**Manuscript submission**

Collabra: Psychology

Editor in Chief

Dear Prof. Vazire:

Attached please find the manuscript “***Caring about (COVID-19 related) social issues signals trustworthiness: Direct and conceptual replication of Zlatev (2019)***” by Angela Dorrough, Nathalie Bick, Lukas Bring, Caroline Brockers, Charlotte Butz and Iris Schneider which we would like to be considered for publication in *Collabra: Psychology*. The manuscript contains four studies. Three of these are studies with samples of predominantly university students (*n* = 1087) and one sample (*n* = 210) is representative for the German population in terms of age and gender.

In all studies, we replicate and extend previous findings on caring and trust, published in *Psychological Science (Zlatev, 2019)*. In line with this work, we consistently find that targets’ level of caring about an issue influenced perceived integrity-based as well as benevolence-based trustworthiness. Importantly, we also extend the original findings by showing that these results are generalizable to various topics, including issues related to the current COVID-19 pandemic (e.g., closing national borders to prevent the spread of the virus). This is especially relevant, because in contrast to many of the issues used in the original research, the issues included in our studies were of direct relevance to the participants, spanning different areas of everyday life.

Furthermore, we also provide behavioral evidence for the effect by showing that money transfers to targets in an incentivized trust game increase with a target’s level of caring about an issue (i.e., installation of wind turbines in the North Sea). Finally, we also include a mini meta-analysis including our and the original findings, and we conclude that the effects of target caring on trustworthiness are small, but very robust.

We believe that our manuscript is of interest for the readership of *Collabra: Psychology* for at least three reasons: First, trust and trustworthiness are constructs that are of high interest and relevance for researchers in different domains, as well as the public and policy makers. Especially during a worldwide crisis, trust (e.g., in scientists and politicians) is essential in fostering adherence to protective policy. Considering the current COVID-19 pandemic, it is no exaggeration to say that in these instances, trust can save lives. This makes it especially important to understand whether previous findings in the area of trust translate also to high stake situations. Second, replication studies like ours provide more certainty about empirical findings and the size, generalizability, and robustness of effects, and can inform theory building and development. Third, our work is a perfect fit to methodological rigor valued by *Collabra: Psychology:* we vigorously pursue open science principles, in line with the journal aims outlined in the submission guidelines. Specifically, our hypotheses were pre-registered and all files (data, analysis scripts and materials) are openly available on the Open Science Framework. Furthermore, our sample sizes were a priori determined to achieve high power for testing the effects reported in Zlatev (2019).

This paper has previously been reviewed by *Social Psychological and Personality Science (SPPS)*. As you can see from the reviewer comments, the reviewers, one of whom was the author of the original paper, highlighted the diligence with which we executed the replication. According to the previous editor’s action letter the reason for not accepting the manuscript was mainly that the relevance of the phenomenon was not considered sufficiently high and that *SPPS* does not have a distinct section for replication studies. Against this background – and as stated in your submission guidelines – we request a streamlined review process. One of the reviewers (Julian Zlatev) signed his review. He and the editor (Igor Grossmann) agreed that their comments can be openly available at *Collabra: Psychology* when the submission is accepted. The other two reviewers are unknown to us.

Although the reviews by Zlatev and the anonymous reviewers were very positive overall, the reviewers had some suggestions on how to improve our manuscript. We addressed most of these issues in a revised version of the paper which we now submit to *Collabra: Psychology*. Please find the point by point response to each of these previous comments and suggestions in the detailed response section below.

We confirm that this work is original, not previously published, and not under consideration by another journal. The rights of the research participants have been protected in the reported studies and all authors have approved the submission of the paper.

Thank you for handling our manuscript.

