

**Exploring Social Psychologists' Local Perspectives on Direct and Conceptual
Replication and their Contribution to Research Quality**

Robert van Ark¹

S3201309

¹Faculty of Behavioural and Social Sciences, University of Groningen, The Netherlands

PSB3N-BT15: Bachelor Thesis

Group: 2121_1a_21

Supervisor: J.M. Hoek

Second evaluator: dr. J.P. Wessel

In collaboration with: Maria Bompa, Kaiti Evgeniou, Rafael Funke, Larissa Hoß, Colm

O'Fuarthain

February 10, 2022

Author Note

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned.

Abstract

While psychology as a whole is said to be in a replication crisis, in particular social psychology has endured heavy criticisms. Calls are being made for conducting more direct replications so as to estimate the reliability of research across psychological fields. However, recently it has been argued that the applicability and value of replication (types) should be assessed locally per (sub)field. This study examines the local role and value of direct and conceptual replication in social psychology. 77 social psychologists from Dutch universities participated in a pilot survey. The survey yielded quantitative and qualitative descriptive data as to the importance of direct and conceptual replication and their relation to research quality. The majority of participants believed both replication types to be important, though more so for conceptual replication. These results contrast the calls for more direct replications across all of psychology, and thus provide initial information that a local assessment of the role and value of replication is warranted in social psychology and beyond. Replication's relation to research quality was found to be distinctly different for both replication types. Hence, awareness of these distinct contributions can help optimize utilization of direct and conceptual replication in social psychology. Notably, nearly half of the sample consisted of PhD students, misrepresenting the overall population. Moreover, using a convenience sample might have biased our sample to strong views on replication. Therefore, caution is required when considering our results as descriptive of the population of social psychologists in the Netherlands.

Keywords: social psychology, direct replication, conceptual replication, localism, research quality

Exploring Social Psychologists' Local Perspectives on Direct and Conceptual Replication and their Contribution to Research Quality

“Reproducibility is the cornerstone of science” (Simons, 2014, p. 76). This idea that replicability¹ is at the foundation of good science is not new. For Dunlap (1926), replicability was a requirement for proof, and a demarcation criterion between science and pseudoscience. However, despite the epistemic authority that is given to replicability across science, replication studies are rarely conducted (e.g. Evanschitzky et al., 2007; Makel et al., 2012; Kelly, 2019; Hardwicke et al., 2021). Moreover, it turned out that many ground-breaking and well-cited findings failed to replicate in the social, behavioural, and biomedical sciences (Romero, 2019; Spellman, 2015). Consequently, a consensus grew in the scientific community on the existence of a replication crisis (Romero, 2019; see Baker, 2016). In psychology, the results from large scale projects such as the Reproducibility Project: Psychology (RP:P; OSC, 2015) drew great attention to the claim that psychology was indeed in crisis (Wiggins & Christopherson, 2019). Its outcomes were darkest for social psychology, whose replication rate was estimated to be as low as 25%. For this reason, but also for revealed cases of fraud (see Shrout & Rodgers, 2018) and fishy practices (e.g. Gelman, 2016), social psychology was heavily criticized for lacking reliable research. Consequently, debates followed on the role and value of replication for social psychology. These debates are dominated by loud and prominent voices, while the opinions of the majority remain unknown. Inspired by calls for a local view on replication (e.g. Guttinger, 2020), the present

¹ The notion of reproducibility and reproducing refers to the possibility (as a quality of a study) and attempt, respectively, to obtain the same numerical results by running the same software and code on similar data (Plessner, 2018; Romero, 2019). However, the term reproducibility is often not intended to imply this notion, but meant synonymously to the term replicability. This holds for the writing quoted from Simons (2014) in the first paragraph of this page (it differs from the said notion in that he does not think experimental units must be exactly the same). If the contrary is not explicitly mentioned, the term reproducibility quoted from another's writing is interpreted as interchangeably with replicability. This paper does not aim to discuss reproducibility in its said distinct notion, and I will therefore solely use the term 'replicability' in my own writing.

study investigates the beliefs of social psychologists on 1) how replication studies should be conducted in social psychology (directly or conceptually), and 2) how these replication types contribute to research quality in social psychology.

Replication in Psychology

However, what is a replication exactly? Across science, interpretations of replicability differ widely (Plessner, 2018). Generally, replication refers to redoing an original experiment, at a varying closeness and thus with different purposes. More specifically, various typologies and definitions for replication have been and continue to be suggested (e.g. Asendorpf et al., 2013; Hudson, 2021). Moreover, social psychologists' beliefs about replication are likely to differ from other psychologists' beliefs. Since objects of study, environmental control, and reliance on statistics differ widely between subfields, so do interpretations of replication and the value they yield (Leonelli, 2018). If replicability would still be approached without differentiation across science, some research fields cannot live up to that form of accountability and the expectations of reliability (Penders et al., 2019). Therefore, the role and value of replication should be assessed locally (Guttinger, 2020).

Throughout psychology, two types of replications dominate: direct and conceptual replication (Guttinger, 2020; Morawski, 2019). A direct replication refers to an attempt to recreate the critical elements of an original study - those elements deemed necessary to replicate the original finding, such as the sampled population and utilized operationalizations - in order to test the validity of its results. In a conceptual replication, procedural changes are made deliberately – such as different sampled populations or utilized operationalizations - in order to test the generalizability and robustness of the original finding. The number of theoretically relevant changes made can range from one to multiple (Fraser et al., 2020; Zwaan et al., 2018).

These two types of replications do in reality not represent such a simple dichotomy, but rather a continuum (Fraser et al., 2020). Yet in practice, direct and conceptual replication are still mainly considered a dichotomy, which is reflected in most publications concerning replication, including those most influential (e.g. Nosek & Errington, 2020; Zwaan et al., 2018).

The relative value of direct and conceptual replication is still heavily debated within psychology (Machery, 2020). These debates involve arguments on one form being superior over the other (e.g. Simons, 2014) or on how to combine both forms optimally (e.g. Earp and Trafimow, 2015). All in all, debates on replication in psychology do not focus on *whether* replication studies should be conducted. Instead, the focus is on *how* they should be conducted.

The Ongoing Debates

A first perspective on *how* replications should be conducted holds conceptual as superior to direct replication. It is argued that a successful conceptual replication (obtaining similar results as the original study) can increase confidence in the underlying theory, since the theory is more likely to be accurate when it generalizes across multiple operationalizations. Direct replications, however, are said to only increase one's confidence in operationalizations, which is of lesser value to scientific advancement and mostly relevant when considering practical applications such as psychotherapeutic solutions (Crandall & Sherman, 2016; Stroebe & Strack, 2014). Furthermore, those supporting conceptual replication's superiority emphasize that psychology's objects of study are dynamic and context-sensitive (Morawski, 2021), especially so for social psychology (Ramscar, 2015). Therefore, direct replications are said to have considerable potential to fail. Such failures would be regarded as uninformative, as using identical procedures in replications cannot inform us on why a finding does not replicate (Stroebe & Strack, 2014).

A second perspective holds direct replication as the superior replication type. Its main function is argued to be that of measuring a finding's reliability and thus correcting for false-positives (Simons, 2014). Only then can we be sure scientific claims are credible (Nosek & Errington, 2020). Moreover, it is argued that because of deliberate procedural changes, a failed conceptual replication cannot tell whether the deliberate changes caused the failed replication effort or that the original finding was found by chance. Thus, conceptual replications cannot tell whether an original finding is reliable, and are therefore argued to be uninformative as a type of replication (Pashler & Harris, 2012). For this reason, some hold that conceptual 'replication' is in actuality no replication, and should be regarded (and named) as a generalizability test rather than a replication of an original study (Nosek & Errington, 2020; Zwaan et al., 2018).

A third perspective takes the middle ground in the debate. Here people argue for the importance of both types of replications. It is argued that direct replication tests the reliability of findings, and thus, "solid starting points for theoretical developments [through conceptual replications]" are built (Asendorpf et al, 2013, p. 140). Earp and Trafimow (2015) share this view, and add that conceptual replication is futile without its direct counterpart. The underlying theory and its boundary conditions cannot be confidently specified when the original finding's reliability is not investigated to begin with.

Replication and Research Quality in Social Psychology

How replication is valued in social psychology cannot be unequivocally deduced from the debates described above. A previous study investigated the value of replication for psychology broadly (Agnoli et al., 2021), where the majority of participating Australian and Italian psychologists reported that they believed replication is very important (76% and 59%, respectively). The local value for social psychology, however, remains unclear. Besides information on social psychologists' perspectives on direct and conceptual replication,

information on the perceived value of replication can help to progress the debates regarding replication.

In the present paper, the local value of direct and conceptual replication is explored by assessing how they contribute to research quality in social psychology. In doing so, the conception of research quality used distinguishes between four different dimensions – each containing myriad aspects and practices (such as replication) that contribute to research quality: solidity (well-founded and soundly conducted research), originality (providing new knowledge), scientific value (relevance for previous as well as future research), and societal value (Aksnes et al., 2019; Gulbrandsen, 2000). Replication can contribute to research quality through the various functions it holds. For example, direct replication might add to scientific value, for it is said to inform, albeit not conclusively (Earp & Trafimow, 2015), on the reliability of an original finding (Simons, 2014).

The present study

The debates on replication's role and value are mainly held by loud and prominent voices in literature and other (online) media. To assess the local role and value of replication in social psychology, however, it is important to get a larger understanding of social psychologists' views on replication. The present study descriptively explored opinions of social psychologists from Dutch universities in order to tighten the information gap between vocal voices' views and those of less prominent academics. Previous research regarding replication in psychology has done so for Australian and Italian psychologists (Agnoli et al., 2021). The present study offers additional value by exploring replication in social psychology locally, by distinguishing clearly between direct and conceptual replication, and by investigating their relationship with research quality. This attempt was led by the following research question: "How do social psychologists believe replication studies should be

conducted: directly and/or conceptually? To what extent are these forms indicative of research quality and why?”

Methods

A new survey was designed to evaluate social psychologists' opinions not only on replication, but also on the reform movement (see Morawski, 2019), open science practices and epistemological beliefs. This survey was based on earlier qualitative work on the replication crisis by other undergraduate students. Furthermore, it was a pilot for future studies that will assess similar subject matters on a larger, international scale.

Researcher Description

As the study was exploratory and descriptive in nature, researcher reflexivity is important. Presently, I am a student at the University of Groningen (UG) finishing the BSc Psychology. Regarding replication in psychology, I follow Asendorpf et al. (2013) in saying that direct replication is important for building reliable research, wherefrom theoretical development can be made through conceptual replications. Also, but not often mentioned elsewhere regarding replication, I believe that the dimension of societal value of research quality is very important, and that we need to know how reliable our research is for it to have societal impact.

Ethical Considerations

Prior to sample collection, the project was approved by the BSS-Psychology Ethics Committee at the University of Groningen. The code for approval is PSY-2122-S-0016. Participation was voluntary and could be ended at any time during the survey. Participants also provided informed consent prior to data collection. Email addresses required for survey dissemination were available publicly. Data was processed anonymously. We did not collect personal data such as name or email address during the survey or metadata such as IP addresses; therefore, a participant's answers could not be linked back to them.

