

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Agnoli, F;Fraser, H;Singleton Thorn, F;Fidler, F

Title:

Australian and Italian Psychologists' View of Replication

Date:

2021-01-01

Citation:

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), <https://doi.org/10.1177/25152459211039218>.

Persistent Link:

<https://hdl.handle.net/11343/289628>

License:

[CC BY NC](#)

Australian and Italian Psychologists' View of Replication



Franca Agnoli¹, Hannah Fraser², Felix Singleton Thorn³,
and Fiona Fidler⁴

¹Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy;

²School of Biological Sciences, University of Melbourne, Melbourne, Australia; ³School of Psychological Sciences, University of Melbourne, Melbourne, Australia; and ⁴School of Biological Sciences and School of Historical and Philosophical Studies, University of Melbourne, Melbourne, Australia

Advances in Methods and
Practices in Psychological Science
July-September 2021, Vol. 4, No. 3,
pp. 1–15
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/25152459211039218
www.psychologicalscience.org/AMPPS



Abstract

Solutions to the crisis in confidence in the psychological literature have been proposed in many recent articles, including increased publication of replication studies, a solution that requires engagement by the psychology research community. We surveyed Australian and Italian academic research psychologists about the meaning and role of replication in psychology. When asked what they consider to be a replication study, nearly all participants (98% of Australians and 96% of Italians) selected options that correspond to a direct replication. Only 14% of Australians and 8% of Italians selected any options that included changing the experimental method. Majorities of psychologists from both countries agreed that replications are very important, that more replications should be done, that more resources should be allocated to them, and that they should be published more often. Majorities of psychologists from both countries reported that they or their students sometimes or often replicate studies, yet they also reported having no replication studies published in the prior 5 years. When asked to estimate the percentage of published studies in psychology that are replications, both Australians (with a median estimate of 13%) and Italians (with a median estimate of 20%) substantially overestimated the actual rate. When asked what constitute the main obstacles to replications, difficulty publishing replications was the most frequently cited obstacle, coupled with the high value given to innovative or novel research and the low value given to replication studies.

Keywords

replication, conceptual, direct, reproducibility, open data, open materials

Received 6/11/20; Revision accepted 7/17/21

The recent crisis of confidence in psychological research (Chambers, 2017; Lilienfeld & Waldman, 2017; Pashler & Wagenmakers, 2012; Zwaan et al., 2018) is also called a *replication crisis* because it emerged from observing that many published findings were difficult or impossible to replicate. In psychology, as in other sciences, “replicability is almost universally accepted as the most important criterion of genuine scientific knowledge” (Rosenthal & Rosnow, 1991, p. 9). Despite the essential role of replicability in establishing the reliability of findings, remarkably few replication studies have been published in psychology and other social science journals (see Schmidt, 2009). Makel et al. (2012) searched 100 highly rated psychology journals over the years 1900 to 2012

and found that only 1.07% of published psychology articles reported replications of prior research. Narrowing the search to more recent research, Hardwicke et al. (2021) found that replications studies were rare (5%) in a random selection of psychology studies published between 2014 and 2017. Similar studies in other fields have also shown very low prevalence of published replication studies, including rates of only 0.13% in education research (Makel & Plucker, 2014) and 1.2% in marketing

Corresponding Author:

Franca Agnoli, Department of Developmental Psychology and Socialization, University of Padova
Email: franca.agnoli@unipd.it



Creative Commons NonCommercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits noncommercial use, reproduction, and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

research (Evanschitzky et al., 2007). These consistent findings of low prevalence of published replication studies have led to concern among some researchers about the reliability of the literature.

Concerned about the replication crisis and uncertain about its severity, researchers initiated systematic efforts to replicate samples of published studies. A 2014 special issue of *Social Psychology* published 15 preregistered replication studies (Nosek & Lakens, 2014), and a Bayesian analysis found that these replications provided little evidence for the effects reported in the original studies (Marsman et al., 2017). The Open Science Collaboration conducted replications of 100 studies published in three important psychology journals and found that the replications produced weak evidence for the original findings; only 36% of the replications obtained statistically significant results compared with 97% of the original studies (The Open Science Collaboration, 2015). The Center for Open Science coordinated an international team of researchers who attempted to replicate 28 classic and contemporary findings in psychology in high-powered preregistered studies. They found that only about half the replications yielded statistically significant results in the same direction as the original study (Klein et al., 2018).

These systematic replication studies are evidence that the crisis of confidence in psychology is justified, but these studies have also inspired debate about what constitutes a replication and what can be concluded from a successful or unsuccessful replication (Nosek et al., 2012). Some researchers observed that a replication that follows the same procedure and uses the same materials as the original study may still differ in essential ways (Crisp et al., 2014; Schwarz & Strack, 2014), and it is evident that an exact replication is impossible (Nosek & Errington, 2020).