Sincerely,

Angela Dorrough, Nathalie Bick, Lukas Bring, Caroline Brockers, Charlotte Butz, & Iris Schneider

*Original comments by the reviewers are in italics.*

**Reviewer 1:**

*To start, I want to disclose that I am the author of the original paper that this paper is replicating. I appreciate the fact that these studies were carefully and thoughtfully designed and analyzed. I also found the paper to be well-written and easy to follow.*

*Overall, I think these studies extend the original work in interesting and helpful ways. The use of a representative sample (in a different country than the original studies) and the wide variety of issues shown to participants are particular strengths. However, I found the distinction made between issues that most people have direct experience with (e.g., COVID-19 precautions) versus those people don’t (e.g., capital punishment) to be a bit less relevant. It’s not clear to me why this would ex ante affect results.*

*One potential reason could be that people might care more about directly relevant issues than indirectly relevant ones. This however seems to be better measured by the “participant caring” question, which is separately analyzed in the SOM. There may be other reasons why this distinction would be important, but if so then I think those reasons should be laid out more explicitly. Also, if this does appear to be an important distinction, then it would be helpful to measure it directly and look for moderation (similar to what is done with perceived controversy in the SOM).*

Since we had a word limit for our previous submission, we reported the results concerning potential moderating effects in the supplementary materials. Based on the reviewer’s comment we now report the respective findings in the main paper in a separate section on pp. 12:

**“Moderator Analyses for all Reported Studies**

In the studies reported in this article we included additional measures for (in part exploratory) moderator analyses. The results of these analyses are reported in what follows.

***Perceived Controversy***

The first potential moderator we examined is the perceived controversy of a social issue that is how controversial an issue is within society. Research in other domains examining the effect of perceived attitudes or attitude involvement has shown that the type of topic matters for the social judgment people make about the attitude holder. For example, discussing a controversial topic provokes anxiety and discomfort and can result in negative evaluation of the interaction partner (Simons & Green, 2018). Furthermore, when actors express their opinion on a social issue, controversy of the topic influences how a person is perceived on various dimensions (e.g., Beall et al., 2017). For example, expressing strong feelings about a controversial topic makes people seem more competent than expressing strong feelings about trivial topics (Pillaud et al., 2018). Thus, it seems that the degree to which a topic is controversial matters for the degree to which caring about a topic and expressing an attitude about a topic influences perception of the interaction partner or the attitude holder.

Thus, for Studies 2-4, we pre-registered two additional hypotheses – namely, that the effects of a target’s level of caring about a social issue on perceived integrity-based trustworthiness and perceived benevolence-based trustworthiness are moderated by perceived controversy of the social issue. To test these hypotheses, perceived trustworthiness was predicted by a target’s level of caring about a social issue, perceived controversy of the social issue as well as the interaction of perceived controversy of the social issue with a target’s level of caring about a social issue controlling for participant-target agreement on the social issue. For Study 1, this analysis was conducted on an exploratory basis. For perceived integrity-based trustworthiness, such an effect has been observed for Study 1 only (*ß* = 0.08, *p* = .025; *p* > .227 for all other studies), whereas for perceived benevolence-based trustworthiness we did not observe such an effect (all *p* > .072). Thus, due to the small effect size and its inconsistency, we cannot make strong claims concerning such a moderation until further studies provide more evidence.

Besides perceived controversy, we explored two additional potential moderators – perceived familiarity of an issue and participant caring – as indicators of the relevance of a topic for participants.

***Perceived Familiarity***

In Study 4 we assessed perceived familiarity of a social issue as an additional potential moderator. In exploratory analyses, we predicted perceived integrity-based trustworthiness and perceived benevolence-based trustworthiness by a target’s level of caring about a social issue, perceived familiarity of the social issue as well as their interaction controlling for participant-target agreement on the social issue. These analyses did not reveal significant interaction effects (both *p* > .081).

***Participant Caring***

We ran additional exploratory analyses concerning a potential moderating effect of participant’s level of caring about the social issue. For perceived integrity-based trustworthiness, we observed such an effect in three out of four studies (Study 2: *ß* = 0.10, *p* < .001; Study 3: *ß* = .09, *p* < .001; Study 4: *ß* = .06, *p* = .006). For perceived benevolence-based trustworthiness, the interaction also became significant in three studies (Study 2: *ß* = 0.09; *p* < .001; Study 3: *ß* = 0.09; *p* < .001; Study 4: *ß* = 0.05; *p* = .022). Although not significant in all cases, it seems that the effect of a target’s level of caring about a social issue on perceived trustworthiness is increased for participants who care about the respective social issue themselves.