Participants

Our target population consisted of social psychologists from Dutch universities. Since the main aim of our pilot study was to receive and integrate the feedback on our survey for the final survey, we aimed for a relatively small minimum sample size. Furthermore, as an important part of our study was the thematic analysis of the open questions, we regarded a minimum of 20 participants as sufficient. We deemed a response rate of 10% as realistic and in turn contacted 246 psychologists. Using a convenience sample, we approached researchers from University of Groningen (UG) (102), VU Amsterdam (27), University of Amsterdam (47), Tilburg University (34), and Radboud University (RU) (36) by extracting email addresses from their department websites. All members from the departments were included, with exception of secretaries and external affiliates; that is, from researchers, to PhD students, to full professors and lecturers. The universities mentioned supra were selected because they clearly separated social psychology from other departments such as organizational psychology and because email addresses were easily extractible. The selection sequence began with the present researchers' own university (UG), since we expected the highest response rate from them, and thereafter continued by decreasing city population size within the Netherlands. After extracting all email addresses from the fifth university (RU), we had obtained 246 email addresses and stopped the sampling procedure.

Procedure

We sent a Qualtrics link and the informed consent form out to our target population via email. This email included information on why the participants were being contacted, namely that we were looking for social psychologists to share their perceptions and opinions on the crisis debate, the reform movement and their methodological proposals. Respondents were also informed that the resulting data would be used for several bachelor theses and may eventually contribute to publication in a scientific journal. The survey ran for three weeks and

reminders to answer the survey were sent one and two weeks after our initial invitation. The invitation emails and informed consent form can be found in Appendix A.

Limitations of the sampling procedure

As we worked with a convenience sample, certain types of responses may be under- or overrepresented. Moreover, it is quite likely that researchers who participated in our study are different from those who chose not to fill out the survey. One possibility is that those with stronger opinions on the reform movement and its practices were more likely to answer. In addition, a heated public Twitter debate regarding the reform movement, escalating when a “reformer” publicly criticized a newly published social psychology study, occurred days prior to survey distribution (Brown, 2021). Due to the perceived denunciative and personal nature of the criticism, and the surrounding tone debate (e.g. Derksen, 2021) underlying reform discussions, polarised opinions might have been exacerbated by the debate. The possibility of bias is taken into account in the discussion of the results.

Survey Description

Although the survey used in the current study is novel, qualitative work in former bachelor and master theses, which investigated psychologists' perspectives on the replication crisis and open science practices, contributed to item generation (Futjes, 2021; Hershler, 2021; Nicolai, 2021; Pool, 2021; Sales, 2021; Schmidt, 2021; Schwarzbach, 2021). We also consulted survey designs which assessed the role of replication in ecology (Fraser et al., 2020) and psychology (Agnoli et al., 2021) and literature considering epistemological and ontological differences between reformers and challengers (Derksen, 2019; Flis, 2019; Morawski, 2019).

The survey consists of four core sections and it was anticipated that it would take 15 minutes for participants to complete it. First, the participants were asked about their epistemological and ontological views regarding (psychological) science. The second section,

most important for the current research purposes, broadly investigated the participants' views on (1) the purposes of new replication studies (generalizability, falsification, and/or confirmation of established results), (2) the importance of conducting direct and conceptual replication, and (3) the extent to which direct and conceptual replication are indicative of research quality, with all three aspects in relation to their field(s) of expertise. Then, with an open-ended question, we asked why participants believed that either direct or conceptual replication, when successful, is or is not indicative of research quality. Further optional open-ended questions asked if there were other important quality indicators outside replication in their field, and gave respondents the opportunity to provide broad thoughts on replication or the replication items. The third block of questions gathered information on open science concepts, practices and applications. Lastly, participants evaluated critical reflections on the reform movement and obstacles to proposed reforms. The complete survey can be found in Appendix E.

Data Analysis

Quantitative Analysis

A quantitative data analysis plan was made before the data were inspected. I did not deviate from this plan during the quantitative analysis.

Seven quantitative items were descriptively analysed: Q24 (generalization), Q25 (falsification), Q26 (confirmation), Q28 (importance of direct replications), Q29 (importance of conceptual replications), Q30 (direct replication as indication of research quality), and Q31 (conceptual replication as indication of research quality). The quantitative items yielded 0-100 scores (100 corresponding to complete agreement and 0 to no agreement at all), or a missing value when a "not applicable to my field" choice was selected. In these cases, these responses were excluded from analysis. The 0-100 scores were analysed using medians,

interquartile range, and total spread. We used boxplots to visualize the scores of the seven replication items, including all data points and highlighted outliers.

In addition to the descriptions of the quantitative items themselves, correlations were considered to yield informative descriptions of patterns between related item pairs. For example, a highly negative correlation ($r < -.7$) between items Q25 (falsification) and Q26 (confirmation) could show a pattern wherein people believe that new replication studies should either falsify *or* confirm existing results, whereas a highly positive correlation ($r > .7$) could show a pattern wherein people believe that new replication studies should both falsify *and* confirm existing results. No predictions were made as to the outcomes of the correlation computations.

The data violated the assumptions of linearity, normality, and absence of outliers for Pearson's correlation (see Figure B1, Figure B2 & Table B1). Therefore, Spearman's rank-order correlation was computed for the relevant item pair that met the assumption of monotonicity: Q30|Q31 (direct and conceptual replication indicative of research quality). A scatterplot is displayed for this pair. Although desired, correlations were not computed for Q28|Q29 (importance of direct and conceptual replication), Q28|Q30 (direct replication being important and an indication of quality), Q29|Q31 (conceptual replication being important and an indication of quality) and each pair in the Q24-Q26 cluster (replication for generalizability, falsification, confirmation), as they did not meet the assumption of monotonicity (see Figure B1 & Figure B3). A detailed, transparent explanation regarding correlation computations is included in Appendix B.

These analyses were conducted for 94 collected responses (response rate of approximately 38%). 17 responses from self-reported non-social psychologists were excluded (including the only dishonest response), yielding a response sample of 77 social psychologists. Of these remaining responses, 15 were incomplete. For median and spread

calculations, we considered all numerical responses as valuable and as such, we did not remove partial responses. However, the computation of correlations required complete responses across the pair and therefore partial responses were not considered for the respective correlation pairs.

We analyzed the data using open-source statistical software programmes R (R Development Team, 2018) and JASP (JASP Team, 2021). The exact lines of code used for analysis in R are included in Appendix C.

Qualitative Analysis

Thematic analysis by nature encourages deviations. Therefore, we explicated below the actual procedure with which the qualitative analysis was carried out.

In cooperation with two other students, a thematic analysis (TA) was performed as described by Braun and Clarke (2006) to investigate why respondents believe replication studies should be conducted directly or conceptually, and why they believe these forms are (not) indicative of research quality. In doing so, we recursively carried out six phases: 1) familiarization with the data, 2) code generation, 3) searching for themes, 4) reviewing themes, 5) defining and naming themes, and 6) writing the report. As our codes and themes were based on the data, the conducted TA is of inductive nature. We described what is explicitly mentioned in the data and did not go beyond what has been said by participants by examining underlying ideas and assumptions.

First, we familiarized ourselves with all open question responses to the survey. Meanwhile, all contributors (Kate Evgeniou, Colm O’Fuarthain, and Robert van Ark) individually highlighted informative responses about replication and generated a non-exhaustive list of initial codes that described the highlighted responses, without attaching it to specific text extracts. Thereafter, all contributors collaborated to generate one collective codebook. We compared the initial codes from each collaborator, checking for differences and

similarities. Similar codes were combined to form one code. Codes that were unique to a collaborator were also included. Thereafter, definitions were created for all codes. This preliminary codebook consisted of 25 codes. Then, all collaborators collectively applied these codes to the text data. Codes were only applied if a consensus was met by all collaborators that the code fit that specific data extract. Although TA is subjective by nature, we refrained from interpreting the text beyond what was explicitly said or in a way that would suit our research interests. During the coding process, eight new codes were added to describe relevant data that existing codes did not yet describe. Thereafter, the collaborators examined the codebook once more. Four codes that were not assigned to any pieces of data were removed from the codebook.

We then began the theme generation process by independently sorting similar codes into themes (i.e. groups of codes) and generating initial theme names. We compared our independent themes, and engaged in discussions over the appropriate final grouping and naming. The process of grouping and naming was done recursively. Simultaneously with these discussions, codes were again assessed on their distinctness and relevance. Thus, four more codes were removed and changed, and the relevant text extracts were reassigned. Hence, we identified six themes from 25 codes. The final codebook containing definitions and exemplars can be found in Appendix D. These are displayed and discussed in the *Results*.

Results

Of all 77 participating social psychologists, 73 participants worked from the Netherlands and four from various other countries. The median number of years that participants had been working in academia was six years (IQR: 7 [4, 11]). Considering current job positions, 35 participants were PhD students, four post-doctorates, 17 assistant professors, eight associate professors, six full professors, and seven respondents identified working in other positions.

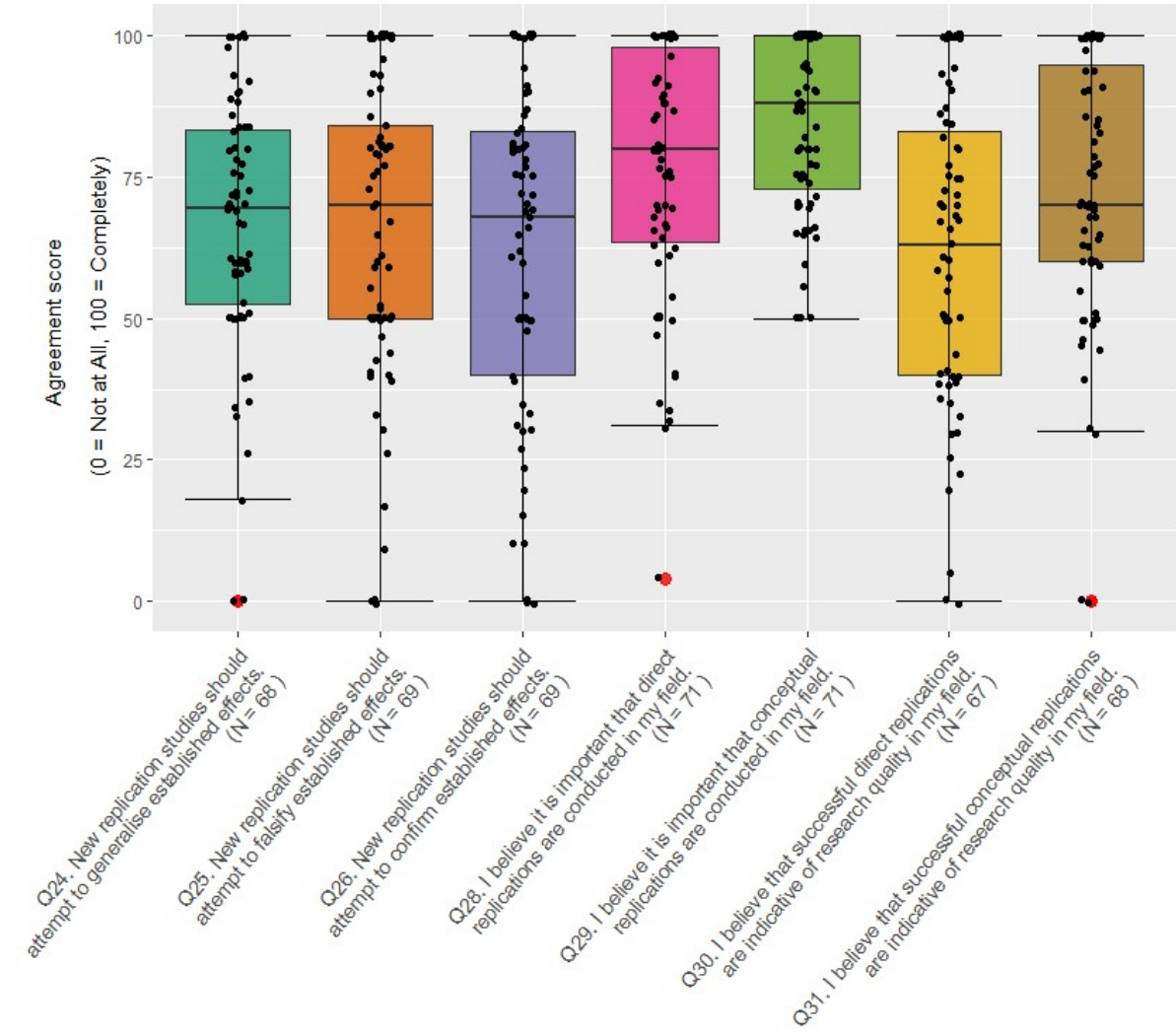
Quantitative Results

Replication

Figure 1 shows the responses on the seven items asking about the value and aims of replication. Each item is discussed in detail below.

