Meta-Psychology, 2024, vol 8, MP.2024.4230
https://doi.org/10.15626/MP.2024.4230
Article type: Replication Article

Open data: Yes
Open materials: Yes
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: Yes

Edited by: Erin M Buchanan
Reviewed by: Savannah Lewis, Tom Rosman
Analysis reproduced by: Lucija Batinović
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/H57N4

Facets of Trust in Science: Researchers can be perceived as ethical and competent despite inconsistent research results

Hilmar Brohmer[1], Simon M. Ceh[1], Marta Czerwonka[2], Sandra Grinschgl[3], Gabriela Hofer[1], Izabela Lebuda[2], and Hildrun Walter[4]
[1]Department of Psychology, University of Graz, AT
[2]Institute of Psychology, University of Wrocław, PL
[3]Department of Psychology, University of Bern, CHE
[4]Center for Society, Science and Communication, University of Graz, AT

The public perception of science and scientists themselves has become a much-debated topic in recent years. In this article, we contribute to a more nuanced understanding of the public's trust in science by focusing on the practices of science, which are often not known by the public. Building on previous research by Ebersole, Axt and Nosek (2016), we conducted a preregistered, quota-sampled survey in Austria (N = 564), where we presented participants with different scenarios about scientific practices. Thereby, we disentangled the perception of scientists–i.e., how competent and ethical they are being perceived–from the confidence in their scientific findings–i.e., how correct their results are being perceived. For instance, when "a researcher X conducted a study with an interesting finding, which he then publishes", this researcher was–in our study– perceived as averagely competent and ethical, and the findings were perceived as neither correct nor incorrect (but somewhere in between). However, if another "researcher Y tried to replicate X's finding, but failed - and X then criticized Y's methodology and dismissed the new study", researcher X was perceived as less competent, less ethical and the original results were perceived as less correct by participants. Importantly, if researcher X "acknowledged Y's methodology" or "investigated the difference between the original study and the failed replication", ratings for X's competence and ethical behavior were higher than for how correct his original results were being perceived. Moreover, the highest competence and ethics ratings were obtained, when researcher X was described to share the methods and data online for transparency. Psychological dispositions of the participants, such as political orientation or motivation for cognition, did not seem to affect these ratings to a large degree. These results are discussed in the light of Mertonian norms of science, which highlight cooperativeness and disinterestedness.

Keywords: trust in science, perception of scientists, open science, constructive replication

Introduction

The public perception of science and scientists is an important topic as science is the key domain of societal progress (Besley et al., 2021; Pinker, 2018). Although there is a strong public support for scientific endeavors in some countries (e.g., (European Commission, 2021a), other regions and countries express lower confidence in science (Rabesandratana, 2019). While this lack of confidence seems to be more pronounced in low-income countries, some variation can be observed across high-income countries as well. For instance, fairly strong parts of the public in Germany and Austria express skepticism towards scientists and lower levels of interest in science compared to other European countries (Cologna et al., 2024; Eberl et al., 2021, European Commission, 2021b; ÖAW, 2022; Wissenschaft im Dialog, 2023). Of particular concern is that over 50 percent of Austrians do not explicitly support basic research and endorse the statement that knowledge about science and research is not relevant in their daily lives (European Commission, 2021b). Such attitudes are alarming because scientific findings have a huge impact on the daily lives of people: Not only did recent advances in research (e.g., developments in artificial intelligence) lead to a digital revolution in all societal areas (see Lee et al., 2023), but science also helped tackle the Covid-19 pandemic in record time (Haseltine, 2021). In the case of the pandemic, it was even shown that science-skeptical people, who often did not comply with scientifically informed measures like physical distancing, suffered more health-related consequences (e.g., Brzezinski et al., 2021).

But how can the public perception of science, scientists, and scientific results be improved? In this constructive replication study (Hüffmeier et al., 2016) we want to build on top of previous work by Ebersole, Axt and Nosek (2016). They showed that the perception of science and scientists depends on internal practices of "how science is done", which may have to be taken into account when communicating science. Additionally, we want to account for some psychological dispositions that may explain differences between people.

The Role of Scientific Practices in the Perception of Science

A crucial leverage point for improving the public's confidence in science would be to take scientific practices into account. Those practices can be regarded as the reflection of internal norms by how researchers formulate research questions, conduct studies, analyze results, share those results with the community, and react to other research findings. Most ideally for the scientific endeavor, the underlying norms – as they were formulated by Merton (1942; Anderson et al., 2007) – would imply that scientists i) conduct research objectively and for the sake of knowledge gain (rather than personal gain) and ii) be equally critical towards others and their own work (rather than protecting their pet theories). This may imply that researchers would appreciate replication studies by others to see if previous results hold true in light of new evidence. They should further iii) assess others' and their work based on the merit of the contributions and iv) the quality of research (rather than others' status and the quantity of their publications), and v) openly and cooperatively share their findings and knowledge with others (rather than keeping them hidden for patents and publications)[1].

In reality, there is often a gap between these norms and actual research practices (Anderson et al., 2007; Macfarlane and Cheng, 2008): Some researchers may conduct research to confirm their ideas only; they have pet theories, which they protect when others want to replicate their research in new studies to gain additional evidence; they aim to publish in "high-impact journals" to enhance their reputation; and they often do not share their data and materials (see also (Houtkoop et al., 2018). Some researchers may also feel pressured to engage in such practices because of the incentive structure (see Chambers, 2017) and publish-or-perish culture (e.g., van Dalen, 2021; van Dijk et al., 2014) in academia. In consequence, scientific progress may be hampered by such counter-normative practices.

The exact practices scientists employ during their research are mostly unknown by the public. However, Ebersole and colleagues (2016) could demonstrate that the communication of these practices, when they actually followed the Mertonian norms, improved the reputation of science and scientists: In a US-wide survey, the authors described the behavior of hypothetical researchers in several scenarios and asked participants how they perceived these researchers with regard to three dimensions: i) competence (or ability), ii) ethical behavior, and iii) the correctness of their results. The former two dimensions fit conceptually to the framework of epistemic trust in experts (Hendriks et al., 2015), which differentiates expertise (reflecting competence) from the often highly correlated integrity and benevolence (reflecting ethical behavior). In general, these dimensions seem to be relevant to understanding the reputation of experts.

Ebersole and colleagues' (2016) baseline scenario described a researcher publishing an interesting finding in a scientific journal. The authors found that participants perceived this researcher as averagely competent and their behavior as averagely ethical (i.e., neither incompetent nor highly competent, and neither unethical nor highly ethical) and their results as neither completely correct nor completely incorrect. However, in the second scenario, when the researcher's study was successfully replicated by another researcher, ratings for the first researcher became more positive on all three dimensions. Likewise, when the findings could not be replicated (as described in the third scenario), the ratings became more negative. Importantly, however, when such an unsuccessful replication attempt led the first researcher to be interested in investigating the differences between their own study and the replication attempt, their previous results were seen as less likely to be true – while the researchers themselves were seen as more competent and ethical compared to the baseline scenario. A similar effect occurred when the first researcher was agreeing with the methods of the replication study. Both behaviors may have signaled a cooperative mindset: Rather than dismissing the new results and protecting their own original results, the researcher showed an interest in researching the phenomenon, which is in line with the Mertonian norms (see Anderson et al., 2007).

In summary, the study by Ebersole et al. (2016)

[1] As a final norm–which is less of interest in the context of our research–they should autonomously decide on the direction of their research (rather than being directed by higher administrative management).

shows that peoples' perception of research results can be disentangled from their perception of the researcher. However, this study is yet to be replicated in a population outside of the US. We intended to achieve this in Austria in a well-powered and diverse sample, while also exploring additional aspects of Mertonian norms that coincide with Open-Science practices, as well as additional psychological factors of the participants that may be associated with confidence in science.

The Extension to Open Science Practices

Another crucial leverage point, which has come into research focus recently, is a more explicit communication of Open Science practices to the public. These often encompass a broad range of topics (e.g., including citizen science and diversity, see UNESCO, 2021). But in terms of the research process itself, Open-Science practices mostly focus on transparency (e.g., via open data and materials), analytical reproducibility, and replicability in follow-up studies with the same design (Crüwell et al., 2019; Robson et al., 2021), connecting them to the Mertonian norms.