***Participant-Target Agreement***

In the original article, Zlatev (2019) reported rather inconsistent findings concerning the interaction between target-participant agreement on a social issue and a target’s level of caring about the social issue in predicting perceived trustworthiness. In contrast, we find that the influence of a target’s level of caring about a social issue on perceived integrity-based trustworthiness (Study 1: *ß* = .09; *p* = .024; Study 2: *ß* = .14; *p* < .001; Study 3: *ß* = .17; *p* < .001; Study 4: *ß* = .07; *p* = .002) and perceived benevolence-based trustworthiness (Study 1: *ß* = .12; *p* = .002; Study 2: *ß* = .11; *p* < .001; Study 3: *ß* = .12; *p* < .001; Study 4: *ß* = .06; *p* =.006) consistently increased with participant-target agreement on the social issue.”

In the discussion section of our revised manuscript, we return to these potential moderators.

p. 17: “We tested perceived controversy of a social issue as a moderator in all studies but due to the inconsistent findings, we cannot make strong claims concerning such a moderation until further studies provide more evidence. However, in exploratory analyses we identified two other moderators that showed quite consistent results across the studies, namely participant’s level of caring about a social issue and participant-target agreement on the social issue. Specifically, the effect of a target’s level of caring about a social issue on perceived trustworthiness increased if the participants cared about the issue. Furthermore, the effect of a target’s level of caring about a social issue on perceived trustworthiness also increased when participant and target had the same stance on the issue (i.e., supported vs. not supported the issue). Future research could investigate additional circumstances under which the effects are more or less pronounced.”

*It may be interesting to compare the relative effect sizes of agreement and caring on the outcome measures across studies. For example, the beta coefficients for agreement are fairly large, relative to caring, in Studies 1 and 2. However, this isn’t the case for Study 3. The only major difference between these studies is the type of issues presented to participants. It seems useful to note that perceptions of trustworthiness as a function of agreement appear to be less pronounced in COVID relevant issues compared to more general social issues (which would not necessarily have been my prediction ex ante). Obviously more work would be needed to test this more systematically, but it could be included as an interesting future direction.*

We thank the reviewer for noting this interesting pattern. We added a comment on these findings, and a call for additional studies to the discussion section of our paper.

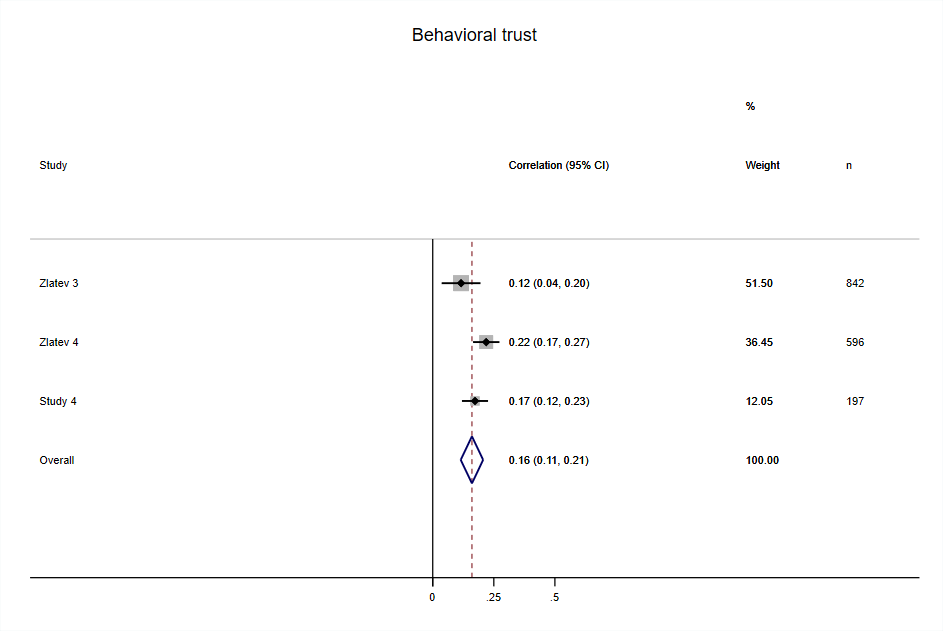
pp. 16: “Interestingly, in all studies but Study 3, whether participant and target agreed on an issue seemed to be of higher importance for perceived trustworthiness than a target’s level of caring about the issue (see beta coefficients in Tables 1-4). The social issues in Study 3 were different in that they concerned COVID-19 related recommendations to prevent the spread of the virus. It is possible that in times of crisis and when the issue is essential to survival, caring about an issue is at least as important as agreeing on the issue. Future studies could investigate this pattern of results in more detail.”