Figure 1

Boxplots for the seven relevant quantitative survey items (with jitter function; outliers are indicated by red dots)



Generalizability, Falsification, and Confirmation. Participants were asked to assess their agreement with purposes that may predicate replication. Results indicate similar patterns of responses on the statements that new studies should attempt to generalize ($Mdn = 69.5$, $N = 68$), falsify ($Mdn = 70$, $N = 69$), and confirm ($Mdn = 68$, $N = 69$) established effects. However, the spread of agreement among respondents was less similar across the three items (generalizability: IQR: 30.75 [52.5, 83.25]; falsification: IQR: 34 [50, 84]; confirmation: IQR: 43 [40, 83]).

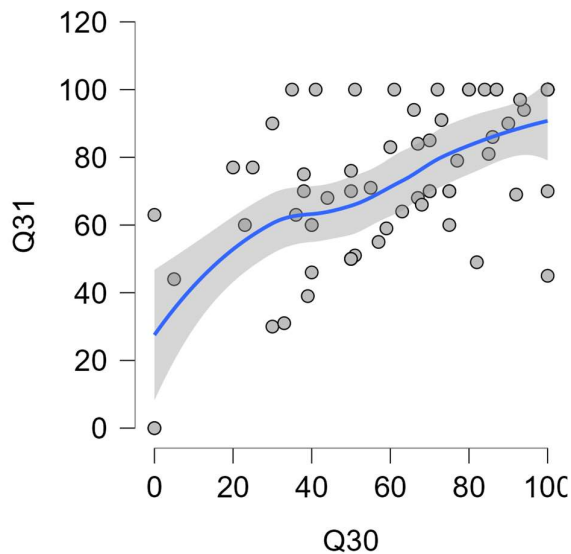
Importance of Direct and Conceptual Replication. There was high agreement that the conduction of direct replication ($Mdn = 80$, $N = 71$) as well as that the conduction of conceptual replication ($Mdn = 88$, $N = 71$) is important in social psychology. Moreover, the spread of agreement among participants was smaller for conceptual replication (IQR: 27 [73, 100]) than for direct replication (IQR: 34.5 [63.5, 98]). Notably, the lowest agreement scores concerning the importance of conceptual replication and direct replication were 50 and 4, respectively. This was not observed in any of the other item scales, which all recorded responses at the lowest possible value: zero.

Successful Replication as an Indication of Research Quality. When participants were asked if successful direct and conceptual replication are indicative of research quality, we estimated the median for direct replication ($Mdn = 63$, $N = 67$) to be lower than that of conceptual replication ($Mdn = 70$, $N = 68$). Again, the spread of agreement among participants was found to be smaller for conceptual replication (IQR: 34.75 [60, 94.75]) than for direct replication (IQR: 43 [40, 83]).

Correlation Pattern. Spearman's correlation coefficient (Spearman's rho) was computed for one item pair. As shown in Figure 2, a moderately strong positive relationship was found between the agreement scores on the statements that successful direct (Q30) and conceptual (Q31) replication are indicative of research quality ($r_s = .55$).

Figure 2

Scatter plot including a smooth regression line for the items on direct (Q30) and conceptual (Q31) replication being indicative of research quality



Thematic Analysis

Using thematic analysis, 25 codes were identified from the text, resulting in six themes (see Table 1).

Table 1

Identified Themes with Pertaining Codes (Number of Occurrences Across Responses between brackets)

Themes	Assigned Codes
Epistemology & Ontology	<i>Social psychology and context-sensitivity (10)</i> <i>Objectivity and truth as foundations for science (7)</i> <i>Research is subjective by nature (6)</i> <i>Universal and stable effects exist in science (1)</i>
Process and Conclusions of Replication	<i>Successful single replication is not conclusive (7)</i> <i>Replication is a learning and quality process (4)</i> <i>Incentives and bias for failed replications (2)</i> <i>Replication should not have purposes (1)</i>
Direct Replication: Functionality	<i>Direct replication for reliability (9)</i> <i>Direct replication for robustness (3)</i> <i>Direct replication indicative of quality of methodology (3)</i>

Direct Replication: Drawbacks	<i>Direct replication is uninformative (5)</i> <i>Direct replication reinforces original biases and mistakes (3)</i> <i>Direct replication not applicable in social psychology (2)</i>
Conceptual replication: Functionality	<i>Conceptual replication for generalizability (12)</i> <i>Conceptual replication for theory development (7)</i> <i>Conceptual replication for theory boundary conditions (4)</i> <i>Conceptual replication for validity (4)</i> <i>Conceptual replication and context-sensitivity (3)</i> <i>Conceptual replication overcomes methodological limitations and bias (3)</i>
Broad judgements regarding both replication types	<i>Conceptual over direct replication (13)</i> <i>Both replication types are uninformative (3)</i> <i>Both replication types are similarly important (2)</i> <i>Direct over conceptual replication (2)</i> <i>Nature of study determines which type of replication (1)</i>

Epistemology & Ontology

Apart from replication's functions, some participants also mentioned their stance on knowledge (generation) and truth. Ten responses identified that social psychology and its effects are inherently context sensitive. This can be problematic as "*some effects were studied 30-40 years ago and ... psych effects are culturally and time dependent.*" Participants thought that this must be taken into account when processes such as replication are considered, especially so for direct replication. Relatedly, six responses expressed that the findings and phenomena in social psychology will always be influenced by the subjective nature of its researchers, and that this should be acknowledged by critics and practitioners alike. Contrastingly, seven responses identified that pursuing objectivity and single truth-seeking were fundamentally valuable and realistic in social psychology and science as a whole. Relating this to replication, one respondent stated that "*if we cannot make any replicable observations, then it is very unlikely that any of our explanations captures parts of the truth*". Similarly, one response maintained that similar to the physical sciences, social psychology does possess stable and universal effects.

Process and Conclusions of Replication

Furthermore, participants mentioned purposes of replication studies and what can be concluded from them. Firstly, four responses stated that doing replication is mainly a learning process (“*It's not the results of the replication that matters. What matters is that we do them and learn from them.*”) that informs us on the research quality. Regarding the latter, if replication efforts fail, they need not say that the initial findings were false, but that other ways of capturing phenomena could be sought after. Furthermore, seven responses stated that outcomes of a single replication study are inconclusive. For both direct and conceptual replication, a single outcome cannot tell us definitively whether a finding is “*true*”, but “*can reveal more nuances to our understanding of the effect*”. Also, one response mentioned that replications should not have purposes such as confirmation, but that one should be open to all possible outcomes. Relatedly, two responses mentioned that currently incentives for failing replications outweigh those for successfully replicating, possibly biasing the replicator towards failing to replicate.

Direct Replication: Functionality and Drawbacks

Participants’ responses reflected two assessments of direct replication: its positive functionalities and its drawbacks. A first function, described in nine responses, is to test the reliability of an original finding. Thus, successful direct replication would serve as “*direct evidence*” of the original findings; that is, that those findings are not found by chance or questionable research. Relatedly, three responses noted that direct replication can test the robustness of findings. Lastly, three responses mentioned that direct replication can test the methodological quality of an original study. If an original study is replicable, it would show that its methodology is sound; especially so when the replication effort was successful.

Two of the three drawbacks are related to each other. Since social psychology was said to be a field where context and meanings continually change, responses mentioned that

direct replication is an uninformative replication type for social psychology (five responses) or that it was not applicable to the field altogether (two responses). The third drawback, mentioned in three responses, was said to be the perpetuation of mistakes and biases of original studies in replication efforts, yielding “*less meaningful*” replication results.

Conceptual Replication: Functionality

The respondents also mentioned various functionalities of conceptual replication. Notably, they did not mention any drawbacks about conceptual replication. First and foremost, conceptual replication was seen as a tool to generalize findings (12 responses). For example, participants noted that when conceptual replications fail, while having sampled from different populations, this is initial evidence the theoretical hypothesis does not generalize to that population. Four responses noted that thus a possible boundary condition is found. Seven responses noted that by seeking for generalizations and/or boundary conditions, conceptual replication can contribute to theory development. Furthermore, four responses discussed the ability of conceptual replication to inform on validity, meaning that by measuring phenomena in different ways, one can increase their confidence that they are effectively measuring what they purport to measure. It was said that without conceptual replications, “*it is difficult to be certain that the study examines the phenomenon of interest.*” Furthermore, three responses mentioned the function of conceptual replication to overcome methodological limitations and biases. By replicating an original study conceptually, and therefore procedurally differently, methodological improvements can be made, and biases in the original set-up would be overcome. Lastly, in three responses it was said that a replication that “*accounts for variations in different studies*” better suits a field wherein “*theories can be quite time, culture and context dependent.*”

Broad Judgements on Direct and Conceptual Replication

Reflecting on replication broadly, most responses (13) believed conceptual replication to be distinctly more valuable than direct replication, alluding to the many functional advantages of conceptual replication, such as finding generalizations and boundary conditions. Relatively fewer responses (2) identified direct replication to be more valuable, as conceptual replications would allow more “*degrees of freedom*” and have an unclear relationship between operationalizations. The latter was said to limit what can be concluded from conceptual replications. Lastly, some (two responses) expressed that both replication types are equally important in social psychology, while pointing to the different functions that both types serve. Others (three responses) expressed that neither replication type is necessarily informative since “*the results of the field are so contextual*” and successful replication is “*meaningless if the original phenomenon/effect is not of theoretical value*”. Lastly, one response mentioned that the decision for a replication type depends on the nature of the study.

Discussion**Replication in Social Psychology**

The first aim of this study was to investigate social psychologists’ perspectives on the conduction of direct and conceptual replications. Almost all participants agreed that both direct and conceptual replication are important to be conducted (median agreement score of 80 and 88, respectively). The qualitative analysis also showed that many participants valued conceptual replication more than direct replication. Moreover, *every* respondent that deemed the question applicable to their field ($N = 71$), agreed to a considerable degree with the statement that conceptual replication is important to be conducted. Thus, the perspective that conceptual replications are uninformative (Pashler & Harris, 2012) was not held by any of our participants.