There is evidence that when researchers adhere to these practices and make their data and methods openly accessible, public trust in science may benefit: In a large US panel study (N = 4464) conducted by the Pew Research Center (Funk et al., 2019) a year before the Covid-19 pandemic, an absolute majority (57 percent) indicated that their trust in scientific findings would increase when scientists made their data openly available. In a German survey (Rosman et al., 2022, Study 1), which was conducted during Covid-19, similar findings emerged: 74 percent from the total sample of N = 504 agreed with the statement that they would trust a study more if the scientists made data and methods openly available.

Notably, few experimental studies have yielded mixed results: First, in Song et al. (2022), participants with an academic background read about scientific findings in short summaries (i.e., as if they were reading a popular-scientific article) and expressed higher trust in the findings when Open-Science practices were highlighted (relative to Open-Science practices not being mentioned). Second, Rosman et al. (2022, Study 2), focusing on the lay population, did not observe beneficial effects of Open-Science practices on trust in the main confirmatory part of their analyses (although they did find some support in their exploratory analysis). Third, using an alternative design with Open-Science badges as markers of practices in published research, Schneider et al. (2022) found some support that such markers have a positive effect for teachers and scientists, but less so among the general public. However, as all three studies also explored different moderator variables, follow-up experiments with a more explicit focus on Open Science may be required (see Hofer et al., 2023).

Participants' Psychological Dispositions and Attitudes

Whether or not scientists are being trusted does not only depend on how and what is communicated, but may also differ between perceivers. From the literature, there are several factors that often show associations with trust in science or science skepticism (e.g., political orientation, belief in conspiracy theories or science literacy). However, it is not known how these psychological dispositions and attitudes of the perceiving person relate to effects of the scientific practices on trust. Thus, we aimed at testing a set of promising factors in an exploratory fashion.

First, participants' political orientation and ideology is consistently related to trust in science, where politically left-oriented and progressive people usually show higher confidence in science than people on the more conservative right site of the political spectrum across countries (e.g., Funk et al., 2020; McCright et al., 2013). This is also true for more specific science-related topics, such as belief in the climate crises or vaccination intention (e.g., Brohmer and Walcher, 2024; McCright et al., 2016). Second, peoples' belief in conspiracy theories seems to play a role for trusting science: Potentially because scientists are often believed to be part of the conspiratorial acts (Brohmer and Walcher, 2024; Brotherton et al., 2013), higher scores on conspiracy scales go hand in hand with higher skepticism towards and misconceptions of science (Spälti et al., 2023; Vranic et al., 2022). Third, we investigated participants' own preference for analytic thinking and motivation for cognition. This preference is often expressed in a motivation to solve complex problems in everyday life (similar to need for cognition, see Blaise et al., 2021). Higher preferences predict a more nuanced understanding of science and its processes (e.g., Čavojová et al., 2023; Feist, 2012), which is not surprising, given that knowledgeable people are also more likely to engage in scientific activities and career paths (e.g., Blotnicky et al., 2018). However, little is known about how people with higher preferences perceive scientists with regard to their ethical behavior, competence and their scientific work. Finally, similar to large-scale surveys (e.g., European Commission, 2021b), we included measures whether participants have a broad and general interest in science (i.e., in their leisure time).

The Present Study

Taken together, in this study, we aimed to assess how research practices affect trust in science and scientists. To achieve this, we followed the design and procedure of the main part of Ebersole and colleagues' original work (2016) in a constructive replication and extension study (Hüffmeier et al., 2016; Nosek and Errington, 2020). The main goal of our study was to see if the findings generalize to Austrians, who have shown comparatively high science skepticism before (European Commission, 2021a). However, what makes this study an incremental constructive replication (Köhler and Cortina, 2023) is that we integrated some minor alterations to extend its explanatory power: First, we made minor changes in the wording and scales, which made them more straight-forward (see Procedure and instrument section and Table S8 for details). Second, we formulated some additional scenarios, which fit the original setup and highlight Open-Science practices. Third, as most scenarios focus on the importance of replication studies (see below), we wanted to see if reflecting on these scenarios affected the participants' perception of replication studies. In line with the findings of Ebersole and colleagues (2016), we formulated the following preregistered hypotheses (see https://osf.io/cba4s/). These hypotheses refer to scenarios, which we presented to the participants. Moreover, ratings of each scenario were compared to ratings of a baseline scenario (sc01) in a within-subjects design, where a hypothetical *researcher X publishes an interesting finding*.

- H1: Following a successful replication by researcher Y (sc02), researcher X will be perceived as a) more competent, b) more ethical and his original results will be perceived as c) more correct.
- H2: Following a non-successful replication by researcher Y (sc03), researcher X will be perceived as a) less competent and his original results will be perceived as c) less correct. We expect b) an effect close to (i.e., equivalent to) zero for how ethical X is perceived.
- H3: Following a non-successful replication by researcher Y, after which X criticized Y's methodology and dismissed the new results (sc04), researcher X will be perceived as a) less competent, b) less ethical and his original results will be perceived as c) less correct.
- H4: Following a non-successful replication by researcher Y, after which X agreed with Y's methodology and said that the original results might not be correct (sc05), researcher X will be perceived as b) more ethical and his original results will be perceived as c) less correct. We expect a) an effect equivalent to zero for how competent X is perceived.
- H5: Following a non-successful replication by researcher Y, after which X started a new study to investigate why there were two different results (sc06), researcher X will be perceived as a) more competent and b) more ethical. We expect c) an effect equivalent to zero for how correct his original results are perceived.
- H6: Following the scenario that X publishes a failed self-replication challenging the original results (sc07), researcher X will be perceived as a) more competent and b) more ethical and his original results will be perceived as c) less correct.
- H7: Following the scenario that after failed self-replication, researcher X decided that the new results are not valid and did not publish them (sc08), researcher X will be perceived as a) less competent, b) less ethical and his original results will be perceived as c) less correct.
- H8: Following the scenario that researcher X did not follow-up on the results and moved on to investigate other things (sc09), researcher X will be perceived as a) less competent, b) less ethical and his original results will be perceived as c) less correct.

As some of the scenarios strongly emphasized the importance of replication studies, we thought that this will affect participants' perception of replication studies in more general terms as well. Similar to other experiments (e.g., Ranney and Clark, 2016), when exposed to science-backed information on a particular topic, participants might then indicate a different attitude towards this topic. We therefore formulated the following hypothesis:

- H9: Participants, who read the scenarios first, will later indicate that replication studies are more important to science compared to participants, who read the scenarios later (and give their rating for the importance of replications first).

In the exploratory part of our study, we wanted to achieve two things. First, we aimed to extend the previous findings by Ebersole and colleagues (2016) by adding three new scenarios, which we also tested against the baseline scenario. Those scenarios address further aspects of scientific practices and Open Science:

Scenario 10 (sc10) describes researcher X, who did not follow-up on previous results, but instead conducts new studies that built on top of non-replicated previous findings. Hence, this scenario addresses the issue that researchers may assume that one-time findings are already sufficient for knowledge gain. Scenario 11 (sc11) stated that after a failed and non-published first study, researcher X succeeds in a second study and publishes it. This is the reversed case of scenario 8 and likewise addresses the file-drawer problem. Finally, in scenario 12, researcher X shares his data and materials for transparency after publishing an interesting finding (sc12).

Second, we wanted to explore whether the perception of researchers following the scenarios may be partly correlated with psychological dispositions and attitudes of participants, which have been shown to be associated with science skepticism or trust in science before (see above; Blaise et al., 2021; Brohmer and Walcher, 2024; Funk et al., 2020). These include 1) belief in conspiracy theories, 2) left-right political orientation, 3) motivation to engage in cognition and analytical thinking, and 4) participants' leisure interest in science.