The debate about the crisis of confidence has also produced concrete proposals (e.g., Chambers, 2017) for improving the quality of psychological research and confidence in published studies. Conducting and publishing replications of published research is among these proposals, as exemplified by systematic replication studies (Camerer et al., 2018; Klein et al., 2018; The Open Science Collaboration, 2015; Soto, 2019). A researcher encouraged to conduct a replication must decide how to conduct the study, that is, how faithfully to repeat the conditions of the original study. A replication study with the same population, methods, and statistical analyses (sometimes called a *direct replication*) is a means of assessing the reliability of the original results. Chambers (2017) asserted that direct replications are needed and would strengthen the foundations of psychological theories. A replication study that alters the population, methods, and/or statistical analyses (sometimes called a *conceptual replication*) is a means of assessing the

generalizability of the original results and may not be informative about the reliability of the original results. Encouragement to conduct replications is ambiguous without consideration of its purpose and the best method to achieve that purpose, and Machery (2020) observed that there has been little discussion in the psychological literature of what constitutes a replication or what purpose it serves.

We surveyed academic psychologists about the meaning of replications and the role that replications play in their research. The survey is largely descriptive and exploratory. It is an attempt to uncover how psychologists define and think about replication, including the role it plays in their own research. The survey questions themselves were adapted from a recent, similar survey of ecology researchers (Fraser et al., 2020), which itself was an attempt to build on much earlier work by sociologists of science, Mulkay and Gilbert (1991).

A central question is what kinds of studies participants consider to be replications. The three main components of a study are the population from which the sample is drawn, the experimental method (including design, procedure, and materials), and the statistical analyses. Some researchers might expect a replication to retain the same population, methods, and statistical analyses used in the original study, whereas others might expand the concept to include modifying one or more of these elements when replicating a study. Modifying the population or methods could be intended to assess the generalizability of the original study, and modifying the statistical analyses could be intended to leverage advances in analytic methods, for example. We posed the question as a choice among alternative descriptions of a replication study, inviting participants to choose as many descriptions as they thought appropriate.

Methodologists view replications as essential in determining the reliability of research findings, and we wanted to explore whether the practices of academic psychologists are consistent with this view. The survey asked whether “an effect or phenomenon needs to be successfully replicated before you believe or trust it.” Then the survey asked participants to describe in their own words what else they look for to determine believability or trustworthiness. The survey also asked whether researchers check whether an effect or phenomenon has been replicated, separately considering both plausible and implausible effects. If, indeed, replications are essential to determining reliability, then we should expect researchers would search for evidence of them.

We expected diverse views about what constitutes a replication study, and we were concerned that this diversity would be a substantial contributor to variability in the responses to our remaining questions about researchers’ opinions and practices. To reduce this

source of variability, we asked survey participants to consider only those replication studies that retained the same population, methods, and statistical analyses when completing the rest of the survey. Such a replication study is the best test of the reliability of the original study.

The survey asked participants to rate the importance of replication in psychology. If academic psychologists are aware of the replication crisis in psychology, concerned about it, and view replications as a way of addressing the crisis, then they should consider replications to be very important.

We explored the views of academic psychologists about the frequency of replications in the published literature in a series of related questions. Participants were asked (a) to estimate the percentage of published studies that are replications, (b) whether enough replication takes place in psychology, (c) whether replications are a good use of resources in psychology, and (d) how often replication studies should be published.

Finally, we focused on the replication practices of academic psychologists and their reasons for adopting these practices. The survey asked how often they or the students they supervise replicate studies. It also asked whether they had a replication study published in the past 5 years. Participants were asked to describe in their own words the main obstacles to replications. We anticipated that this question would elicit descriptions of systemic obstacles (e.g., Fanelli, 2010; Romero, 2017) that have suppressed publication of replications in psychology journals.

Our survey covered two broad groups of academic psychologists: those affiliated with Australian universities and with Italian universities. These two groups were selected because they are the academic communities of the authors and because their similarities and differences could provide interesting comparisons. Both groups include more than 1,000 psychologists at varied stages of academic career, and they constitute two different linguistic/cultural and historical academic settings. Psychology in Australia was originally closely linked to Britain but in the 1960s began a shift in interest in U.S. psychological research (Buchanan, 2012), and, of course, the primary language was always English. Psychology in Italy was more closely tied to European psychological research, and Italian psychologists still teach primarily in Italian. Today, however, their research publication output is primarily in English (because of a recent strong push toward international publications), and consequently, both groups are publishing in international English-language journals. Because both groups participate in the same international academic world, we anticipated that their views about replication would be similar and that any differences might be attributable to different national experiences.