Consistent with literature holding conceptual replication as superior (Crandall & Sherman, 2016; Stroebe & Strack, 2014), participants most often mentioned the ability of conceptual replication to test the generalizability of findings across different operationalizations and populations, thereby contributing to theory development. Thus, conceptual replication is important for explaining psychological capacities through theoretical development. This is a pressing argument for conceptual replication's importance in social psychology, since it is said that we are losing sight of what to actually explain: psychological capacities, not effects (Van Rooij & Baggio, 2021). A considerable group of participants also believed conceptual replication to be superior as it better suits a context-sensitive field such as social psychology. This assessment possibly results from the belief that ever-changing contexts and time influence both operationalizations - rendering replications with identical operationalizations impossible (Stroebe & Strack, 2014) - and objects of study (Morawski, 2021) – so that we need overarching theories not “statistical phenomena” (Stroebe & Strack, 2014, p. 7).

Moreover, participants stated that conceptual replication overcomes methodological mistakes and biases, whereas direct replication would reinforce them. However, a few other participants believed the opposite, saying that direct replication can test the methodological quality of a replicated study. This disagreement is also found in literature concerning replication. Among others, Crandall and Sherman (2016) posit that a direct replication perpetuates flaws in the original study design, whereas others state that direct replication tests the methodological quality of the replicated study (e.g. Zwaan et al., 2018).

Furthermore, a considerable group of participants, including some holding conceptual replication as superior, mentioned the distinct virtue of direct replication in its proposed ability to test the reliability of original findings. Specifically, they noted that direct replication

can distinguish real findings from false-positives (see Simmons et al., 2011) and that it increases the credibility of unexpected and new findings.

Furthermore, some noted that direct replication can test the robustness of a finding. I refrain from interpreting this, though, as the term has no uniform meaning. Zwaan et al. (2018) use the term interchangeably with ‘reliability’ and link robustness indeed to direct replication. Crandall and Sherman (2016), however, mention that direct replication can test the robustness of operationalizations, not findings, and that conceptual replication tests the robustness of findings.

Replication and Research Quality

The second aim was to investigate if and why direct and conceptual replication contributes to research quality. Both successful direct and conceptual replication were assessed to be indicative of research quality (median agreement score of 63 and 70, respectively). These assessments were generally more critical than assessments of the importance of the two replication types. This seems to imply that additional research quality demands more than *successful* replication. This view was held by at least some participants, which is reflected in their saying that single successful replications are not conclusive as to an original study’s reliability and quality. This can possibly be explained by the idea that replication efforts can never be conclusive, but only informative, as underlying factors such as auxiliary assumptions make absolute falsification or confirmation hardly possible (Earp & Trafimow, 2015).

Furthermore, the more critical assessments could mean that participants believed that, although successful direct and conceptual replication are indicative of research quality, the importance of conducting replications is also determined by other factors. For example, the state of overall theoretical development could be viewed as poor (participant: “*I think we in social psychology need to realize that most of our work mainly lacks theory*”), resulting in a

contextual factor that gives additional importance to conceptual replication beyond its contribution to research quality.

There is theoretical literature that does relate replication to research quality (e.g. Leonelli, 2018). However, *how* replication exactly contributes to research quality is not discussed. In order to investigate how replication exactly contributes to research quality with a uniform and explicated definition, a common model is used that distinguishes between four dimensions: solidity, originality, scientific value, and societal value (Aksnes et al, 2019; Gulbrandsen, 2000).

The functions of testing generalizability and identifying boundary conditions can contribute to research quality through the dimension of scientific value, as it sheds additional light on the theoretical hypotheses of earlier research and direct future investigative attempts. Moreover, the function of building theories is argued to increase scientific advancement (Crandall & Sherman, 2016), and contributes to scientific value, as it increasingly enables us to explain the world and its phenomena. Furthermore, the participants noted that conceptual replication can increase our confidence that we are validly studying the constructs we purport to. Thus, conceptual replication contributes to research quality by improving on the solidity and plausibility of the constructs and manipulations used in social psychology.

Regarding direct replication, many participants thought its main function is testing the reliability of findings. Reliable research is a point from which future studies can build, both through reference and by building on it theoretically. Thus, it contributes to research quality through the dimension of scientific value. Moreover, reliable research contributes to research quality by being societally valuable, since for practical matters such as policy recommendations and psychotherapeutic solutions reliable study outcomes are important (Crandall & Sherman, 2016).

That the most frequently mentioned functions of direct and conceptual replication (respectively: testing reliability and generalizability & boundary conditions) both contribute to the dimension of scientific value seems to imply that replication's greatest contribution to research quality lies in this dimension. Hence, what constitutes the dimension of scientific value (the relevance for previous & future research and scientific explanation as a whole) seems to be most important when discussing replication. This is useful information about what underlies the importance accredited to direct and conceptual replication, and acknowledging it can help move the debates on replication in social psychology forward. Still, both direct and conceptual replication offer distinctly valuable contributions to research quality through three dimensions, and awareness of these distinct contributions can improve the use of replication in social psychology and beyond.

Implications and Future Directions

The present study provided initial information as to how social psychologists believe replications should be conducted *locally* (Guttinger, 2020). Calls for more replication studies are made (Agnoli et al, 2021), though there is considerable disagreement on the role of replication in psychology (Wiggins & Christopherson, 2019). The described overall assessment contrasts the loud and prominent voices in psychology broadly, which tend to favor direct replication (e.g. Nosek & Errington, 2020; Simon, 2014). This contrast is important initial information that implies that the role of replication should indeed be considered locally, as opposed to broadly, in both psychology and other sciences. Future research should investigate the local role for a greater, international sample, possibly while developing this pilot survey further.

Moreover, conceptual replication is said to be valued most by the senior researchers and direct replication is said to be valued more by the younger generations (Crandall & Sherman, 2016; Morawski, 2021). Interestingly, this was not reflected in our sample

comprising relatively many early-career researchers. Possibly, this inconsistency is due to the limitation of the sampling procedure that people with particular opinions could have been more likely to participate; those against the calls for direct replication in this case. However, this demarcation between senior researchers and younger generations is often made, and information against it should be taken seriously and further investigated.

Furthermore, within the majority of participants that assessed conceptual replication to be superior, there was a considerable group that also valued direct replication, rendering two subgroups within the said majority. The difference between those two subgroups seems to be their beliefs on the applicability of direct replication in the context-sensitive field that social psychology is. Future research could investigate where the difference in opinions about direct replication's value arises. Possibly, these two subgroups of social psychologists study different objects which are (not) believed to lend themselves to direct replication. In so doing, demarcation criteria for direct replication's applicability in social psychology could be investigated.

Limitations

Using a convenience sample may have resulted in a potentially biased sample, even more so due to a heated Twitter debate occurring prior to survey distribution. This debate was mentioned by participants in open response questions, where one participant acknowledged that her responses might have been different if asked before the Twitter debate. Moreover, our sample consisted for nearly half of PhD students, misrepresenting the population of social psychologists from Dutch universities. For both reasons, and since the study is descriptive in nature, caution is warranted when interpreting these results as a description of the population overall.

Another limitation was the absence of a definition of research quality accompanying the questions. This was bypassed by using a common model of research quality. Hence, the

conclusions are not based on the participants' own conception. Future research that investigates how replication contributes to research quality could investigate the participants' accompanying conception of research quality, possibly including more valuable contributions of replication to research quality.

Concluding Remarks

Overall, both direct and conceptual replication were valued among our sample of social psychologists from Dutch universities, though generally conceptual replication was valued higher. These assessments contrast the calls that are made for more direct replications in psychology. Therefore, the results provide initial information that a local approach to replication is warranted in social psychology and beyond. Furthermore, the functions of direct and conceptual replication were argued to contribute to research quality in distinct ways. Awareness of these distinct contributions can improve the use of direct and conceptual replication in social psychology and beyond.

References

- Agnoli, F., Fraser, H., Singleton Thorn, F., & Fidler, F. (2021). Australian and Italian Psychologists' View of Replication. *Advances in Methods and Practices in Psychological Science*, 4(3), 25152459211039218.
- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, 18(3), 91-93.
- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *Sage Open*, 9(1), 2158244019829575.
- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27(2), 108–119. <https://doi.org/10.1002/per.1919>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Brown, N. [@sTeamTraen]. (2021, November 10). *It's 2021 and social psychologists are still publishing papers like this: Women wearing not much clothing on a night out don't get cold because they believe that they "look hot" (geddit?)* [Tweet]. Twitter. <https://twitter.com/sTeamTraen/status/1458217774677954560>
- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66, 93–99. <https://doi.org/10.1016/j.jesp.2015.10.002>
- Derksen, M. (2019). Putting popper to work. *Theory & Psychology*, 29(4), 449–465.

- Derksen, M., & Field, S. M. (2021). The tone debate: knowledge, self, and social order. *Review of General Psychology*, 00(0), 1-12.
- Dunlap, K. (1926) The experimental methods of psychology. In: *Psychologies of 1925*, ed. C. Murchison, pp. 331–53. Clark University Press.
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, 621–621.
<https://doi.org/10.3389/fpsyg.2015.00621>
- Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007). Replication research's disturbing trend. *Journal of Business Research*, 60(4), 411-415.
- Flis, I. (2019). Psychologists psychologizing scientific psychology: an epistemological reading of the replication crisis. *Theory and Psychology*, 29(2), 158–181.
<https://doi.org/10.1177/0959354319835322>
- Fraser, H., Barnett, A., Parker, T. H., & Fidler, F. (2020). The role of replication studies in ecology. *Ecology and Evolution*, 10(12), 5197-5207.
- Furtjes, J. (2021). *Where to go from here? Researcher perspectives on the replication crisis and the state of the academic world* [Unpublished undergraduate thesis]. University of Groningen.
- Gelman, A. (2016, July 5). *Gremlins in the work of Amy J. C. Cuddy, Michael I. Norton, and Susan T. Fiske*. Statmodeling Columbia.
<https://statmodeling.stat.columbia.edu/2016/07/05/30596/>
- Gulbrandsen, J. M. (2000). *Research quality and organisational factors: An investigation of the relationship*. Trondheim: NTNU.
- Guttinger, S. (2020). The limits of replicability. *European Journal for Philosophy of Science*, 10(2). <https://doi.org/10.1007/s13194-019-0269-1>
- Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., & Ioannidis, J.