Materials and Methods

Power Calculation and Sample Description

We planned to recruit a quota-representative sample via a panel provider, consisting of 500 to 600 participants. A sample size within this range would have enabled us to find effects similar to the original study (see https://osf.io/ex36u). We also decided to use equivalence testing (Lakens, 2017) as some effects were expected to be close to zero. Hence, we set reasonably small equivalence bounds for the effect sizes. Those bounds function as thresholds: effects within these bounds (i.e., they do not pass the threshold and are close to zero) may be interpreted as evidence for a null effect, whereas effects above (or below) these bounds are evidence in favor of an effect. Specifically, as we expected a sample size of more than 500 participants, this would be sufficient for equivalence test bounds of $\Delta d = \pm 0.20$ for paired-samples (or one-sample) t-tests, which often count as small effects. These bounds would imply statistical power of 99.53% ($\alpha = .05$, $N = 500$) to 99.89% ($\alpha = .05$, $N = 600$; based on a power analysis in the TOSTER package, Lakens, 2017). For the independent-samples t-test of Hypothesis 9, we had to set a slightly higher bound of $\Delta d = \pm 0.26$, which was still sufficiently powered (power = 80%).[2] Statistically significant effects ($p < .05$) were only counted as evidence when their point estimate was lying above this threshold (for similar approaches see Brohmer et al., 2023; Hofer et al., 2022).

Our panel provider Talk Online Panel (https://talkonlinepanel.com/at) distributed the survey between February 21 and 28, 2024, to $N_{\text{raw}} = 1158$ Austrian participants. Of those, $n = 818$ reached the attention check (see below), which was failed by $n = 218$. Finally, $n = 36$ participants indicated at the end of the survey that they got distracted by their surroundings "quite a bit" or "all the time". After excluding these participants in line with our preregistration, we were left with a final sample of $N_{\text{final}} = 564$, which was in line with our preregistration. This sample consisted of 48.76% men ($n = 275$), 50.89% women ($n = 287$), and 0.35% gender-diverse or non-binary people ($n = 2$) and had an average age of 50.54 years ($SD = 16.67$). As set in the quotas, most participants were of Austrian nationality (94.86%, $n = 535$), had no high school diploma (70.74%, $n = 399$) and lived on the countryside with less than 10,000 inhabitants in the municipal (52.93%, $n = 298$). Except for the mean age, which was a bit higher than expected, all parameters approached the representative target (see Table S7). As an optional measure, most participants indicated their net income group, which was above € 2000 per month for the majority (52%, $n = 293$).

Procedure and Instruments

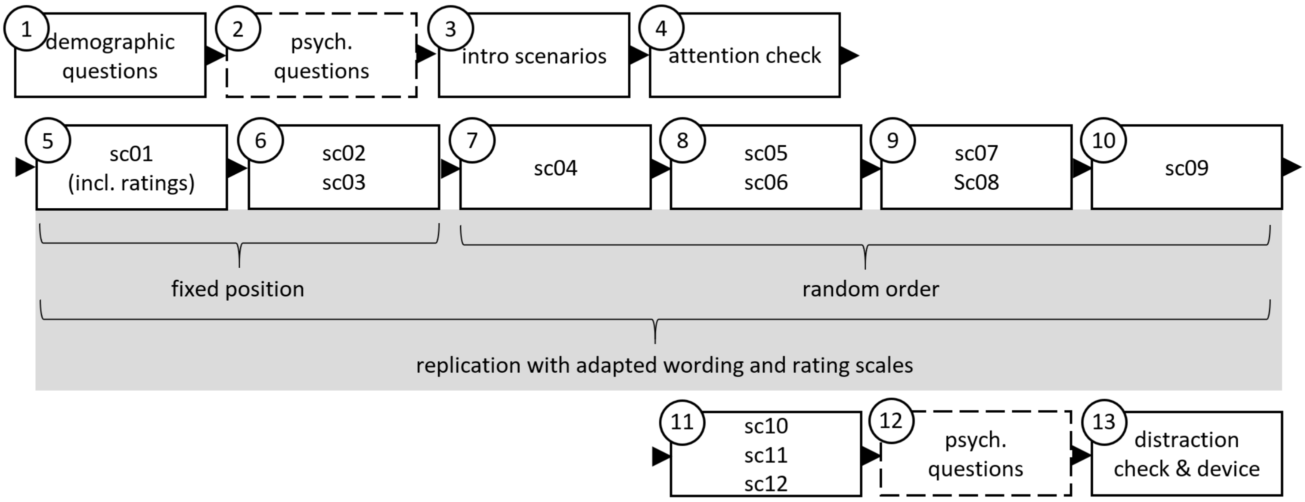
We set up a survey in German using the online software LimeSurvey (2023) and followed the original survey closely. Participants gave their informed consent and answered some demographic questions (see above).

Next, the psychological disposition variables were randomly either shown before or after the scenarios describing behaviors of researchers (to check for order effects, which were largely not present, see results section) and had to be answered on seven-point rating scales (1 = "do not agree at all" / "not at all" to 7 = "fully agree" / "very much"; find the list of variables here: https://osf.io/537cj). We presented participants with two items on their motivation for cognition and analytic thinking (cog1 and cog2; example item: "In general, I enjoy thinking about a problem, regardless of whether I can solve it."; $r = .55$; adapted from Blaise et al., 2021) and two items on their generic conspiracy belief (consp1 and consp2; example item: "Certain significant events have been the result of the activity of a small group who secretly manipulate world events."; $r = .50$,

[2] In the preregistration, we incorrectly indicated the wrong bound of $\Delta d = \pm 0.223$, which does not correspond to our expected minimum sample of $N = 500$ (or $n = 250$ per group). The following command in the TOSTER package will yield the correct bound of $\Delta d = \pm 0.223$: "powerTOSTtwo ($\alpha = .05$, statistical_power = .80, $N = 250$)"

Figure 1

Set up of the Survey. Note: page 2 and 12 were shown in random order (between-subjects condition); page 11 represents the extension of the scenarios on a fixed position; scenario descriptions (sc01-12) are provided Figure 2 and 3.



Brotherton et al., 2013; Gemenis and Littvay, 2023). As a proxy for participants' contact points with science in their daily lives (adapted from European Commission, 2021b), we asked three questions on whether participants enjoy to engage in technical topics in their leisure time (sci1), whether they are staying informed about science in their leisure time (sci2), and whether they think scientific knowledge is not important in their daily lives (know; reversed scaled). Relevant to testing H9, we included two questions on how participants perceive replication studies, where the first question was of interest (rep_imp): "To what extent would you say that scientific replication studies are important for scientific progress?". The second question (rep_cre) read "To what extent would you say that replication studies are creative scientific work?", which we did not have particular expectations for.

Next, we contextualized the scenarios, where we described participants in three short paragraphs the typical work of scientists (i.e., they test their ideas in studies, publish results in scientific journals, and occasionally replicate studies), which were followed by an attention-check question: "Which of the following four points is NOT part of the working methods of science?" (correct answer: "Researchers can publish their research in scientific journals without peer review anytime."). In the next part, they read German-translated versions of the original scenarios. Upon initial correspondence with the original authors, we implemented some minor changes, which would improve the clarity of the study.[3] A crucial change that we implemented across all scenarios was that we gave the hypothetical researcher "X" and "Y" the common Austrian family names "Huber" and "Weber", respectively, to enhance comprehensibility. For the remaining in this article, we will stick to "X" and "Y".