*It might also be worth discussing some of the supplementary analyses (e.g., moderation by perceived controversy) a bit more in the main text of the paper. Even though the results are often inconclusive, they are still asking interesting questions and are extending the original work in a novel way.*

As we already mentioned in response to a comment above, we followed this suggestion and moved the findings concerning perceived controversy and participant caring to the main article (p. 13).

*It would be helpful to include a forest plot for the third meta-analysis (on behavioral measures) along with the forest plots for the first two (figures 1 and 2). It would also be helpful to include a table listing the issues used in each study so that they are all easily available in one place.*

We followed the reviewer’s advice and added a third forest plot to the paper (p. 16).



A table listing all issues used in each study can be found in the supplementary materials (Table S1).

*For Study 4, I assume the transfer measure was kept as a 0-3 variable, rather than analyzed as a binary outcome (transferred something vs. transferred nothing). If that’s the case, it may also be interesting to look at whether most of the effect is explained by people giving something vs. nothing, or whether there’s an additional effect of amount transferred (conditional on transferring something).*

We rerun the respective analyses with a binary outcome (transferred something vs. transferred nothing), as with amount of transfers (excluding zero transfers). In both analyses we find a significant effect of target caring on trust. We report these findings briefly in the main paper.

p. 11: “As displayed in Table 4, we find the effect of target caring for all three outcomes. All conclusions remain unchanged when controlling for participant age and gender as well as social issue using dummy variables (SOM Table S2). The conclusions for behavioral trust remain unchanged when excluding transfers of zero or using a binary outcome (transfer yes vs. no).”

*Small typo: “In sum, Study 3 will give us more insight into the generalizability of the effect over populations and situations.” (pg 9) -this should be “Study 4”*

We changed “Study 3” to “Study 4”.

**Reviewer 2:**

*The manuscript “Caring about (Covid0-19 related) social issues as a signal of trustworthiness: a direct as well as conceptual replication of Zlatev (2019)” details four studies and a mini meta-analysis. The paper boasts four replications and extensions of Zlatev’s original paper. The studies appear to be well-conducted, appropriate and add to the literature (in terms of supporting the original paper and extending it to a novel topic). The mini meta-analysis is particularly useful and I applaud the author(s) for adding this to the paper!*

*1. The mini-metanalysis should be called out as its own study (not subsumed under Study 4). It doesn’t just summarize the manuscript’s effect sizes, but those of Zlatev’s as well. I would also recommend adding it to the abstract. Relatedly, there’s no Figure/Table for the behavioral trust meta-analysis – please add for comparison/full reporting. If there’s too few to warrant a Figure/Table, then please report the effect size, CI, and N for each sample.*

Based on this suggestion, we moved the mini-meta to a separate chapter and refer to it in the abstract (“To provide best estimates for the effect of a target’s level of caring on perceived trustworthiness, we report results of a mini meta-analysis including our findings as well as the findings of the original research.”). We also added a forest plot for behavioral trust which has been also requested by Reviewer 1.