- P. A. (2021). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). *Perspectives on Psychological Science*, (20210308). <https://doi.org/10.1177/1745691620979806>
- Hershler, A. (2021). *In crisis do we unite? Debates on the nature and impact of the credibility revolution*. [Unpublished undergraduate thesis]. University of Groningen.
- Hudson, R. (2021). Explicating Exact versus Conceptual Replication. *Erkenntnis*, 1-22.
- JASP Team (2021). JASP (Version 0.16)[Computer software].
- Kelly, C. D. (2019). Rate and success of study replication in ecology and evolution. *PeerJ*, 7, e7654.
- Leonelli, S. (2018). “Re-Thinking Reproducibility as a Criterion for Research Quality.”
- Machery, E. (2020). What is a replication?. *Philosophy of Science*, 87(4), 545-567.
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7(6), 537-542.
- Morawski, J. (2019). The replication crisis: how might philosophy and theory of psychology be of use? *Journal of Theoretical and Philosophical Psychology*, 39(4), 218–238. <https://doi.org/10.1037/teo0000129>
- Morawski, J. (2021). How to True Psychology’s Objects. *Review of General Psychology*, 10892680211046518.
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *Plos Biology*, 18(3), 3000691. <https://doi.org/10.1371/journal.pbio.3000691>
- Open Science Collaboration. (2015). Psychology. estimating the reproducibility of psychological science. *Science (New York, N.y.)*, 349(6251), 4716. <https://doi.org/10.1126/science.aac4716>
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments

- examined. *Perspectives on Psychological Science*, 7(6), 531–536.
- Penders, B., Holbrook, J. B., & De Rijcke, S. (2019). Rinse and repeat: Understanding the value of replication across different ways of knowing. *Publications*, 7(3), 52.
- Plessner, H. E. (2018). Reproducibility vs. replicability: a brief history of a confused terminology. *Frontiers in neuroinformatics*, 11, 76.
- Pool, A. (2021). *You, Science and The Replication Crisis – A Thematic Analysis Attempting To Understand What Replication Means To Researchers*. [Unpublished undergraduate thesis]. University of Groningen.
- R Development Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramscar, M. (2015, August 15). *The unspeakable in the pursuit of the unrepeatable*. Ramscar.<https://ramscar.wordpress.com/2015/08/05/the-unspeakable-in-pursuit-of-the-unrepeatable/#more-704>
- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, 14(11), e12633.
- Sales, S. B. H. (2021). *Researchers’ experiences of the replication debate and their well being*. [Unpublished undergraduate thesis]. University of Groningen.
- Schmidt, H. (2021). *ECRs stuck between two worlds*. [Unpublished undergraduate thesis]. University of Groningen.
- Schmidt, S. (2009). Shall we really do it again? the powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100.
<https://doi.org/10.1037/a0015108>
- Schwarzbach, N. R. (2021). *Motivations for and reservations about the use of open science practices: An interview study*[Unpublished master’s thesis]. University of Groningen.
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction:

- broadening perspectives from the replication crisis. *Annual Review of Psychology*, 69(1), 487–510. <https://doi.org/10.1146/annurev-psych-122216-011845>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1), 76–80.
- Spellman, B. A. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science*, 10(6), 886–899. <https://doi.org/10.1177/1745691615609918>
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59–71.
- Van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*, 16(4), 682–697.
- Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: an overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, 39(4), 202–217. <https://doi.org/10.1037/teo0000137>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41. <https://doi.org/10.1017/S0140525X17001972>

Appendix A

Invitation Emails and the Informed Consent Form

Initial Invitation Email

Dear [title+ name],

We are contacting you, because we are doing a pilot study for a large-scale study about perceptions of the replication/credibility crisis and the 'reform movement'. In this context, social psychology is a field that is often *talked about*, but in our opinion, not *talked to* enough. We are curious how you, as a social psychologist, have experienced the crisis debate, the reform movement and the proposed changes. The results of this survey will facilitate a critical evaluation of the aims and accomplishments of the reform movement. Because this is a pilot survey, we are especially interested in your feedback about our questions (content, wording, etc.).

We kindly invite you to take part in the survey via this link:

https://rug.eu.qualtrics.com/jfe/form/SV_8quywigev6mhQa2

Participation will take approximately 15 minutes. Your contribution would be greatly appreciated!

In the attachment of this email, you can find more information about the study. Feel free to reply to this email if you have questions or concerns. If you would like to be kept up to date about this research and its results, please send us an e-mail at perceptions.of.reform@rug.nl.

Kind regards,

Robert van Ark, Maria Bompá, Kaiti Evgeniou, Colm Ó Fuartháin, Rafael Funke and Larissa Hoß

Research team:

Joyce Hoek, MSc

Nina Schwarzbach, MSc

Sarahanne Field, MSc

Merle Pittelkow, MSc

Dr. Rink Hoekstra

Prof. dr. Don van Ravenzwaaij

Faculty of Behavioural and Social Sciences, Rijksuniversiteit Groningen, the Netherlands

Reminder 1

Dear [title+ name],

A week ago we contacted you because of our survey about “perceptions of the reform movement”, and we highly appreciate your participation. In case you did already fill out the survey: thank you very much! Please disregard this email. Unfortunately, we cannot remove you from our mailing list, since participation is anonymous.

In case you have not filled out the survey, we would kindly like to remind you that participation in our survey is still possible.

You can participate in the survey using the following link:

https://rug.eu.qualtrics.com/jfe/form/SV_8quywigev6mhQa2

In response to previously raised concerns:

- We invited 250 people to this pilot survey. Therefore, it would be difficult to trace back your identity on the basis of demographic data we ask for.
- If you'd like to give more detailed feedback verbally or via email, please do not hesitate to contact us.
- Some said that the survey takes longer than 15 minutes. Please take into consideration that it might take up to 30 minutes depending on how detailed your answers are.

Thank you in advance,

Robert van Ark, Maria Bompá, Kaiti Evgeniou, Colm Ó Fuartháin, Rafael Funke and Larissa Hoß

Research team:

Joyce Hoek, MSc

Nina Schwarzbach, MSc

Sarahanne Field, MSc

Merle Pittelkow, MSc

Dr. Rink Hoekstra

Prof. dr. Don van Ravenzwaaij

Faculty of Behavioural and Social Sciences, Rijksuniversiteit Groningen, the Netherlands

Reminder 2

Dear [title+ name],

We would like to remind you one last time about our survey about “perceptions of the reform movement”. You still have time to fill it out until December 8th, after which the survey will close. Your participation is still highly appreciated!

In case you did already fill out the survey: thank you very much! Please disregard this email. Unfortunately, we cannot remove you from our mailing list, since participation is anonymous.

You can participate in the survey using the following link:

https://rug.eu.qualtrics.com/jfe/form/SV_8quywigev6mhQa2

In response to previously raised concerns:

- We invited 250 people to this pilot survey. Therefore, it would be difficult to trace back your identity on the basis of demographic data we ask for. In addition, we’ve decided not to publish the data of this pilot survey on OSF or any other open data platform.
- If you’d like to give more detailed feedback verbally or via email, please do not hesitate to contact us.
- Some said that the survey takes longer than 15 minutes. Please take into consideration that it might take up to 30 minutes depending on how detailed your answers are.

Thank you in advance,

Robert van Ark, Maria Bompá, Kaiti Evgeniou, Colm Ó Fuartháin, Rafael Funke and Larissa Hoß

Research team:

Joyce Hoek, MSc

Nina Schwarzbach, MSc

Sarahanne Field, MSc

Merle Pittelkow, MSc

Dr. Rink Hoekstra

Prof. dr. Don van Ravenzwaaij

Faculty of Behavioural and Social Sciences, Rijksuniversiteit Groningen, the Netherlands

Informed Consent**“PERSPECTIVES OF THE REPLICATION CRISIS, SCIENCE AND THE REFORM MOVEMENT”**

Welcome and thank you very much for participating in our survey. For more information about this pilot study, please refer to the study information form in the email or contact us at: perceptions.of.reform@rug.nl

Please read the information below and indicate whether you agree with it before continuing with this survey. You have the right to take a screenshot of this information.

- I have read the information about the research. I have had the opportunity to ask questions about it.
- I understand what the research is about, what is being asked of me, which consequences participation can have, how my data will be handled, and what my rights as a participant are.
- I understand that participation in the research is voluntary. I myself choose to participate. I can stop participating at any moment. If I stop, I do not need to explain why. Stopping will have no negative consequences for me.

I consent to participating in this study

Appendix B

Correlation computations

This appendix concerns itself with the figures and tables that correspond to the relevant assumption checks for Pearson's r and Spearman's rank-order correlations.

For Pearson's correlation, there are three key assumptions that must be met; 1) that a linear relationship exists between the two chosen variables, 2) that both variables are approximately normally distributed, and 3) that no significant outliers exist. To test these assumptions with our data in R, we 1) assessed scatterplots for linear relationships, 2) conducted Shapiro-Wilks tests and generated Q-Q plots to test normality, and 3) identified outliers from descriptive boxplots (see Figure 1). The data violated all three assumptions. First, the scatterplots failed to indicate any clear linear relationships between all relevant variables, as can be seen in Figure B1. Secondly, all Shapiro-Wilks tests returned significant evidence to reject the null hypothesis, as seen in Table B1, meaning that a normality assumption should be rejected. The violation of normality assumption was also reflected in the Q-Q plots (see Figure B2). Finally, outliers were observed for item Q24, Q28, and Q31.

Spearman's rank-order correlation (Spearman's rho) is a commonly used correlation when the assumptions of Pearson's r are notably violated (Akoglu, 2018). To test the assumption of monotonicity for Spearman's rho, we inspected the scatterplots in Figure B1. We determined visually that only the pairs Q28|Q30, Q29|Q31, and Q30|Q31 could meet the monotonicity assumption of Spearman's rho. We turned to software programme JASP to add a smooth regression line to the scatterplots in order to test the assumption confidently. Therefrom, we concluded that only item pair Q30|Q31 met the assumption of monotonicity (see Figure B3).

Figure B1

Scatter plot matrix for all relevant quantitative survey items

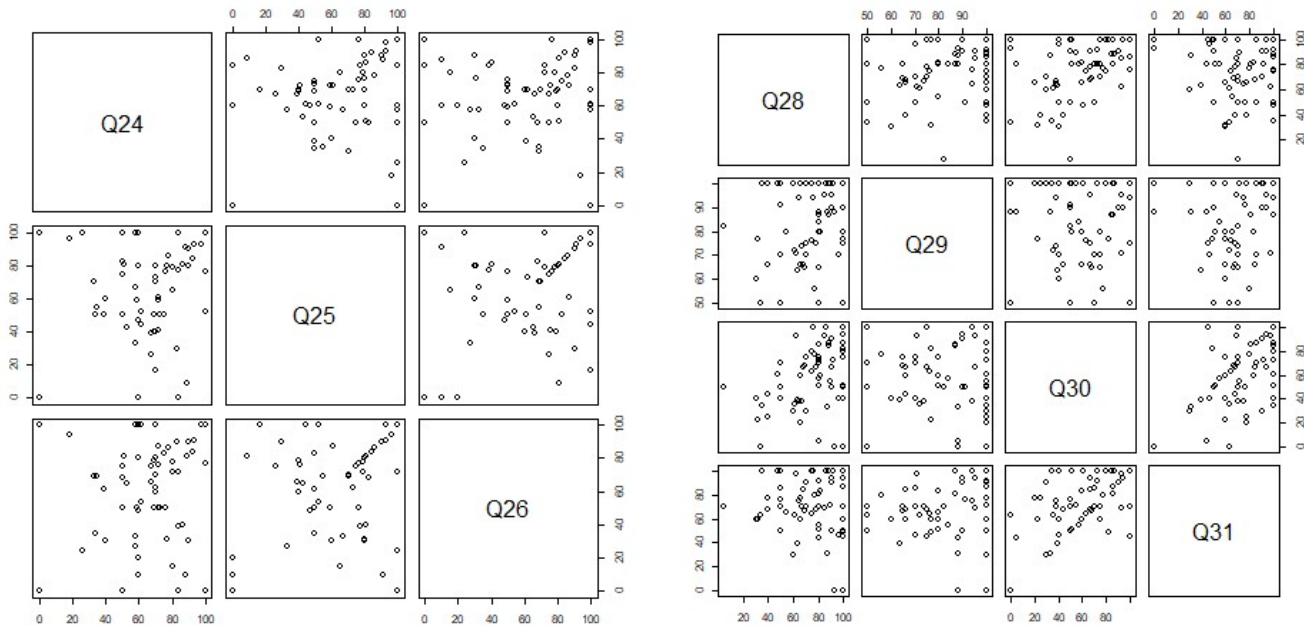


Table B1

Results for the Shapiro-Wilks test for normality assumption in the chosen items

Item	Statistic (<i>W</i>)	p-value
Q24	0.95	0.008
Q25	0.927	< 0.001
Q26	0.93	< 0.001
Q28	0.904	< 0.001
Q29	0.866	< 0.001
Q30	0.954	0.015
Q31	0.909	< 0.001