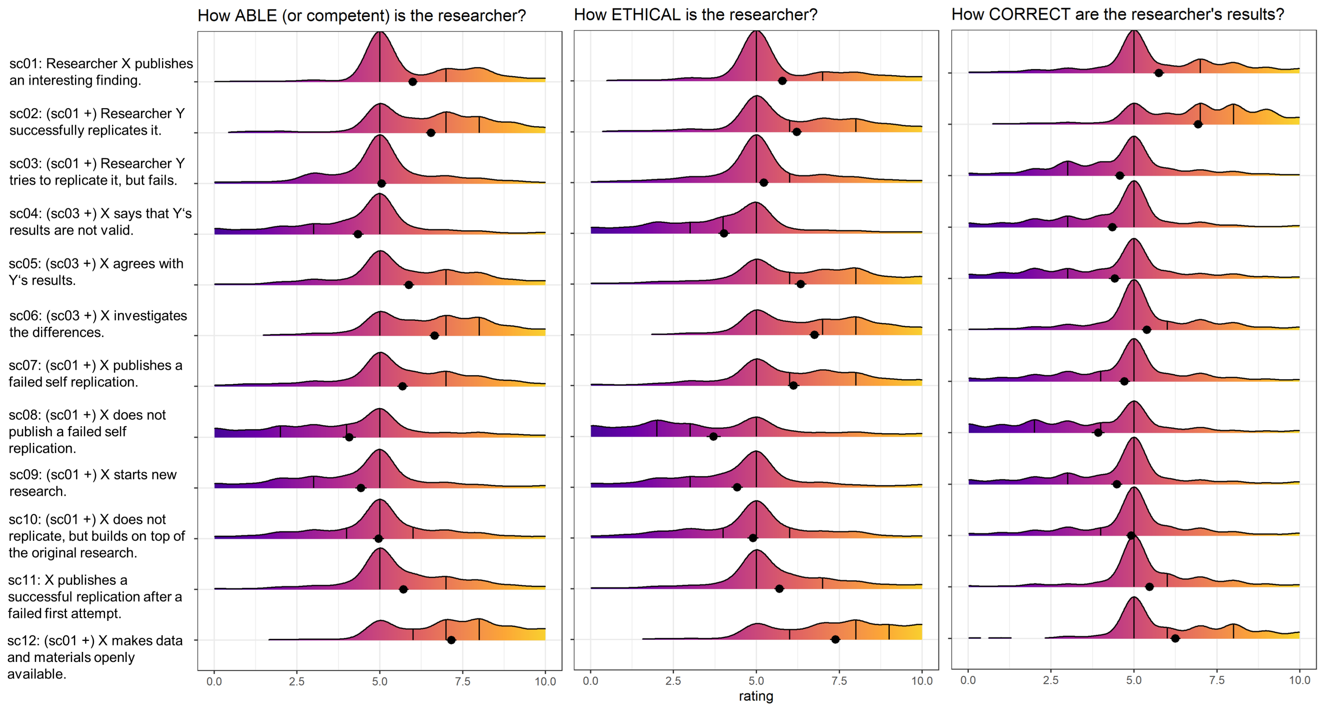
The baseline scenario (sc01), where a hypothetical researcher X published an interesting result, was always presented on page 1, and we tested the other scenarios against this baseline in a within-subjects approach. Hence, sc01 was followed by sc02 and sc03, which introduced the idea of successful and failed replication studies. The following six scenarios (sc04 to sc09) highlighted different behaviors of the researchers and were grouped according to common themes and presented in a random order. For instance, in scenario 4, it was described that researcher X published an interesting result (as in scenario 1), but researcher Y did not find the same result in a replication study. In consequence, X criticized Y's methods and dismissed the new results. The final three scenarios (sc10 to sc12) were not part of the original study and, hence, were always shown at the end (for the complete design, see Figure 1).

Each scenario was accompanied by three questions

[3] Those minor changes were that 1) our revised formulation of the scenarios refers more consistently to "replication studies", whereas in the original study the formulation of "reproducibility" may have been more ambiguous; 2) we used rating scales that consistently used 11 points and were horizontally presented rather than in a drop-down menu (see original setup here: https://osf.io/zmxcy/; see all differences in Table S8).

Figure 2

Descriptive results for the rating domains. Note: ridge plots with density curves show distributions of answers; vertical lines in the plot depict the interquartile distance with the median in the middle; point estimates depict the mean with the 95% confidence interval; see also Table S1.



on how participants perceived researcher X with regard to their competence ("How able or competent is Researcher X?"), ethical behavior ("How ethically does Researcher X behave?") and the correctness of X's original results ("How likely is it that the result found by Researcher X is indeed correct?"), which were single-item measures, rated on eleven-point scales (from 0 to 10; competence: 0 = "one of the worst researchers" to 10 = "one of the best researchers"; ethics: 0 = "one of the most unethical researchers" to 10 = "one of the most ethical researchers"; correctness: 0 = "definitely not correct" to 10 = "definitely correct").

Preprocessing and Data Analysis

The data was preprocessed and analyzed in R (R Core Team, 2020; find a list of all packages used here: https://osf.io/wjv9a).

We first applied our preregistered exclusion criteria (see sample description above), followed by analyzing the demographic information. To test Hypotheses 1 to 8, we calculated Cohen's d and its 95% confidence interval (CI) for each scenario compared to the baseline scenario and conducted paired-samples t-tests. For the exploratory analysis, we calculated composite scores for highly correlated scale items ($r > .41$, Lovakov and Agadullina, 2021). Then we split the sample into a training and test dataset (appr. 60% and 40%). We only interpreted effects as evidence if we found them in the training set and test set.[4] We conducted correlations and independent-samples t-tests for these additional analyses.

[4]We did not register the exact split beforehand, but utilizing a larger training and smaller test set is generally recommended (Dobbin and Simon, 2011): one can identify potential effects in the larger training set, which then need to hold in the smaller test set. This functions as a robustness check as small samples are usually more prone to false-negative errors in general (Button et al., 2013). Note that such approaches are usually used for exploratory machine learning with much larger samples (e.g., Szabelska et al., 2021). Given that effect sizes stabilize when samples reach N = 200 to 250 (Schönbrodt and Perugini, 2013), we reasoned that our test set would still be sufficiently large to detect effects when we use a 60:40 split.

Results

An initial overview of the scenario ratings is provided in Figure 2. The left panel shows the rating for how able (or competent) researcher X is being perceived, the middle panel shows how ethical they are being perceived, and the right panel depicts how correct their results are being perceived. The density distributions suggest that most participants picked the midpoint (i.e., average competence and ethical behavior, and equally likely correct or incorrect results). However, two scenarios–sc02 and sc12–showed consistently higher means and medians across all three domains, whereas other scenarios–particularly sc04 and sc08–showed somewhat lower ratings than the midpoint. Moreover, in three scenarios–sc05 to sc07–the perceived competence and ethical behavior of the researcher seemed to be detached from the perceived correctness of the results–at least from a descriptive perspective.

Confirmatory Analyses

To test Hypotheses 1 to 8, we conducted paired-samples *t*-tests with sc01 as baseline scenario and equivalence bounds at $\Delta d = \pm 0.20$. As shown in Figure 3, all but two scenarios resulted in full support of the respective hypotheses, whereas the remaining two yielded partial support (see also Table 1 and S1 for detailed statistical information, https://osf.io/s9j4d):

- H1: a successful replication attempt by Y led to higher competence, ethics and correctness ratings for X (sc02: $0.29 \leq ds \leq 0.65$, all $ps < .001$, full support).
- H2: a non-successful replication attempt led to lower competence and correctness ratings, but also to lower (instead of null equivalent) ethics ratings for X (sc03: $-0.53 \leq ds \leq -0.31$, all $ps < .001$, partial support).
- H3: Critique of researcher X about Y's methods after Y's failed replication led to consistently lower ratings for X (sc04: $0.59 \leq ds \leq 0.70$, all $ps < .001$, full support)
- H4: Agreement of researcher X with Y's methods after Y's failed replication (sc05) led to a null-equivalent competence rating ($d = -0.05$, $p = .205$), a higher ethics rating ($d = 0.23$, $p < .001$), and a lower correctness rating ($d = -0.51$, $p < .001$, full support)
- H5: Researcher X's attempt to study the differences between his and Y's results after Y's failed replication (sc06) led to higher competence ($d = 0.33$, $p < .001$) and ethics ratings ($d = 0.47$, $p < .001$) and a null-equivalent correctness rating ($d = -0.17$, $p < .001$, full support)
- H6: Researcher X's publication of a failed self-replication (sc07) led to a lower correctness rating ($d = -0.42$, $p < .001$), but also to null-equivalent (instead of positive) ratings for competence ($d = -0.14$, $p = .001$) and ethics ($d = -0.14$, $p < .001$, partial support).
- H7: A failed self-replication that is dismissed by X led to consistently lower ratings for X (sc08: $-0.73 \leq ds \leq -0.66$, all $ps < .001$, full support)
- H8: Not following up on previous results, but investigating new things led to consistently lower ratings for X (sc09: $-0.67 \leq \text{ds} \leq -0.55$, all $ps < .001$, full support)

Overall, our results mostly replicated the original findings by Ebersole et al. (2016). Particularly, we showed again that researchers can be perceived as ethical and competent, even if their findings are not perceived as correct, which can be seen for H4 and H5. This disentanglement effect was also present for H6, although not as pronounced as in the original study. Notably, for the partially supported H2, negative effects became more pronounced as compared to the original study.

