission during this time, you can log in via the journal website.

Kind regards,

Yoel Inbar

yoel.inbar@utoronto.ca