Figure B2

QQ plots for all relevant quantitative survey items

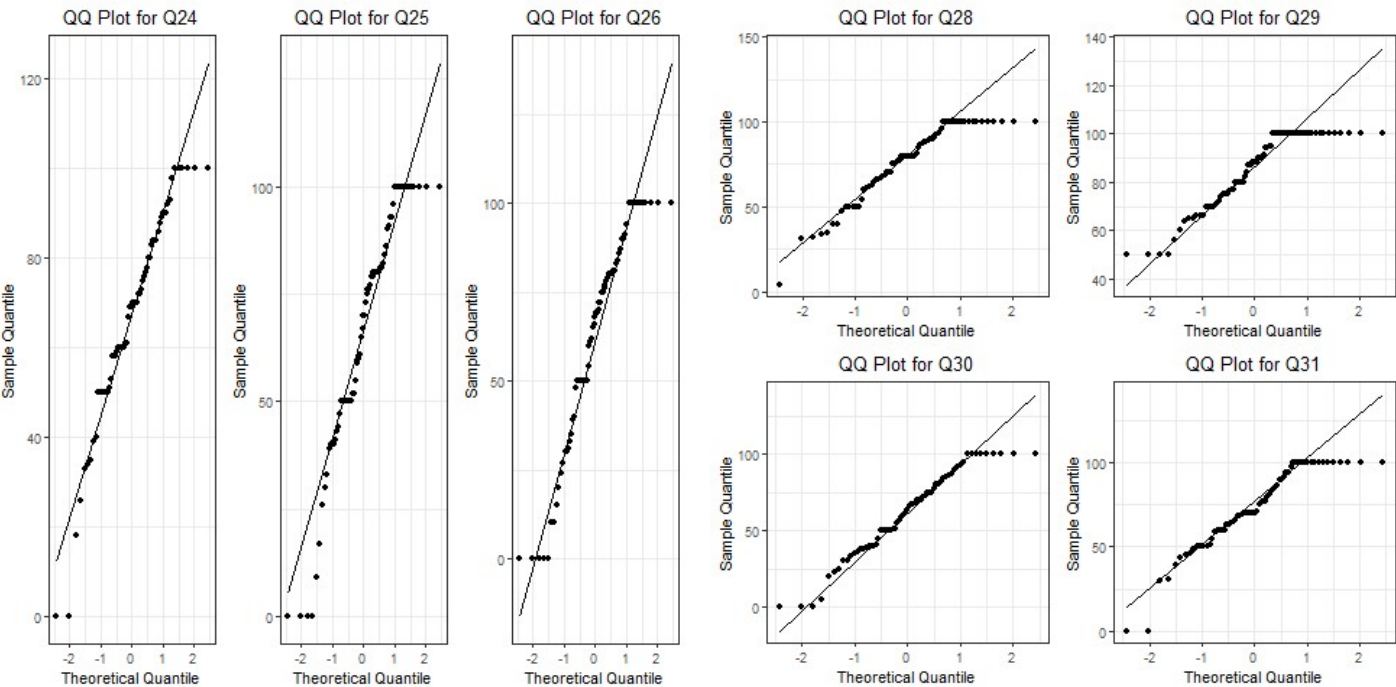
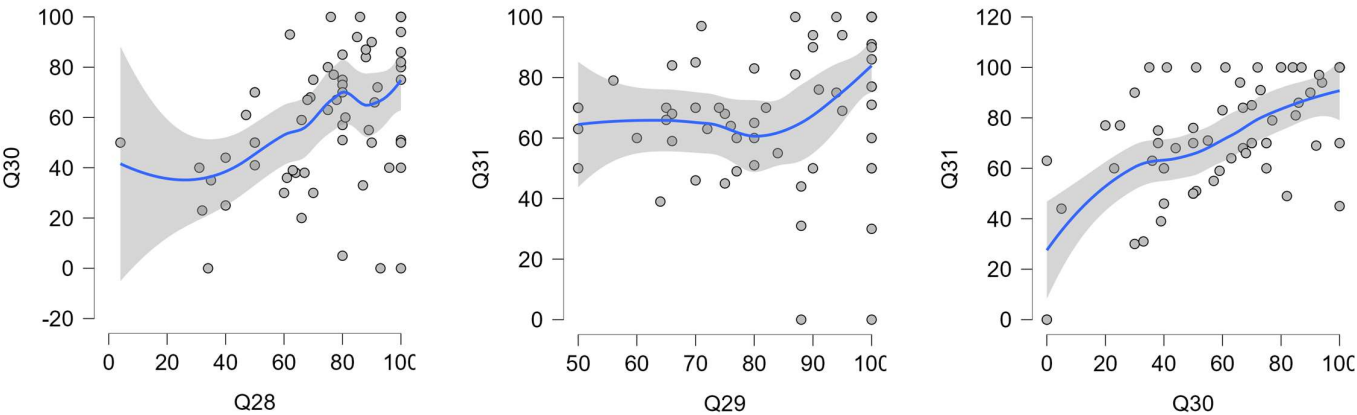


Figure B3

Scatterplots with regression line for the relevant three item pairs



Appendix C

R code used for quantitative analyses

```
#install.packages("ggplot2")
#install.packages("tidyverse")
#install.packages("hrbrthemes")
#install.packages("viridis")
#install.packages("reshape2")
#install.packages("grid")
#install.packages("gridExtra")

library(ggplot2)
library(tidyverse)
library(hrbrthemes)
library(viridis)
library(reshape2)
library(grid)
library(gridExtra)

## Setting working directory and preparing data for analysis
setwd("D:/R")
mydata = read.csv("DATAAREALREAL.csv", header = TRUE)
mydata2 = mydata[, c("fieldofexpertise", "Q99_1", "Q99_5", "Q99_6",
"Q99_7", "Q99_8", "Q110_1", "Q110_5", "Q123", "Q124")]
mydata2 = dplyr::filter(mydata2, grepl('24', fieldofexpertise)) # exclude non-social psychologists
mydata2 = dplyr::filter(mydata2, Q123 == "1" | Q123 == "" & Q124 == "1" | Q124 == "") #Excluding
possible dishonest responding and lack of attention
mydata2 = mydata2[, c("Q99_1", "Q99_5", "Q99_6", "Q99_7", "Q99_8", "Q110_1", "Q110_5")] #
removing exclusion columns
colnames(mydata2) = c('Q24', 'Q25', 'Q26', 'Q28', 'Q29', 'Q30', 'Q31')
mydata2 = as.data.frame(sapply(mydata2, as.numeric))

## Demographics
# Preparing Data
demogr = mydata[, c("fieldofexpertise", "workingcountry", "currentjob", "workinacademiayears",
"Q123", "Q124")]
demogr = dplyr::filter(demogr, grepl('24', fieldofexpertise)) # exclude non-social psychologists
demogr = dplyr::filter(demogr, Q123 == "1" | Q123 == "" & Q124 == "1" | Q124 == "") #Excluding
possible dishonest responding and lack of attention
demogr = demogr[, c("fieldofexpertise", "workingcountry", "currentjob", "workinacademiayears")] #
removing exclusion columns

# In which country are participants working
workingcountry_NL = sum(demogr$workingcountry == '144') # Number of participants working in
the Netherlands
workingcountry_notNL = sum(demogr$workingcountry != '144') # Number of participants working
outside the Netherlands

# How many years have participants been in academia
```



```

yearsinacademia = (as.data.frame(sapply(demogr$workinacademiayears,as.numeric))) # Putting the
Data as numerical
median_yearsinacademia = apply(yearsinacademia, 2, median, na.rm = TRUE)
all_quartiles_yearsinacademia = apply(yearsinacademia, 2, quantile, na.rm = TRUE)
Quartiles_1_yearsinacademia = all_quartiles_yearsinacademia[2,]
Quartiles_3_yearsinacademia = all_quartiles_yearsinacademia[4,]

# What is their current profession
num_UndergradStudent = sum(demogr$currentjob == '4')
num_ResearchAssistant = sum(demogr$currentjob == '7')
num_JuniorResearcher = sum(demogr$currentjob == '8')
num_PhDStudent = sum(demogr$currentjob == '9')
num_Postdoc = sum(demogr$currentjob == '10')
num_AssistantProf_UD = sum(demogr$currentjob == '11')
num_AssociateProf_UHD = sum(demogr$currentjob == '12')
num_FullProf = sum(demogr$currentjob == '13')
num_Other = sum(demogr$currentjob == '14')
#CurrentJob_Counts = c(num_UndergradStudent, num_ResearchAssistant,
#num_JuniorResearcher, num_PhDStudent, num_Postdoc,
# num_AssistantProf_UD, num_AssociateProf_UHD, num_FullProf)

## Calculating medians, IQR, quartiles, and a minimum value

# Medians
d_Medians = apply(mydata2,2,median, na.rm = TRUE)

# IQRs, along with Q1 and Q3
d_IQRs = apply(mydata2,2,IQR, na.rm = TRUE)
all_quartiles = apply(mydata2,2, quantile, na.rm = TRUE)
Quartiles_1 = all_quartiles[2,]
Quartiles_3 = all_quartiles[4,]

# Minimum value
complete_mydata2_min29 = mydata2[complete.cases(mydata2),5]
complete_mydata2_min28 = mydata2[complete.cases(mydata2),4]
min(complete_mydata2_min29)
min(complete_mydata2_min28)

## Boxplot chart
mydata2_m = melt(mydata2) #Change data structure for function ggplot
mylabels = c("Q24. New replication studies should \n attempt to generalise established effects.",
             "Q25. New replication studies should \n attempt to falsify established effects.",
             "Q26. New replication studies should \n attempt to confirm established effects.",
             "Q28. I believe it is important that direct \n replications are conducted in my field.",
             "Q29. I believe it is important that conceptual \n replications are conducted in my field.",
             "Q30. I believe that successful direct replications \n are indicative of research quality in my
field.",
             "Q31. I believe that successful conceptual replications \n are indicative of research quality in
my field.")
item_obs = c(sum(mydata2$Q24 > -1, na.rm = TRUE),