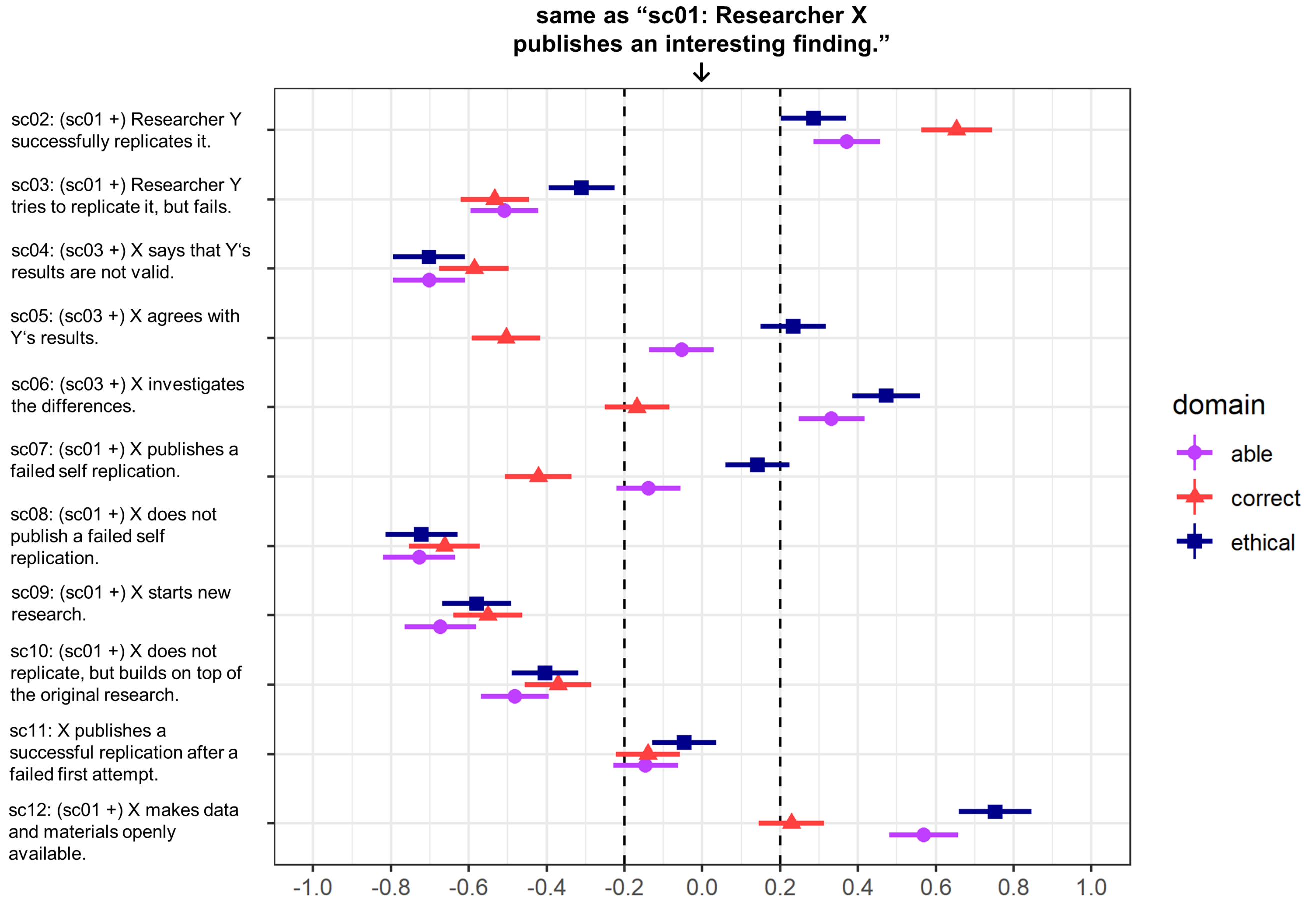
Finally, we tested H9, which stated that the order of the scenarios affected the perceived importance of replication studies. Indeed, participants who read the scenarios first ($M = 6.16$, $SD = 1.15$, $n = 278$), saw higher importance in replication studies than participants who answered this question before reading the scenarios ($M = 5.60$, $SD = 1.36$, $n = 286$; $b = 0.55$, $SE = 0.11$, $t(562) = 5.22$, $p < .001$, $d = 0.44$, 95%CI [0.27, 0.61]; see Figure S1). The effect size's point estimate was above the equivalence bound of d = 0.26. Additionally, the order of the variables did not affect to what degree participants perceived replication studies as creative work (and the order did not affect any other variable substantially, either, see Table S2 and S6).

Exploratory Analyses

As the final three scenarios (sc10-12, see also Figure 3) were newly added to this study, we tested them in our preregistered exploratory split-sample approach against the baseline scenario. A random split of the data was executed and we saved the seed for the purpose of reproducibility (see the code here: https://osf.io/gxmyb). This yielded a training set consisting of $N = 342$ cases (61%) and a test set consisting of $N = 222$ cases (39%).

Figure 3

Main results. Note: sc02 to sc09 are confirmatory; sc10 to sc12 are exploratory; point estimates per domain contain a 95% CI; results are in Cohen's d metric; see original effects here: https://osf.io/ex36u.



In scenario 10, participants read about researcher X conducting studies that build on top of the original, but not replicated, results. This led to consistently more negative ratings for competence, ethics and correctness of the results in the training set, which could then be confirmed in the test set ($-0.62 \leq ds \leq -0.31$, $ps < .001$). In scenario 11, participants read about X, who first finds no interesting results, but then publishes a successful replication. This led to consistent null-equivalent ratings in the training set ($-0.09 \leq ds \leq -0.03$, $p > .104$), but they could not be confirmed in the test set, where ratings were null-equivalent for ethics ($d = -0.07$, $p = .302$), but more negative for competence ($d = -0.25$, $p < .001$) and correctness ($d = -0.23$, $p = .001$). Finally, in scenario 12, participants read about researcher X, who openly shares his original data and materials, which led to consistently higher ratings in both the training and test set ($0.21 \leq ds \leq 0.77$, $ps > .002$; see Table S4 for detailed statistics).

Next, we explored the additional psychological dispositions as preregistered. We provide descriptive statistics and correlations of the training and test sets in Table 2 and only describe statistically significant ($p < .05$) correlations that occurred in both of the sets: Political orientation (higher values = more right-wing) correlated positively with belief in conspiracies and negatively with leisure interest in scientific topics (sci2). Motivation for cognition correlated positively with leisure interest in scientific topics and leisure interest in technical ideas (sci1), and negatively with indifference about science in daily life (know). The latter also correlated negatively with both leisure interest variables and positively with conspiracy beliefs.

We then tested associations with the baseline ratings, as well as with the most positively and most negatively rated scenarios. In line with our main results, the most consistently positive ratings were obtained for sc02 and sc12 and the most consistently negative ratings were

Table 1

Hypothesis tests for all dimensions. Note: a visualization can be found in Figure 3 and descriptive statistics can be found in Table S1. A reviewer suggested that a Bonferroni α-error correction to adjust for multiple testing; applying this correction (α = .05 / 24 = .002) does not affect the conclusions as all previously significant p-values remain statistically significant after the correction.