*2. The paper felt long. In places, words could be cut without loss of meaning (e.g., the title could be “Caring about (Covid0-19 related) social issues signals trustworthiness: Direct and conceptual replication of Zlatev (2019)”). Mostly, there were some technical details that could either be reworded or moved to a supplemental file.*

We changed the title as suggested by the reviewer. Since the other reviewers were rather in favor of adding more information to the paper, we decided not to move any methodological/technical details to the supplementary materials. If the editor thinks we should shorten the paper, we are of course happy to do so.

*3. Zlatev (2019) tested for interactions (largely finding none) between caring and agreement. Yet, I don’t see any tests of interactions in the current manuscript. Was this done on purpose? (and if so, please provide the rationale) or was this an omission?*

Due to the limited scope for the first submission, we did not report the respective findings. Based on this comment by the reviewer we now report these findings in the newly added sections on potential moderating effects (pp. 12).

*4. Also related to the regression analyses, how was target caring coded? I’m assuming agreement was dummy coded (yes = 1; no = 0???).*

Yes, agreement was coded as follows: yes = 1; no = 0. Target caring ranged from 1-100. We added this information to the manuscript as well:

See for example p. 5: “As pre-registered and in accordance with Zlatev (2019), we ran two linear regression models predicting perceived integrity-based trustworthiness (Table 1, Model 1) and perceived benevolence-based trustworthiness (Table 1, Model 2) by a target’s level of caring about a social issue (1-100) and agreement (yes = 1 vs. no = 0) between participant and target on the social issue”

*5. In studies 2, 3, and 4, the first section under methods should be a description of the participants (matches the norm and the reporting in study 1).*

The revised version of the manuscript includes a participant section for each study.

*6. The Supplemental File should follow the order in which items are called out in the text, much like an Appendix. This revision would make it easier to follow and find sections.*

The updated supplementary materials follow the order in which the items are mentioned in the text.

*Minor issues:*

*On page 2, final paragraph: “three important ways” is stated but there are four items listed in that paragraph.*

We changed this to “five important ways” (p. 2) since we now also highlight that we investigate potential moderating effects.

*Page 4, not sure what is meant by “see next subchapter”.*

We changed this to “see method section for the exact wording of these questions” (p. 4).

*Typo on page 9 – I believe you mean Study 4, not 3, in the last sentence of the discussion section. In the same paragraph, “second” is called out but I don’t see “first” called out.*

The reviewer is right concerning the typo. We fixed it accordingly (p. 9). We also elaborate more on what we mean by “second”: “A second potential limitation of the studies reported so far is that most of our samples consisted of primarily students” (p. 9).

*No need to repeat the average effect sizes for the meta-analysis in the discussion. An evaluation of their size (e.g., small, medium, vs. large) is more meaningful/less redundant.*

To address this comment, we added the evaluation of the effect sizes to the main text. However, we decided to still report the exact number in brackets since they can only be estimated from the figures.

p. 14: “All effect sizes were transformed into Pearson’s r. Analyses reveal small effects for perceived integrity-based (weighted mean effect size of *r*M = .24) and perceived benevolence-based (*r*M = .20) trustworthiness as well as behavioral trust (*r*M = .16). Thus, pooled evidence across all studies points to a small effect of a target’s level of caring about a social issue on perceived trustworthiness”

In the general discussion (pp. 16) we now omit the exact numbers (also to avoid repetition).

*Supplemental File, Table 1 – you can reduce some of the clutter by including M (SD) under the column header “perceived controversy” (in the same cell) and then report all numbers that way (e.g., 7.21 (2.12)) and remove all M = and SD = from the cells.*

We changed the table according to this suggestion (see Table S1).

*The reference for G Power is missing from the Supplemental file.*

We added this reference to the main paper since we now report power considerations more prominently in the main article.