Method

Data, materials, and online resources

Comprehensive materials, the code and data required to reproduce our results computationally, and deidentified qualitative responses are available at <https://osf.io/sdkm6/>.

Ethical approval

Ethical review and approval were obtained from the University of Melbourne Human Research Ethics Committee (Ethics ID 1749316.1) and from the University of Padova Ethics Committee for Psychological Research (Ethics ID 2635) to accommodate differences in the requirements of the different universities and countries.

Participants

The Australian survey population was all the psychology researchers at the 37 Australian universities that identify psychology departments on the university websites, excluding the authors' university. We extracted email addresses for all the research staff of these psychology departments, which yielded 1,442 email addresses of university psychology researchers working in Australia. No information was extracted about the academic rank of these researchers, but junior staff and postgraduate students are likely to be underrepresented in our sample because they are less likely to be represented on university websites.

The Italian survey population was identified by a website (www.miur.gov.it) maintained by the Ministry of Education, Universities and Research of the Italian government and included all 1,052 academic psychologists working full-time at Italian public and private universities in March 2018. This website included their names, positions, and affiliations. Email addresses were obtained by searching the websites of their affiliated universities. These psychologists were employed as researchers (equivalent to assistant professors in some other systems), associate professors, or full professors. The ministry's website described the population as including 340 researchers (113 male and 227 female), 452 associate professors (188 male and 264 female), and 260 full professors (134 male and 126 female). Each academic psychologist was also affiliated with one of the following eight scientific disciplines: general psychology, psychobiology and physiological psychology, psychometrics, developmental and educational psychology, social psychology, industrial and organizational psychology, dynamic psychology, and clinical psychology.

Survey instrument

Two versions of the survey were created, one in English and one in Italian (available at <https://osf.io/sdkm6/>). As mentioned earlier, the questions themselves were adapted from a similar survey of ecology researchers (Fraser et al., 2020). Two bilingual Italian-native speakers translated the English-language survey into Italian, and two bilingual English-native speakers back translated it to confirm that the content of the two surveys were as consistent as possible. The surveys were administered via Qualtrics (Provo, UT) in a fixed order and included multiple-choice questions about (a) the types of studies participants consider replications, (b) whether replication is necessary for an effect or phenomenon to be believed or trusted, (c) how often participants check whether an effect or phenomenon has been replicated if it seems plausible, (d) how often participants check whether an effect or phenomenon has been replicated if it seems implausible, (e) how important replication is in psychology, (f) whether there is enough replication taking place in psychology, (g) whether replication is a good use of resources in psychology, (h) how often replication studies should be published in psychology, (i) how often participants replicate studies, and (j) whether participants have had a replication study published in the past 5 years.

The survey also included a slider-bar numeric-response question asking participants to estimate the percentage of published studies in psychology that are replications and three free-text response questions about (a) what participants look for to determine believability or trustworthiness, (b) what participants consider the main obstacles to replication, and (c) other comments.

Procedure

In Australia, an email message was sent on October 18, 2018, to all 1,442 members of the Australian survey population inviting their participation in the survey, but 22 messages were returned as undeliverable. A second message was sent repeating the invitation 11 days later, and 25 of these messages were returned as undeliverable. These invitation letters included a link to the English-language survey. There were 244 (17%) visits to the survey, 215 participants (15%) who answered some questions, and 198 participants (14%) who finished the survey.

In Italy, an email message was sent in October 2018 to all 1,051 members (excluding one of the authors) of the Italian survey population inviting their participation in the survey, but 11 messages were returned as undeliverable. A second message was sent repeating the invitation about 2 weeks later, and 12 messages were returned as undeliverable. All invitation letters included a link to the Italian-language survey. Data collection was

terminated after a month. There were 293 (28%) visits to the survey website, 274 participants (26%) who answered some questions, and 237 participants (23%) who finished the survey.

Results

What is a replication study?

Participants were asked to identify what they considered to be a replication study in psychology by selecting as many options as they wanted among 10 descriptions. Table 1 lists these 10 descriptions in the order they were presented and the frequencies and percentages of Australian and Italian psychologists who selected each description. The most frequently selected description by both Australian (83%) and Italian (68%) psychologists was “Redoing an experiment with the same population, the same methods, and the same statistical analyses,” which describes a direct replication. The second most frequently selected description by Australian (65%) and Italian (54%) psychologists was “Redoing an experiment with a different population, leaving all else the same.” One or both of these two options was selected by 98% of Australian participants and 96% of Italian participants; 50% (Australian) and 26% (Italian) selected both options, 34% (Australian) and 41% (Italian) selected only the first option (the direct replication), and 15% (Australian) and 28% (Italian) selected only the second option. Those participants who selected only the second option apparently think that replication requires changing something, at least the population.