Dimension	vs. sc01	t	df	p	b	SE	Cohen's d
able	sc02 able	8.811	563	<.001	0.546	0.062	0.371
	sc03 able	-12.079	563	<.001	-0.94	0.078	-0.509
	sc04 able	-16.679	563	<.001	-1.654	0.099	-0.702
	sc05 able	-1.268	563	0.205	-0.119	0.094	-0.053
	sc06 able	7.89	563	<.001	0.654	0.083	0.332
	sc07 able	-3.284	563	0.001	-0.309	0.094	-0.138
	sc08 able	-17.274	563	<.001	-1.915	0.111	-0.727
	sc09 able	-15.997	563	<.001	-1.567	0.098	-0.674
ethical	sc02 ethical	6.774	562	<.001	0.43	0.063	0.285
	sc03 ethical	-7.371	563	<.001	-0.559	0.076	-0.31
	sc04 ethical	-16.687	563	<.001	-1.762	0.106	-0.703
	sc05 ethical	5.547	563	<.001	0.55	0.099	0.234
	sc06 ethical	11.222	563	<.001	0.966	0.086	0.473
	sc07 ethical	3.369	563	<.001	0.339	0.101	0.142
	sc08 ethical	-17.148	563	<.001	-2.074	0.121	-0.722
	sc09 ethical	-13.772	563	<.001	-1.363	0.099	-0.58
correct	sc02 correct	15.511	563	<.001	1.179	0.076	0.653
	sc03 correct	-12.666	563	<.001	-1.174	0.093	-0.533
	sc04 correct	-13.928	563	<.001	-1.402	0.101	-0.586
	sc05 correct	-11.973	562	<.001	-1.334	0.111	-0.505
	sc06 correct	-3.983	563	<.001	-0.365	0.092	-0.168
	sc07 correct	-9.994	561	<.001	-1.036	0.104	-0.422
	sc08 correct	-15.743	563	<.001	-1.83	0.116	-0.663
	sc09 correct	-13.107	563	<.001	-1.27	0.097	-0.552

obtained for sc04 and sc08. The results implied that for the baseline scenario sc01, only participants' motivation for cognition positively correlated with higher competence, ethics, and correctness ratings (training: $.126 < r < .163$, $p < .05$; test: $.171 < r < .234$, $p < .05$). This motivation for cognition effect was also present for the positively rated scenarios sc02 (training: $.113 < r < .150$, $p < .05$; test: $.188 < r < .326$, $p < .05$) and sc12 (training: $.198 < r < .249$, $p < .05$; test: $.144 < r < .172$, $p < .05$). No other variables showed correlations with the ratings in the test set following a significant correlation in the training set, neither for the positive nor for the negative scenarios (see Tables S5 and S6).

Discussion

The public perception and reputation of science and scientists partially depends on how research practices are communicated. Following the work of Ebersole and colleagues (2016), we aimed at replicating and extending their study in the Austrian population. Based on previous surveys, Austrians might be an interesting population as they often score somewhat lower in trust in science than other European countries (European Commission, 2021b). In our study, we presented participants with several scenarios, which described scientists behaving differently following the publication of research results and independent or self-replication attempts.

Similar to the original study, we could demonstrate that scientists are less favorably perceived with regard to their competence, ethical behavior, and the correctness of their results (compared to a baseline scenario), when they simply dismissed failed replication attempts, either their own, or their colleagues'. This was also the case if scientists did not follow up with replication studies, but rather conducted studies building on top of the previous findings or entirely new research. However, if scientists showed collaborative behavior when their results could not be replicated (i.e., by acknowledging and investigating the differences between their own study

Table 2

Descriptive statistics and Pearson correlations for psychological variables. Note: training data results are below the diagonal, test data results are above the diagonal of the correlation table; bold correlations ($r > \pm 0.14$) were statistically significantly different from zero in both the training and test set ($p < .05$); see Figure S2, https://osf.io/s9j4d; pol = political orientation, cog = motivation for cognition scale, consp = generic conspiracy belief scale, sci1 = leisure interest in technology and creation, sci2 = leisure interest in scientific topics, know = knowing about science has no relevance in daily life.

Variable	Training			Test			Correlations					
	N	M	SD	N	M	SD	pol	cog	consp	sci1	sci2	know
pol	331	4.99	1.84	218	4.74	1.92	—	-0.134	**0.244**	-0.067	**-0.297**	0.096
cog	342	5.17	1.19	222	5.25	1.27	0.02	—	0.004	**0.400**	**0.451**	**-0.244**
consp	342	3.67	1.44	222	3.51	1.50	**0.202**	-0.05	—	0.019	-0.101	**0.188**
sci1	342	3.96	1.78	222	3.86	1.86	0.041	**0.418**	-0.005	—	**0.377**	**-0.261**
sci2	342	4.28	1.61	222	4.39	1.75	**-0.141**	**0.449**	-0.074	**0.264**	—	**-0.308**
know	342	3.49	1.78	222	3.33	1.77	0.111	**-0.320**	**0.237**	**-0.230**	**-0.432**	—

and the replication), they were perceived as just as competent and ethical or even more competent and ethical compared to the baseline. Importantly, our work also shows that describing the purpose of replication studies to the participants increased their appreciation of such studies for scientific progress.

Besides those replicated effects, we observed some differences compared to Ebersole et al. (2016) in the data patterns: In our Austrian sample, a failed replication by another researcher (sc02) resulted in the first researcher being perceived as somewhat less ethical than in the original US sample. What is more, the negative competence and correctness ratings were even more pronounced in our study, implying that Austrians may be potentially more critical towards studies, when independent researchers fail to replicate them. Interestingly, this claim is also supported by the results for scenario 7: Contrary to the original study, the publication of a failed self-replication did not lead to notable (positive) changes in the perceived competence and ethical behavior ratings (but only lower correctness ratings). This may imply that the Austrian public perceives replication studies as relevant, but not as something particularly outstanding for the researcher, who conducted both studies. Instead, it might make the original researcher seem rather suspicious when it requires another researcher to find out that a study is not replicable.

Overall, the most relevant implication of Ebersole et al. (2016) could also be confirmed for the Austrian population: The reputation of researchers, with regard to their perceived ethical behavior and expertise, relies less on the perceived correctness of research results, but rather on how open and collaboratively the researchers deal with them, especially if replication study results do not align with previous findings. In this regard, these results corroborate that the public appreciates when scientists follow the Mertonian norms (Anderson et al., 2007; Merton, 1942; see also Philipp-Muller et al., 2022) in their behavior. Specifically, signaling a cooperative mindset and being self-critical (e.g., by acknowledging potential issues with one's own research, sc05) or being interested in knowledge gain through one's research (e.g., by investigating differences in studies, sc06) strengthened the perception of these researchers as being ethical. In the same vein, these findings are in line with the epistemic trust framework (Hendriks et al., 2015): researchers being perceived as more ethical may go hand in hand with higher perceived integrity and a feeling that they work for the benefits of the society (i.e., benevolence).

The message that the public appreciates researchers independent of their research results is crucial as inconsistent results in research belong to the everyday professional life of a researcher: It is normal and expected that studies do not always yield the hypothesized results, just as it is normal and expected that replication studies do not always replicate the original finding (be it due to differences in the research design, population, time points or simply measurement error). Such inconsistent results sometimes inform better methods and may contribute to theoretical progress in the long run (e.g., Chambers, 2017, Chapter 3). In other words, communicating the scientific practices behind research results can add credibility to the researchers, contrary to researchers prematurely dismissing failed independent replications or hiding failed studies in the file drawer, which is known to be detrimental to science (see Munafò et al., 2017; Rosenthal, 1979).

The main message of the exploratory results is that if scientists share data and materials openly–i.e., they practice Open Science–this has an overall positive effect on their ratings, especially on how their behavior is perceived from an ethical lens (also regarding Mertonian norms, see above). Those results are in line with similar previous findings (Funk et al., 2019; Song et al., 2022) and therefore underline the importance of Open Science practices, not only for the research process, but also for how science is being perceived (but see also Rosman et al., 2022).

Interestingly, those ratings were unrelated to psychological dispositions. Only participants with a higher motivation to engage in cognition and analytical thinking seemed to be somewhat more favorable of scenarios that could be replicated or that involved Open Science practices. In other words, all results seemed to be largely similar, independent of participants' individual political orientation, their inclination to believe in conspiracy theories, or their general science interest (or lack of) in their daily lives (with all caveats regarding purely observational data).

Limitations and Future Directions

This study did not come without limitations. First, as in the original study, all scenarios were preceded by an explanation of how science works. Together with the scenarios, this might have influenced the perception of the importance of replication studies as shown with the result for Hypothesis 9. Such detailed general explanations cannot precede actual science communication pieces (e.g., popular scientific articles, short-form videos with scientific contents) as this would likely decrease the interest in such formats. Rather, specifics about the practices, which signal Mertonian norms and Open Science may have to be directly integrated into the format in which science is communicated.

Second, as in the original study, the scenarios describe scientific practices, which are common for some of the empirical, mainly quantitative and natural sciences, but not necessarily for all scientific disciplines. For instance, practices in law or the humanities are more qualitatively oriented, where an in-depth understanding is more often focused on case studies (e.g., White and Cooper, 2022). Hence, it remains unknown whether these ratings generalize to all scientific disciplines. In the same vein, it is unknown whether more topic-related scenarios, which may spark societal debate (e.g., climate change or vaccination), would affect the ratings.

