Reply to PCIRR Stage 2 decision letter #726:
Monin and Miller (2001) replication and extension

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response to each item. Please note that the editor's and reviewers' comments are in bold with our reply underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/aZABLKBxZhyd

A track-changes manuscript is provided with the file: "PCIRR-S2-RNR-Monin-Miller-2001-manuscript-v1-QX-track-changes.docx" (https://osf.io/9ngez)

We note that we are not familiar with the titles and ranks of the reviewers, and looking for that information proves tricky. To try and err on the side of caution, we refer to all reviewers with the rank Dr./Prof. We apologize for any possible misalignments and are happy to amend that in future correspondence.

Reply to Editor: Dr./Prof. Chris Chambers

> **Two of the reviewers from Stage 1 kindly returned to evaluate your Stage 2 submission. As you will see, the reviewers are broadly satisfied with the manuscript and offer some minor suggestions for presentational clarifications, as well as a potential issue to address in the Discussion. Following submission of your revision and response, I anticipate being able to issue a final positive recommendation of your Stage 2 manuscript without further in-depth review.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit. We have responded to the reviewers' comments below and revised accordingly.

Other than changes based on the valuable suggestions, we also revised the part in the Introduction where we reviewed prior meta-analyses on moral licensing. These changes were due to an update of Rotella et al. (2023), a pivotal working meta-analysis. The update removed a few statistics reported in previous versions and opted for different meta-analytic strategies. Given the update, the statistics that we cited initially may not best reflect the findings of Rotella et al. (2023), and changes are therefore necessary to ensure consistency between our review and the literature.

Reply to Reviewer #1: Dr./Prof. Marek Vranka

I have just finished going through the Stage 2 submission of "Licensing via Credentials: Replication of Monin and Miller (2001)," and I must say, I'm thoroughly impressed! The authors have done a commendable job and fully adhered to the preregistered plan. I have only minor comments and suggestions, which are detailed in the enclosed Word document. Great work to everyone involved!

Thank you for the positive and supportive opening note and the constructive feedback.

1. Citation count: Not sure if it is necessary to update this, as it could be needlessly confusing: this part should stay as close to the Stage 1 version as possible.

We do not strongly prefer a change, but we see the point of keeping it exactly as in Stage 1. So we reverted back to what it is at IPA.

2. "from one's feeling/perception that they have established": Belief?

We prefer "feeling/perception" as "belief" implies relatively stronger awareness, which we believe is not emphasized in the theorization. We decided to stick with "perception."

3. Deviations: Maybe make clear these are deviations from the original study that is being replicated and not deviations from the Stage 1 protocol?

This is a good point. We have clarified.

4. "however, differed across conditions": Add ";" here?

There are statistics after the bracket so ";" is not needed here.

5. "if the results differed substantially": This can likely be dropped, as it is just a bit confusing (now you know where the results are presented).

Indeed. We have deleted this text.

6. "though that is possible": Past tense here?

We have revised.

7. "was not robust to different analyses": Can you dig a bit deeper into this? Is the effect significant only when the reputation concern variable is in the model? If I am not mistaken, the reputation concern was associated

in the opposite direction with the DV, so its presence in the model might explain the difference. Of course, this is all speculative, but the results begs the question of what is causing it.

Thank you for raising this point. Yes, there was only a credential effect when reputational concern was included in the model, but reputational concern itself did not predict the DV (here: gender or ethnicity attitudes; reputational concern was (only in the racism domain) associated in the opposite direction when the DV is gender or ethnicity preferences). We revised the text into the following:

> "This unexpected result, however, was not robust to different analytical strategies. We observed a statistically significant credential effect only when reputational concern—which nonetheless did not predict gender or ethnicity attitudes itself—was included in the model. Therefore, this evidence should be interpreted with caution."

8. "such as sampling error": Not sure what exactly do you mean by this and I would drop it.

We revised it into "such as characteristics specific to our sample."

9. "In this case, it appears reasonable to predict that if one were to observe a moral consistency effect": I would argue that because people see Black people as equally able, there is nothing to be consistent about, as picking the best Black candidate is completely unremarkable. This could be in line with expressing preference for Whites in DV, as the issue in that scenario could be seen as mainly caused by the hostile environment and not an inability. For women, selecting a female candidate might make an issue of positive discrimination more salient for participants, which could explain the effects in both ethnicity and gender scenarios, as some participants might perceived their decision along similar lines (i.e., to select an candidate for the unfriendly environment in order to change it for the better). Of course, it is only a speculation, you may or may not consider it. However, it would be good (in my opinion) to at least offer some possible explanations of the observed results (i.e., the heterogenity and the effect in the opposite direction).

Thank you for offering your interpretation of the results. We do find it a plausible interpretation, which unfortunately cannot be corroborated here. We noted this interpretation in the discussion:

> "One Stage 2 reviewer offered us one such plausible explanation. It starts from the observation that participants agreed much more strongly that Black people are as able as White people to do any kind of job than they did with the two genders. Given this

difference, one can reasonably speculate that (1) because Black people were perceived as equally able as White people, choosing the Black candidate was unremarkable and (from the perspective of the recruiter) would not necessarily make themselves appear moral, and (2) selecting a female candidate could have made "positive discrimination" salient to participants, and due to its controversy (Noon, 2010), participants went on to endorse more strongly that they would not consider gender and even ethnicity identities (for those who saw the ethnicity scenario following the non-sexist credential manipulation) in hiring. Unfortunately, we do not have other data that could not speak directly for or against this explanation."