*Reviewer 3:*

*In the current manuscript, the authors provide direct and conceptual replications of Zlatev's (2019) intriguing finding that people will trust a person even when that person disagrees with them on controversial social attitudes. The work is diligent and described with clarity, but I do not think it rises to the level of contribution associated with SPPS. To be sure, replication is important, even crucial, and the mix of direct and (arguably) conceptual replication is a plus, but unless there is a special program at SPPS devoted to publishing replications, I do not think this manuscript rises to the requisite standard. It should be published somewhere, but perhaps not here.*

*The manuscript does provide one intriguing extention of the Slatev work, examining trust at the behavioral level via the traditional trust game. However, how that game worked is unclear. The authors never really describe how their transfer and return of "Talers" will be converted to actual euros, if at all, even in the supplemental materials. They are also opaque about where exactly their participants in Studies 2 and 3 come from. They mention that these are student samples in the introduction, but never specify or confirm that fact in the methods.*

On p. 12 we now state that 1 Taler amounts to 30 Cents (“1 Taler = 0.30 €”). Furthermore, based on this comment and one suggestion by Reviewer 2 we added a participant section to each study. This section includes the recruiting method and also mentions the percentage of students in the respective sample for Studies 2 and 3. For Study 1 we do not have information on whether the participants were students. However, the sample for Study 1 was recruited in a similar manner as the samples for Studies 2 and 3.

p. 6: “A total of 284 participants completed the study. According to the participant code, one participant took part twice. Thus, we excluded the second participation, resulting in *N* = 283 participants (196 female, 87 male; 18-81 years of age; 41.7% students)”

p. 9: “A total of 294 participants completed the study. Three participants participated twice – again we excluded the second participation, resulting in *N* = 291 participants (208 female, 80 male, 3 non-binary; 18-65 years of age; 60% students). “

*Other detail would be informative. How correlated were integrity- and benevolence-related trust? I presume they were very correlated. The authors never really test whether their findings interact with specific social issue, which is a surprise given their focus on generality.*

We now report the correlation between integrity- and benevolence-based trustworthiness for each study.

See for example p. 4: “Perceived integrity-based trustworthiness was measured using a German translation of three items (*M* = 4.28, *SD* = 1.43, α = .78) taken from Zlatev (2019): “This person has a great deal of integrity,” “I can trust this person’s word,” and “This person cares about honesty and truth.” Perceived benevolence-based trustworthiness was measured using a German translation of three items (*M* = 4.01, *SD* = 1.49, α = .79; correlation between perceived integrity- and perceived benevolence-based trustworthiness: *r* = .68) taken from Zlatev (2019)”.

In the supplementary materials we report analyses including dummy variables for the social issues as well as the interactions with a target’s level of caring about a social issue. Results are robust to these additions. Furthermore, compared to the grand mean over all issues we find interaction effects for single social issues. Since they are inconsistent and difficult to interpret we would rather not report them in. If the editor thinks these findings are particularly interesting, we are happy to add a section discussing this to the paper.

*To rise to the level of contribution associated with this journal, and many others, the authors would be wise to identify some theoretically-driven "twist" that informs our understanding of Zlatev's effects. For example, are there issues in which caring fails to produce trust in a theoretically meaningful way. For example, what if the issues were economic rather than social; that shifts the entire set of norms people would use to judge others. Would people trust a person who really cared about keeping the minimum wage low, or cutting capital gains taxes on rich people? Extending the phenomenon to behavioral trust, as the authors have done here, is a start, but a more ambitious and extensive foray into this matter would have been more impressive.*

We think it is a great idea to also look at economic issues in future studies. It seems possible that if someone cares about cutting taxes for rich people, this has a negative effect on benevolence, but not so much on integrity-based trust. But of course, without additional data, this is only speculation. We add a respective call for future research to the discussion section of our article.

pp. 17: “Future research could investigate additional circumstances under which the effects are more or less pronounced. One idea is to include economic issues instead of or in addition to social issues in future research. It is conceivable that if someone cares about the issue of (not) cutting taxes for rich people for example, this has a negative effect on benevolence-based (since benevolence for example conveys a target’s kindness or positive intentions), but not so much on integrity-based trustworthiness (i.e., a target’s perceived honesty or virtuousness).”