Third, this study did not involve an experimental between-subjects manipulation as all participants read all scenarios followed by the baseline scenario on the fixed first position. As previous experimental research with between-subjects conditions found mixed results on, for instance, the importance of communicating Open Science practices (Rosman et al., 2022; Song et al., 2022), we suggest more research in this direction to understand when and under which conditions communicating such practices is useful, which would also allow more causal conclusions.

Fourth, although we argue that the wording of our rating scales for competence, ethics and correctness capture relevant aspects of trust in science, more nuances could be helpful for a deeper understanding. For instance, following Hendriks et al. (2015), it would be theoretically possible to differentiate benevolence (i.e., whether a researcher is ethical in their behavior) from integrity (i.e., whether a researcher has sincere and just motives) in the behavior of scientists. Although we argue that knowledge about integrity may also require much more contextual background information than can be provided in scenario studies like ours.

Fifth, we adapted this study to Austria, which has a tradition of science skepticism among European countries (European Commission, 2021b). Still, in many other, primarily developing countries, science skepticism and ignorance are sometimes even more pronounced (Rabesandratana, 2019). Hence, to uncover and understand potential cultural differences, replication studies on other continents are needed (for a recent large-scale study on the matter see Cologna et al., 2024).

Finally and related to the previous limitation, we want to recognize that it is not just the communication of science that may require adjustment to uphold public trust in science. Public trust in science also depends on an effective science education–such as in schools or university hands-on labs for pupils–which starts in childhood (Krüger et al., 2022; Riccardi, 2023). If children and adolescents learn early about scientific methods (see also the first point above), they might be better at evaluating information on science. The factors and variables in this study could not address these more complex underlying conditions.

Conclusion

In this constructive replication, we show that researchers can, indeed, be perceived as ethical and competent by the public, even if their research results do not replicate in follow-up studies. However, this is only the case when they follow the Mertonian norms in a way that signals a cooperative mindset and an interest in knowledge gain. Importantly, when researchers apply Open Science practices, they are perceived as most ethical and competent. These findings highlight the importance of including information on scientific practices when communicating scientific findings to the public. By doing so, trust in science might be increased.

Author Contact

- Hilmar Brohmer, corresponding author, https://orcid.org/0000-0001-7763-4229, hilmar.brohmer@uni-graz.at
- Simon M. Ceh, https://orcid.org/0000-0003-0534-3697, simon.ceh@uni-graz.at
- Marta Czerwonka, https://orcid.org/0000-0001-8247-3749, marta.czerwonka@uwr.edu.pl
- Sandra Grinschgl, https://orcid.org/0000-0001-6666-9426, sandra.grinschgl@unibe.ch
- Gabriela Hofer, https://orcid.org/0000-0003-4407-1487, gabriela.hofer@uni-graz.at
- Izabela Lebuda, https://orcid.org/0000-0002-4715-1928; izabela.lebuda@uwr.edu.pl
- Hildrun Walter, https://orcid.org/0000-0001-8720-3668, hildrun.walter@uni-graz.at

Acknowledgements

The authors thank Brian Nosek for his feedback on the survey materials.

In memory of Monika Brohmer.

Conflict of Interest and Funding

The authors declare no conflict of interest. This study received funding by the Erzherzog Johann Zukunftsfonds (WKO [Economic Chamber Styria, Austria; awarded to HB]). Furthermore, GH was supported by the Post-Doc Track Programme of the OeAW (Austrian Academy of Sciences), and MC received funding through the National Science Center Poland, 2022/44/C/HS6/00134.

Author Contributions

HB conceptualized the study, preregistered the study, collected and analyzed the data, managed the data, and wrote the first draft to the manuscript. MC, IL, and HW gave comments on the preregistration and suggested changes to the survey and on the manuscript. SMC and SG commented on the materials and gave substantial feedback on the manuscript. GH gave substantial feedback on the manuscript and checked the analysis script.

Open Science Practices

- anonymized data and code: https://doi.org/10.17605/OSF.IO/UZTJ7
- study materials: https://doi.org/10.17605/OSF.IO/VWQBM;
- preregistration: https://doi.org/10.17605/OSF.IO/CBA4S

This article earned the Preregistration+, Open Data, Open Materials, and Open Code badge for preregistering the hypothesis and analysis before data collection, and for making the data, materials, and code openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References

Anderson, M. S., Martinson, B. C., & De Vries, R. (2007). Normative dissonance in science: Results from a national survey of us scientists. *Journal of Empirical Research on Human Research Ethics*, *2*(4), 3–14. https://doi.org/10.1525/jer.2007.2.4.3

Besley, J. C., Lee, N. M., & Pressgrove, G. (2021). Reassessing the variables used to measure public perceptions of scientists. *Science Communication*, *43*(1), 3–32. https://doi.org/10.1177/1075547020949547

Blaise, M., Marksteiner, T., Krispenz, A., & Bertrams, A. (2021). Measuring motivation for cognitive effort as state. *Frontiers in Psychology*, *12*, 785094. https://doi.org/10.3389/fpsyg.2021.785094

Blotnicky, K. A., Franz-Odendaal, T., French, F., & Joy, P. (2018). A study of the correlation between stem career knowledge, mathematics self-efficacy, career interests, and career activities on the likelihood of pursuing a stem career among middle school students. *International Journal of STEM Education*, *5*, 1–15. https://doi.org/10.1186/s40594-018-0118-3

Brohmer, H., Köstenbaumer, D., Thaller, A. E., Fleiß, E., Athenstaedt, U., & Posch, A. (2023). Changing behavior while having climate change in mind? an investigation of social-psychological predictors for specific pro-environmental behaviors in the energy domain. *Energy Efficiency*, *16*(7), 80. https://doi.org/10.1007/s12053-023-10157-x

Brohmer, H., & Walcher, S. (2024). "I Have Some Serious Doubts About this Vaccine..." – Generic Conspiracy Beliefs Predict the Acceptance of the Covid-19 Vaccination. *Open Psychology*, *6*(2), 20220138. https://doi.org/10.1515/psych-2022-0138

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology*, *4*, 46573. https://doi.org/10.3389/fpsyg.2013.00279

Brzezinski, A., Kecht, V., Van Dijcke, D., & Wright, A. L. (2021). Science skepticism reduced compliance with covid-19 shelter-in-place policies in the united states. *Nature Human Behaviour*, *5*(11), 1519–1527. https://doi.org/10.1038/s41562-021-01227-0

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Čavojová, V., Šrol, J., & Ballová Mikušková, E. (2023). Scientific reasoning is associated with rejection of unfounded health beliefs and adherence to evidence-based regulations during the covid-19 pandemic. *Current Psychology*, 1–15. https://doi.org/10.1007/s12144-023-04284-y

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press. https://doi.org/10.2307/j.ctvc779w5

Cologna, V., Mede, N., Berger, S., Besley, J., Brick, C., Joubert, M., & Teran-Escobar, C. (2024). Trust in scientists and their role in society across 67 countries [Preprint]. https://www.researchgate.net/publication/378011710_Trust_in_scientists_and_their_role_in_society_across_67_countries

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. *Zeitschrift für Psychologie*, *227*(4). https://doi.org/10.1027/2151-2604/a000387

Dobbin, K. K., & Simon, R. M. (2011). Optimally splitting cases for training and testing high dimensional classifiers. *BMC Medical Genomics*, *4*, 1–8. https://doi.org/10.1186/1755-8794-4-31

Eberl, J. M., Greussing, E., Huber, R. A., & Mede, N. G. (2021). Wissenschaftsbezogener Populismus: Eine österreichische Bestandsaufnahme [science-related populism: An austrian evaluation] [Retrieved from Universität Wien]. https://viecer.univie.ac.at/corona-blog/corona-blog-beitraege/blog124/

Ebersole, C. R., Axt, J. R., & Nosek, B. A. (2016). Scientists' reputations are based on getting it right, not being right. *PLOS Biology*, *14*(5), e1002460. https://doi.org/10.1371/journal.pbio.1002460

European Commission. (2021a). *Europeans strongly support science and technology according to new eurobarometer survey* [Press release by the European Commission, September 23, 2021]. https://ec.europa.eu/commission/presscorner/detail/en/IP_21_4645