10. "We do not suggest that the manipulation was certainly effective": I am bit confused by this sentence, did you mean "ineffective"?

Yes, that was what we meant. In view of the change above, we also changed this part. The revised text are pasted below after (11) since these comments address the same part.

11. "preferences are consistent with a priori expectations": Maybe be more clear about what expectations you have in mind here.

The revised text are pasted below:

"If true, the above explanation implies that the manipulation did not work to provide moral credentials and thus had low validity. Nonetheless, even if the explanation is not true, we may still not be able to conclude that the manipulation worked as intended. We can conclude based on the exploratory questions that participants' perceptions of different hiring decisions and preferences are consistent with a priori expectations (e.g., choosing the Black outstanding candidate is moral; not choosing him implies racism; hiring preferences neutral to candidates' gender is perceived to be the least prejudiced). It was however possible that this pattern of perceptions only emerged when participants actively reflected on those decisions and preferences. In other words, merely making a particular hiring decision may not be sufficient to prompt participants to think about what the decision implies and hence unable to create a moral credential that has an appreciable downstream effect, at least in our setting where situational reputational concern is low. Similarly, even if the hiring preference dependent measure can afford a moral credential effect in principle, participants might not be reflecting what those preferences mean for their self-images or appear to others when expressing them, making moral credentials less relevant. To what extent the manipulation creates moral credentials and makes participants believe that others would consider them non-sexist or non-racist is an important open question that our data could not address."

Reply to Reviewer #2: Dr./Prof. Štěpán Bahník

The report generally follows the pre-registered procedure and any possible deviations are sufficiently justified. Therefore, I have just a few minor comments.

Thank you for the positive and supportive opening note and the constructive feedback.

(1) I believe that according to the APA style commas should be included before and after statistics (e.g. "… and $t(868) = −3.01$, $p = .006$ for the gender…" should be "… and, $t(868) = −3.01$, $p = .006$, for the gender…").

Thank you. We are unfamiliar with this being an APA 7th Manual requirement and could not find that in the guide. We would please rather keep things as they are. We would gladly revise if given clear editorial guidelines, and if needed will attend to this in the copyediting stages.

(2) The results were quite hard to follow for me. This might be a matter of preference, so feel free to ignore the recommendation, but I believe that the results would be easier to understand if they used more concrete language. For example, "Also, domain-inconsistent credentials did not show support for an effect compared with controls…" could be something like "Also, participants who had domain-inconsistent credentials did not significanlty differ in their hiring preferences from participants without credentials…" or even "Also, participants presented with a Black (or female) star applicant did not significanlty differ in their hiring preferences of males (or Whites)—that is, in the different domain—from participants presented with a White male star applicant…"

Thank you for raising this concern. We have thoroughly revised the presentation of results according to your suggestions.

(3) I wonder whether a possible explanation for the lack of the credential effect might be a different perception of choosing the star applicant now and 20 year ago. It is possible that 20 year ago participants were more likely to feel that they are choosing the applicant "despite" their race or sex and now, they might be more likely to feel that they are choosing the applicant "because" of their race or sex. This would be partly suggested by the higher proportion of participants selecting the black/female star applicant than white male star applicant in the current study, while there was no significant difference in the original study (even though the study does not report the proportion of participants choosing the star applicant

> **in each condition). Unfortunately, we do not have data about the perception of morality of choosing the star applicant in the original study, but is there some result that would suggest against this explanation? The discussion now mentions the possibility of cultural change as well as of the effectiveness of the manipulation, but I think this is a slightly different point than the one already made in the discussion. That is, it relates to the change in the effectiveness of the manipulation rather than of the change in the perception of the dependent variable.**

We believe your point makes good sense and in a way resonates with the alternative explanation put forward by Dr. Vranka above: choosing an applicant "despite" their racial or gender identity could be considered moral, particularly given a (supposedly) more racist/sexist societal background back then, whereas choosing an applicant "because" of the identities may be viewed negatively by those who, e.g., disfavor affirmative actions. Your point that the manipulation loses effectiveness is also valuable. We incorporated both insights into the discussion section. Here's the relevant text:

> "Assuming our manipulation was ineffective, we are faced with two possibilities: either the manipulation was similarly ineffective in the original study, and as such, the original results were likely false positives, or its effectiveness has diminished over time. In other words, it could be the case that the manipulation just cannot provide moral credentials for those in experimental conditions, or it could be that the manipulation provided moral credentials before but not in our study. Regarding the latter possibility, another Stage 2 reviewer raised a point—echoing the insight of the aforementioned reviewer—that is worthy of further exploration: participants in the original study could have felt that they were choosing the star applicant despite ethnic or gender identities, whereas in comparison, those in the current study might have felt that they were choosing the star because of these identities. This change likely mirrors broader social, cultural, and institutional changes within U.S. society. Nonetheless, while selecting candidates without bias against historically marginalized identities may generally be viewed as morally commendable, actively favoring candidates based on these identities can spark controversy, as seen in recent backlashes against Affirmative Action programs (Liptak, 2023). Consequently, in our study, participants—especially those who went through the non-sexist credential manipulation—might have felt choosing a candidate for a disadvantaged identity could be seen as contentious even when the choice had little to do with identity. They were thus prompted to explicitly affirm gender or ethnicity equality in later scenarios. It is important to note that this speculation can be true despite that we observed little difference between participants' evaluations of selecting the Black star candidate and of selecting the female star candidate. Participants can hold certain

personal norms and at the same time worry whether the personal norms diverge from social norms (Bicchieri et al., 2014)."