```

```

sum(mydata2$Q25 > -1, na.rm = TRUE),
sum(mydata2$Q26 > -1, na.rm = TRUE),
sum(mydata2$Q28 > -1, na.rm = TRUE),
sum(mydata2$Q29 > -1, na.rm = TRUE),
sum(mydata2$Q30 > -1, na.rm = TRUE),
sum(mydata2$Q31 > -1, na.rm = TRUE))
mylabels = paste(mylabels, "\n (N =", item_obs, ")") # Adding the number of observations to the
labels

ggplot(mydata2_m, aes(x = variable, y = value, fill = variable)) +
  geom_boxplot(alpha = 0.8, varwidth = TRUE, outlier.colour = "red", outlier.fill = "red", outlier.size =
3) +
  stat_boxplot(geom='errorbar') +
  geom_jitter(width = 0.08) +
  scale_fill_brewer(palette = "Dark2") +
  scale_x_discrete(labels = mylabels) +
  xlab("Questions") +
  ylab("Agreement score \n \n (0 = Not at All, 100 = Completely)") +
  xlab("") +
  theme(plot.margin = unit(c(0.5,0.5,0.5,2),"cm"), legend.position = "none", axis.text.x =
element_text(angle = 50, hjust=1, size = 10.5))

## Assumptions check for Pearson correlation coefficient
# Checking linearity assumption by means of scatterplots, visualised in a matrix

# Q24-26
png(file = "Linear Check Q24-26 scatterplot matrix.png")
pairs(~Q24+Q25+Q26, data = mydata2, main = "Scatter Plots, Q24-Q26")
dev.off()

# Q28-31
png(file = "Linear Check Q28-Q31 scatterplot matrix.png")
pairs(~Q28+Q29+Q30+Q31, data = mydata2, main = "Scatter Plots, Q28-Q31")
dev.off()

## Checking normality assumption by means of QQ plots and Shapiro-Wilks Test
# Shapiro-Wilks Test - when it is not significant, we can assume normality
Shap_Q24 = shapiro.test(mydata2$Q24)
Shap_Q25 = shapiro.test(mydata2$Q25)
Shap_Q26 = shapiro.test(mydata2$Q26)
Shap_Q28 = shapiro.test(mydata2$Q28)
Shap_Q29 = shapiro.test(mydata2$Q29)
Shap_Q30 = shapiro.test(mydata2$Q30)
Shap_Q31 = shapiro.test(mydata2$Q31)

# Q24-26 side by side QQ plots
qq24 = ggplot(mydata2, aes(sample = Q24)) + stat_qq() + ggtitle("QQ Plot for Q24") +
xlab("Theoretical Quantile") + ylab("Sample Quantile")+stat_qq_line() + theme_bw() +
theme(plot.title = element_text(hjust = 0.5))

```

```

qq25 = ggplot(mydata2, aes(sample = Q25)) + stat_qq() + ggtitle("QQ Plot for Q25") +
xlab("Theoretical Quantile") + ylab("Sample Quantile") + stat_qq_line() + theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
qq26 = ggplot(mydata2, aes(sample = Q26)) + stat_qq() + ggtitle("QQ Plot for Q26") +
xlab("Theoretical Quantile") + ylab("Sample Quantile") + stat_qq_line() + theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
png(file = "Normality Check Q24-26 matrix.png")
grid.arrange(qq24, qq25, qq26, ncol = 3, nrow = 1, top = textGrob("Normality Check Q24-26", gp =
gpar(fontsize=20, font=1)))
dev.off()

# Q28-31 side by side QQ plots
qq28 = ggplot(mydata2, aes(sample=Q28)) + stat_qq() + ggtitle("QQ Plot for Q28") +
xlab("Theoretical Quantile") + ylab("Sample Quantile") + stat_qq_line() + theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
qq29 = ggplot(mydata2, aes(sample=Q29)) + stat_qq() + ggtitle("QQ Plot for Q29") +
xlab("Theoretical Quantile") + ylab("Sample Quantile") + stat_qq_line() + theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
qq30 = ggplot(mydata2, aes(sample=Q30)) + stat_qq() + ggtitle("QQ Plot for Q30") +
xlab("Theoretical Quantile") + ylab("Sample Quantile") + stat_qq_line() + theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
qq31 = ggplot(mydata2, aes(sample=Q31)) + stat_qq() + ggtitle("QQ Plot for Q31") +
xlab("Theoretical Quantile") + ylab("Sample Quantile") + stat_qq_line() + theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
png(file = "Normality Check Q28-Q31 matrix.png")
grid.arrange(qq28, qq29, qq30, qq31, ncol = 2, nrow = 2, top = textGrob("Normality Check Q28-31", gp =
gpar(fontsize=20, font=1)))
dev.off()

## Spearman's correlation coefficient

# Assumptions of Pearson's coefficient were not met. For the pair of variables
# that was thought to have a monotonic relationship, Spearman's correlation
# coefficient is computed.

# Q30 & Q31
mydata2_3031 <- mydata2[,c(6,7)]
mydata2_complete_3031 <- mydata2_3031[complete.cases(mydata2_3031),]
cor(mydata2_complete_3031, method = "spearman")

```

Appendix D

Codebook including definitions and exemplars

Table D1

Codebook, definitions and exemplars used for the thematic analysis

Codes	Definition	Exemplars
Successful single replication is not conclusive	The success/failure of a single replication (both direct and conceptual) study should not tell us anything about the quality of the study.	<i>"...a successful or unsuccessful replication may not necessarily mean that a phenomenon is not true, but reveal more nuances to our understanding of what we study."</i>
Replication is a learning and quality process	Replication as a whole is a credible research quality method. Replication has important benefits for the field of social psychology that does not depend on the results of single studies, which themselves are not so beneficial and meaningful	<i>"It's not the results of the replication that matters. What matters is that we do them and learn from them."</i>
Incentives and bias for failed replications	The idea that the incentives for failed replications may outweigh successful replications for researchers, which may consciously or unconsciously affect their replication efforts	<i>"The underlying problem is that failed replications are seen as more newsworthy than successful replications, so that replicators can have more impact if their replication attempts fail."</i>
Replication should not have purposes	Here is the idea that when replication has goals, such as finding out if the outcome of a single study is or is not reproducible, it limits the productive output of this enterprise.	<i>"I believe replicators (or scientists in general) should not have such goals. The goal should be to establish whether a particular effect replicates, and the replicator should be open to all possible outcomes."</i>
Direct replication for reliability	Successful direct replication studies can say something about the reliability (a measure) of our results	<i>"Id say direct replication would be the first step, to ensure the reliability of the effect...."</i>
Direct replication for robustness	Successful direct replication studies can say something about the robustness (a characteristic) of our field and our theories.	<i>"Direct replication rate should be diagnostic of the robustness of findings published in a field."</i>
Direct replication indicative of quality of methodology	Successful direct replication studies can say something about the quality of the measures and methodology we use in social psychology	<i>"Direct replications also have their value and can indicate the stability of the work and the quality of research protocols (can someone replicate the work)."</i>
Direct replication is uninformative	Direct replication does not provide any new information and cannot prove anything, so it yields uninformative results	<i>"There are a lot of factors which might influence a direct replication to not be successful (...) a failed direct replication does not tell us all that much about the effect."</i>
Direct replication reinforces original biases and mistakes	As direct replication aims to follow the original study in as close a manner as possible, what ends up happening is that the mistakes and biases of the original researcher also end up in the new replication study, eliminating the possibility of improving the quality of our research.	<i>"Using the same materials/populations as before, as in direct replications, only further reinforces/empowers the potential biases involved in the original research."</i>
Direct replication not applicable in Social Psychology	Due to the context-sensitivity of the social sciences, the conditions between two different studies will always be different even if the methods are followed as closely as possible. Thus, direct replication cannot say anything about the research quality of our results, and can be considered not applicable in the field.	<i>"... there are cases when direct replication is difficult because of changed context or meaning..."</i>

Conceptual replication for generalizability	Conceptual replication increases the generalizability (for example, contexts, different populations, and different operationalizations) of phenomenon that have been studied	<i>"Conceptual replication is more important for generalizability than direct ones because it gives a sense that the way we study things can be applied to other contexts or samples or methods."</i>
Conceptual replication for validity	Conceptual replication for (construct) validity, in the sense that conceptual replication measures phenomena in different manners and thereby increases our confidence that we are effectively capturing the phenomenon we purport to.	<i>"In my opinion, conceptual replications are extremely important. Without them, it is difficult to be certain that the study examines the phenomenon of interest."</i>
Conceptual replication for theory boundary conditions	The success or failure of conceptual replication studies can inform us on the boundaries of our theories.	<i>"... conceptual replications can add important information on boundary conditions and extensions."</i>
Conceptual replication for theory development	Conceptual replication can be a more effective form of replication for building theories, which is considered a desired facet for science in social psychology.	<i>"For an effect to be meaningful it should be present in more than one study. Conceptual replications are thus important as they indicate in what context something is and is not present, which we can build on theoretically."</i>
Conceptual replication and context-sensitivity	Conceptual replication is more appropriate for social psychology due to the context sensitive nature of the field.	<i>"Especially conceptual replications are important as social psychological theories can be quite time, culture and context dependent."</i>
Conceptual replication overcomes methodological limitations and bias	By testing in a different manner than the original study, conceptual replication overcomes methodological limitations and bias in original studies.	<i>"Conceptual replications can overcome the methodological limitations or unique methodological features of the previous studies."</i>
Conceptual over direct replication	Responses that indicate that conceptual replication is distinctly more important than direct replication	<i>"Conceptual replications do the same [as direct replication] AND are indicative of whether the result is something generalizeable [sic] and not specific to the exact methods used in the original experiment."</i>
Both replication types are similarly important	To code when a response indicates that both replication types are similarly important.	<i>"I'd say direct replication would be the first step, to ensure the reliability of the effect, followed by conceptual replication for validity."</i>
Both replication types are uninformative	To code when a response indicates that both replication types, even if successful, do not provide any information or meaning.	<i>"Successful replications, whether direct or conceptual, can be meaningless if the original phenomenon/effect is not of theoretical value."</i>
Direct over conceptual replication	To code when a response indicates that direct replication is distinctly more important than conceptual replication.	<i>"Direct replication rate should be diagnostic of the robustness of findings published in a field. Conceptual replications are limited by the (typically) unclear correspondence in validity of measurements across studies purporting to test the same hypothesis."</i>
Nature of study determines which type of replication	When the nature of study determines which type of replication.	<i>"I think it depends heavily on the nature of the study. If we're talking about decision making processes, observing behavior in experiments, etc., direct replications have value. If we're talking about field experiments, conceptual replication might be more fruitful, as of, for example, cultural differences."</i>
Social psychology and context-sensitivity	The idea that social psychology as a field deals with context-sensitivity in terms of phenomena and dynamic systems.	<i>"Basically a complexity perspective confronts you with the possibility of fundamental uncertainty. Replication might be only possible in more stable situations of complex systems, and hence is not a good concept to study more turbulent stages in social systems."</i>
Objectivity and truth as foundations for science	The idea that in science objectivity and (single) truth are valuable and realistic cornerstones to hold	<i>"Replication is important, because if we cannot make any replicable observations, then it is very unlikely that any of our explanations captures parts of the truth."</i>
Research is subjective by nature	This code defines the idea that an ideal of an objective researcher and science is not realistic, and acknowledging that subjectivity of the scientists will always be contained within science and our findings	<i>"I think that there will always be an element of subjectivity in the kind of research we do ..."</i>

Universal and stable effects exist in science	The idea that there exists objects and effects in psychological science that are stable and universal.	<i>“I think there are phenomena in reality that have stable characteristics but that can never be described in any words.”</i>
---	--	--

Appendix E

Complete Survey

Start of Block 0: Informed Consent

Welcome and thank you very much for participating in our survey. For more information about this pilot study, please refer to the [Study information form](#) or contact us at: perceptions.of.reform@rug.nl. The study will take approximately 15 minutes, contains 11 sections and is best completed on a computer. Please read the information below and indicate whether you agree with it before continuing with this survey. You have the right to take a screenshot of this information. I have read the information about the research. I have had the opportunity to ask questions about it. I understand what the research is about, what is being asked of me, which consequences participation can have, how my data will be handled, and what my rights as a participant are. I understand that participation in the research is voluntary. I myself choose to participate. I can stop participating at any moment. If I stop, I do not need to explain why. Stopping will have no negative consequences for me. I consent to participating in this study:

- ☐ Yes, I consent to participation.
- ☐ No, I do not consent to participation.
-

End of Block 0: Informed Consent

Start of Block 1: Demographics

First, we'd like to ask you for some demographic data.

Q1. In what country are you currently working?

▼ Afghanistan ... Zimbabwe

Q2. What is your broad field of expertise?