European Commission. (2021b). *Special eurobarometer 516 – april-may 2021: European citizens' knowledge and attitudes towards science and technology*. European Union. https://doi.org/10.2775/071577

Feist, G. J. (2012). Predicting interest in and attitudes toward science from personality and need for cognition. *Personality and Individual Differences*, *52*(7), 771–775. https://doi.org/10.1016/j.paid.2012.01.005

Funk, C., Johnson, C., & Hefferon, M. (2019). 5 key findings about public trust in scientists in the U.S. [Report by the Pew Research Center]. https://www.pewresearch.org/short-reads/2019/08/05/5-key-findings-about-public-trust-in-scientists-in-the-u-s/

Funk, C., Tyson, A., Kennedy, B., & Johnson, C. (2020). Science and scientists held in high esteem across global publics [Report by the Pew Research Center]. https://www.pewresearch.org/science/2020/09/29/science-and-scientists-held-in-high-esteem-across-global-publics/

Gemenis, K., & Littvay, L. (2023). Covid-19 conspiracy beliefs and government rule compliance [European Social Survey Report]. https://www.europeansocialsurvey.org/news/article/round-10-covid-19-questions-finalised

Haseltine, W. A. (2021). How covid changed science [May 25, 2021]. https://www.scientificamerican.com/article/how-covid-changed-science/

Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring laypeople's trust in experts in a digital age: The muenster epistemic trustworthiness inventory (meti). *PLOS One*, *10*(10), e0139309. https://doi.org/10.1371/journal.pone.0139309

Hofer, G., Brohmer, H., Grinschgl, S., Kanatschnig, T., Korda, Ž., & Potthoff, J. (2023). Can open science practices improve trust in science? [Pre-Study Poster presentation, Society for the Improvement of Psychological Science (SIPS) 2023 Meeting]. https://osf.io/4s6av

Hofer, G., Macher, S., & Neubauer, A. C. (2022). Love is not blind: What romantic partners know about our abilities compared to ourselves, our close friends, and our acquaintances. *Journal of Research in Personality*, *98*, 104211. https://doi.org/10.1016/j.jrp.2022.104211

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70–85. https://doi.org/10.1177/2515245917751886

Hüffmeier, J., Mazei, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology*, *66*, 81–92. https://doi.org/10.1016/j.jesp.2015.09.009

Köhler, T., & Cortina, J. M. (2023). Constructive replication, reproducibility, and generalizability: Getting theory testing for jomsr right. *Journal of Management Scientific Reports*, *1*(2), 75–93. https://doi.org/10.1177/27550311231176016

Krüger, J. T., Höffler, T. N., & Parchmann, I. (2022). Trust in science and scientists among secondary school students in two out-of-school learning activities. *International Journal of Science Education, Part B*, *12*(2), 111–125. https://doi.org/10.1080/21548455.2022.2045380

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lee, P., Goldberg, C., & Kohane, I. (2023). *The ai revolution in medicine: Gpt-4 and beyond*. Pearson. https://www.pearson.com/en-us/subject-catalog/p/the-ai-revolution-in-medicine-gpt-4-and-beyond/P200000011399/9780138279516

Limesurvey GmbH. (2023). *Limesurvey: An open source survey tool*. LimeSurvey GmbH, Hamburg, Germany. http://www.limesurvey.org

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, *51*(3), 485–504. https://doi.org/10.1002/ejsp.2752

Macfarlane, B., & Cheng, M. (2008). Communism, universalism and disinterestedness: Re-examining contemporary support among academics for merton's scientific norms. *Journal of Academic Ethics*, *6*, 67–78. https://doi.org/10.1007/s10805-008-9055-y

McCright, A. M., Dentzman, K., Charters, M., & Dietz, T. (2013). The influence of political ideology on trust in science. *Environmental Research Letters*, *8*(4), 044029. https://doi.org/10.1088/1748-9326/8/4/044029

McCright, A. M., Dunlap, R. E., & Marquart-Pyatt, S. T. (2016). Political ideology and views about climate change in the european union. *Environmental Politics*, *25*(2), 338–358. https://doi.org/10.1080/09644016.2015.1090371

Merton, R. K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology*, *1*(1-2), 115–126. https://heinonline.org/HOL/P?h=hein.journals/jolegpo1&i=115

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, *18*(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

OeAW - Austrian Academy of Sciences. (2022). *Oeaw wissenschafts-barometer pilotstudie 2022* [Report by Gallup International]. https://www.oeaw.ac.at/wissenschaftsbarometer/ergebnisse/2022

Philipp-Muller, A., Lee, S. W., & Petty, R. E. (2022). Why are people antiscience, and what can we do about it? *Proceedings of the National Academy of*

Sciences, *119*(30), e2120755119. https://doi.org/10.1073/pnas.2120755119

Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress*. Penguin UK. https://www.penguin.co.uk/books/267227/enlightenment-now-by-pinker-steven/9780141979090

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rabesandratana, T. (2019). These are the countries that trust scientists the most—and the least [June 18, 2019]. *Science*. https://doi.org/10.1126/science.aay4391

Ranney, M. A., & Clark, D. (2016). Climate change conceptual change: Scientific information can transform attitudes. *Topics in Cognitive Science*, *8*(1), 49–75. https://doi.org/10.1111/tops.12187

Riccardi, P. (2023). How to restore trust in science through education. *Nature Physics*, *19*(2), 146–147. https://doi.org/10.1038/s41567-023-01941-8

Robson, S. G., Baum, M. A., Beaudry, J. L., Beitner, J., Brohmer, H., Chin, J. M., ..., & Thomas, A. (2021). Promoting open science: A holistic approach to changing behaviour. *Collabra: Psychology*, *7*(1), 30137. https://doi.org/10.1525/collabra.30137

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rosman, T., Bosnjak, M., Silber, H., Koßmann, J., & Heycke, T. (2022). Open science and public trust in science: Results from two studies. *Public Understanding of Science*, *31*(8), 1046–1062. https://doi.org/10.1177/09636625221100686

Schneider, J., Rosman, T., Kelava, A., & Merk, S. (2022). Do open-science badges increase trust in scientists among undergraduates, scientists, and the public? *Psychological Science*, *33*(9), 1588–1604. https://doi.org/10.1177/09567976221097499

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Song, H., Markowitz, D. M., & Taylor, S. H. (2022). Trusting on the shoulders of open giants? open science increases trust in science for the public and academics. *Journal of Communication*, *72*(4), 497–510. https://doi.org/10.1093/joc/jqac017

Spälti, A. K., Lyons, B., Stoeckel, F., Stöckli, S., Szewach, P., Mérola, V., ..., & Reifler, J. (2023). Partisanship and anti-elite worldviews as correlates of science and health beliefs in the multi-party system of spain. *Public Understanding of Science*, *32*(6), 761–780. https://doi.org/10.1177/09636625231154131

Szabelska, A., Pollet, T. V., Dujols, O., Klein, R. A., & IJzerman, H. (2021). A tutorial for exploratory research: An eight-step approach [Preprint]. https://doi.org/10.31234/osf.io/cy9mz

UNESCO. (2021). Unesco recommendation on open science [Programme and Meeting Document]. https://doi.org/10.54677/MNMH8546

van Dalen, H. P. (2021). How the publish-or-perish principle divides a science: The case of economists. *Scientometrics*, *126*(2), 1675–1694. https://doi.org/10.1007/s11192-020-03786-x

van Dijk, D., Manor, O., & Carey, L. B. (2014). Publication metrics and success on the academic job market. *Current Biology*, *24*(11), R516–R517. https://doi.org/10.1016/j.cub.2014.04.039

Vranic, A., Hromatko, I., & Tonković, M. (2022). "i did my own research": Overconfidence, (dis)trust in science, and endorsement of conspiracy theories. *Frontiers in Psychology*, *13*, 931865. https://doi.org/10.3389/fpsyg.2022.931865

White, R. E., & Cooper, K. (2022). Hermeneutic inquiry. In R. E. White & K. Cooper (Eds.), *Qualitative research in the post-modern era* (pp. 59–91). Springer. https://doi.org/10.1007/978-3-030-85124-8_3

Wissenschaft im Dialog. (2023). Science barometer 2023 [Report retrieved from https://wissenschaft-im-dialog.de/projekte/wissenschaftsbarometer/#erhebung-2023].