- ☐ Social psychology
 - ☐ Developmental psychology
 - ☐ Industrial and organizational psychology/ work psychology
 - ☐ Environmental psychology
 - ☐ Experimental psychology
 - ☐ Personality psychology
 - ☐ Clinical (neuro) psychology
 - ☐ Cognitive psychology
 - ☐ Quantitative psychology
 - ☐ Biological psychology
 - ☐ Political psychology
 - ☐ Other, namely: _____
-

Q3. What is your current job position?

- ☐ (Undergrad) student
- ☐ Research Assistant
- ☐ Junior researcher
- ☐ PhD student
- ☐ Postdoc
- ☐ Assistant professor/UD
- ☐ Associate Professor/UHD
- ☐ Full professor
- ☐ Other, namely: _____

Q4. How long have you been working in academia? (years)

End of Block 1: Demographics

Start of Block 2: Terms

To have a consistent and shared understanding throughout the survey, we would like to clarify what the terms mean to us. Throughout the survey, you can always go back to these definitions using a pop-up button found at the bottom.

Direct replication: The attempt to conduct a study in a manner as close to the original as possible (the same population, methodology, and statistical analyses).

Conceptual replication: The attempt to test the same theoretical process or effect as an existing study, or understand boundary conditions of given phenomena, but that uses methods that vary in some way from the previous study.

Successful replication: When the replication study yields results which are sufficiently similar to the original study in terms of the strength of the effect and whether the effect goes in the same direction as the original. 'Sufficiently similar' varies, and is usually defined by the replicating author.

Open science: Open science aims to make science more transparent. Open science practices include among others: preregistration, registered reports, open data, open peer review, and open access publishing.

Metascience: The study of research itself, often with the aim of improving its practice. Meta-researchers study the scientific community and its actors, their methods and reporting, reproducibility, evaluation, behavior, and incentives.

Reform movement: There are many different words describing groups of people that are promoting change in science, including 'meta-science movement', 'open science movement' or 'reformer movement'. In the following we summarize people sharing concern with regards to improving science through either meta-scientific or transparent/open science practices as the 'reform movement'.

Q5. Optional: Do you have feedback on these definitions?

From now onwards, we will refer mostly to the reform movement. You can always go back to the definitions if you are unsure about the terms used in the survey.

End of Block 2: Terms

Start of Block 3: Reform movement

The next questions will be about how the aims of the reform movement resonate with you and your research practices.

Q6. Please indicate the extent to which you...

Not at all

Completely

...identify with the reform movement



Q7.

Do you agree with this statement: "I am part of the reform movement."?

- ☐ Yes
- ☐ No
- ☐ Don't know

Q8. Optional: Do you have any thoughts with regard to your identification with the reform movement you'd like to add here?

Q9. Optional: Do you have feedback on the questions about identification with the reform movement?

End of Block 3: Reform movement

Start of Block 4: Epistemology/Ontology

We would like to know more about how you think about science and knowledge in general. Please indicate how the following statements relate to your research.

Please indicate how the following statements relate to your research:

	Not at all	Completely
Q10. "For every phenomenon that I study, there are multiple valuable truths."		
Q11. "In my field of research, scientists can ultimately get to/reach the truth."		
Q12. "In my field of research, results depend on the perception of the researcher."		
Q13. "Science should be organized in such a way as to reduce scientists' biases."		
Q14. "In my field of research, the effects are dependent on the time period in which these studies took place rather than universal."		
Q15. "In my field of research, the effects are dependent on the culture where the study took place rather than universal."		
Q16. "In my field of research, the effects are dependent on the experimental setup rather than universal."		
Q17. "It is possible to specify all the boundary conditions that enable a theory to hold true."		
Q18. "Conducting a scientific study requires constant adaptation of the methods used."		
Q19. "The expertise of an individual scientist is important to study a phenomenon."		

Q20. Optional: Do you have any thoughts you'd like to add here?

Q21. Optional: Do you have feedback on the questions about science and knowledge in general?

End of Block 4: Epistemology/Ontology

Start of Block 5: Research Quality

The current survey includes some questions about the quality of research. First, we would like to know what you think of the current state of research quality in your field.

Please indicate the extent to which you agree with the following statement:

Not at all

Completely

Q22. "I think that research quality in my field is something that needs to be improved."

Q23. Optional: Can you elaborate?

End of Block 5: Research Quality

Start of Block 6: Replication







The next couple of questions will be about replication.

Please indicate the extent to which you agree with the following statement:

Not at all

Completely



Not applicable

Q24. "New replication studies should attempt to generalise established effects."	
Q25. "New replication studies should attempt to falsify established effects."	
Q26. "New replication studies should attempt to confirm established effects."	
Q27. "Original researchers of a study should participate in the process of replication."	
Q28. "I believe it is important that direct replications are conducted in my field."	
Q29. "I believe it is important that conceptual replications are conducted in my field."	

We would now like to ask some questions about replication and research quality.

Please indicate the extent to which you agree with the following statement:

Not at all Completely Not applicable

Q30. "I believe that successful direct replications are indicative of research quality in my field."	
Q31. "I believe that successful conceptual replications are indicative of research quality in my field."	

Can you elaborate on your previous two answers?

Q32. Why do you think that successful replication is, or is not, indicative of research quality in your field of research? Please indicate what type of replication you are talking about (i.e., direct, conceptual or any other form)?

Q33. Optional: Which quality indicators other than replication do you think are important in your field of research?

Q34. Optional: Do you have any thoughts you'd like to add here?

Q35. Optional: Do you have feedback on the questions about replication?

End of Block 6: Replication



Start of Block 7: Open Science Ideas

The next couple of questions are about your ideas of open science in general.

Please indicate the extent to which you agree to the following statements:

Not at all

Completely

Q36. I think that science in general should be transparent and open if possible.	
Q37. Generally, I think that the more transparent and open the research process is, the higher its quality and reliability.	

Q38. Optional: Do you have any thoughts you'd like to add here?

Q39. Optional: Do you have feedback on the questions about open science ideas?

End of Block 7: Open Science Ideas

Start of Block 8: Open Science Practices

The next couple of questions are about your thoughts on the practical application of open science.

Q40. Please give an estimate on how many hours of (informal) training on open science practices you have received.

Please indicate the extent to which you agree with the following statements:

Very Little Very Much Not applicable

Q41. "I feel like I have received sufficient (informal) training on how to practice open science."	<div></div>
Q42. "My working environment/colleagues encourage me to use open science methods to conduct my research."	<div></div>

Q43. Which of the following practices are you currently using in your research?

	Never	Rarely	Sometimes	Mostly	Always	I don't know what this means	Not applicable
Preregistration	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Registered reports	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open access publishing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open materials (code, metadata)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open peer review	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q44. Optional: Alternatively, which other open science practice are you currently using in your research?

	Never	Rarely	Sometimes	Mostly	Always
Other practice:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q45. Which of the following practices would you like to use (more) in your future research?

- ☐ Preregistration
- ☐ Registered reports
- ☐ Open access publishing
- ☐ Open data
- ☐ Open materials (code, metadata)
- ☐ Open peer review
- ☐ Other, namely: _____
- ☐ None

Q46. Optional: What would you need to practice (open) science the way you'd like to?

End of Block 8: Open Science Practices

Start of Block 9: Critique

From interviews, we gathered some information about how the reform movement is perceived. We will now like to know how much you agree with the next statements.

Q47. Please indicate the extent to which you agree with the following statement:

Not at all Completely Not applicable

"I have the feeling that people in the reform movement understand the practices of my field."



Q48. Optional: Please explain why (not)?

Please indicate the extent to which you agree with the following statement:

Not at all Completely Not applicable


Q49. "I feel like the reform movement addresses the most pressing issues regarding scientific quality in my field."



Q50. Optional: Please explain why (not)?

Please indicate the extent to which you agree with the following statement:

Not at all Completely Not applicable

Q51. "The proposed solutions solve the problems in my field sufficiently."	
--	--

Q52. Optional: Please motivate your answer.

Q53. The reform movement prioritizes some solutions over others. Please rank how **you think** the reform movement prioritizes the following issues (1=most priority, 16=least priority):

- _____ Preregistration/registered reports
- _____ Data/code sharing
- _____ Research methods other than inferential (qualitative, descriptive, exploratory)
- _____ Improving statistics (bayesian statistics vs NHST etc)
- _____ Theory or construct development
- _____ Bigger sample sizes
- _____ Slow science
- _____ Managing competitive culture in academia
- _____ More collaboration
- _____ More direct replication
- _____ More conceptual replication
- _____ Increasing diversity within universities
- _____ Increasing the importance of societal impact
- _____ More freedom to pursue your scientific interests
- _____ More job security
- _____ Nuanced reporting of results

Q54. Are you sure you finalised the ranking?

- ☐ Yes, I am
- ☐ No, I am not

Q55. Optional: What problems with regard to the quality of research in your field is the movement missing?

Q56. Optional: Do you have feedback on the questions about the priorities of the reform movement?

End of Block 9: Critique

Start of Block 10: Important Issues To Be Addressed

Q57. In order to improve **research quality** in your field, multiple solutions are suggested. Please rank how important **you think** they are to improve research quality in your field (1=most important, 16=least important).

- _____ More focus on preregistration/registered reports
- _____ More focus on data/code sharing
- _____ More focus on research methods other than inferential (qualitative, descriptive, exploratory)
- _____ More focus on improving statistics (Bayesian statistics and/or NHST etc.)
- _____ More focus on theory or construct development
- _____ More focus on bigger sample sizes
- _____ More focus on slow science
- _____ More focus on managing competitive culture in academia
- _____ More focus more collaboration
- _____ More focus on direct replication
- _____ More focus on conceptual replication
- _____ Increasing diversity within universities
- _____ Increasing the importance of societal impact
- _____ More freedom to pursue your scientific interests
- _____ More job security
- _____ More focus on nuanced reporting of results

Q58. Are you sure you finalised the ranking?

- ☐ Yes, I am
- ☐ No, I am not









Q59. Optional: Did we forget something?

Q60. Optional: Do you have feedback on the questions about the important issues to be addressed?

End of Block 10: Important Issues To Be Addressed

Start of Block 11: Obstacles to Implementation

Researchers also report various obstacles to reforming science. How much do you agree with the following statements?

	Not at all	Completely	Not applicable
Q61. "Open science does not sufficiently take into account privacy issues for studies with sensitive data."			
Q62. "Open sciences practices are too time-consuming."			
Q63. "At this moment, open science practices are not rewarded or incentivised enough."			
Q64. "Practicing open science gives me a competitive advantage over other scientists."			
Q65. "Practicing open science gives me a competitive disadvantage over other scientists."			
Q66. "The critique about my field of research from the reform movement makes me feel like I have to prove my innocence."			
Q67. "The tone of the members of the reform movement should be more nuanced."			
Q68. "I am less likely to engage with the propped reform practices because I feel the reform movement is prejudiced toward my field of research."			

Q69. Optional: Do you want to elaborate on any of your answers with regard to obstacles for reform?

Q70. Optional: What other obstacles for changing the practices of your field do you see?

End of Block 11: Obstacles to Implementation

Start of Block 12: Feedback

You've now reached the end of the survey.

Q71. Would you like to give more specific feedback on the survey?

Q72. I have honestly answered the questions above.

☐ Yes

☐ No

Q73. I paid attention filling in this survey.

☐ Yes

☐ No

Please press → to submit your answers. You cannot change your answers anymore after submitting.

End of Block 12: Feedback