The third most frequently selected description by Australian (48%) and Italian (22%) psychologists was “Redoing an experiment with different statistical analyses, leaving all else the same.” All but three Australian participants and all but one Italian participant who chose this description also chose one or both of the preceding two descriptions.

Relatively few participants chose any of the four descriptions that included changing the methods; mean selection rates for these descriptions were 14% for Australians and 8% for Italians. In addition, relatively few chose either of the two descriptions that involved reanalyzing previously collected data; mean selection rates for these two descriptions were 11% for Australians and 15% for Italians.

As these comparisons indicate, Australian and Italian psychologists were similar in their selection of descriptions of a replication study. There were, however, two main differences worth noting. First, Australians were more likely than Italians to select descriptions that involve changing the statistical analyses. The description “Redoing an experiment with different statistical analyses,

Table 1. Alternative Descriptions of a Replication Study and the Percentages (With 95% Confidence Intervals) of Participants That Selected Them (Australian $N = 214$, Italian $N = 250$)

Description	Australian		Italian	
	<i>n</i>	Percentage (95% CI)	<i>n</i>	Percentage (95% CI)
Reanalyzing previously collected data with the same statistical analyses	28	13 (9, 18)	31	12 (9, 17)
Reanalyzing previously collected data with different statistical analyses	20	9 (6, 14)	42	17 (13, 22)
Redoing an experiment with the same population, the same methods, and the same statistical analyses	178	83 (78, 88)	169	68 (62, 73)
Redoing an experiment with a different population, leaving all else the same	138	65 (58, 71)	136	54 (48, 61)
Redoing an experiment with different methods, leaving all else the same	40	19 (14, 24)	30	12 (9, 17)
Redoing an experiment with different statistical analyses, leaving all else the same	102	48 (41, 54)	56	22 (18, 28)
Redoing an experiment changing the population and methods but using the same statistical analyses	28	13 (9, 18)	18	7 (5, 11)
Redoing an experiment changing the methods and statistical analyses but keeping the same population	24	11 (8, 16)	14	6 (3, 9)
Redoing an experiment changing the population and statistical analyses but using the same methods	81	38 (32, 45)	42	17 (13, 22)
Redoing an experiment changing all three of these (population, methods, and statistical analyses) at the same time	22	10 (7, 15)	13	5 (3, 9)

Note: Confidence intervals (CI) are Wilson score intervals with Yates's continuity correction for binomial proportions (Newcombe, 1998; Wallis, 2013).

leaving all else the same" was selected by 48% of Australians but only 22% of Italians. Likewise, the description "Redoing an experiment changing the population and statistical analyses, but using the same methods" was selected by 38% of Australians but only 17% of Italians.

Second, Australian participants selected more options than Italian participants. As shown in Figure 1, 25% of Australian participants chose only one option, 22% chose two options, and 22% chose three options, whereas 41% of Italian participants chose only one option, 28% chose two options, and 14% chose three options. Only 3% of Italian participants and 10% of Australian participants chose six or more options. Possibly this indicates that the concept of replication has a more restricted meaning to Italian psychologists.

Is replication required for belief or trust?

Participants were asked whether "an effect or phenomenon needs to be successfully replicated before you believe or trust it." Figure 2 presents the percentages of "yes," "maybe," and "no" responses. Note that 19% of Italian psychologists (compared with only 8% of Australian psychologists) said that their belief or trust in an effect does not depend on the existence of a replication, in contrast to claims that replicability is universally

regarded as the most important criterion of scientific knowledge (Rosenthal & Rosnow, 1991).

What else do you look for to determine believability or trustworthiness? We asked participants to describe in a free-text format what else they look for to determine believability or trustworthiness; 193 Australian and 193 Italian participants responded to this question. The Australian responses ranged from one to 477 words in length, with a median length of 14 words, and the Italian responses ranged from one to 73 words in length, with a median length of eight words.

We analyzed these texts using the open-coding method of grounded theory, a qualitative method for identifying content categories (Corbin & Strauss, 2015). No categories were defined a priori. Instead, categories were identified by analysts who read the text responses and created categories that describe the content of the text. The analysts iteratively refined and adjusted the categories to capture the topics of all the text entries. A native English-speaking analyst studied the Australian responses and a native Italian-speaking analyst studied the Italian responses and generated categories that summarized the topics addressed by the participants. These analysts compared and integrated the English- and Italian-language categories, which resulted in 55 detailed categories of content that fall within nine higher level categories (e.g., the two detailed categories "Clarity of

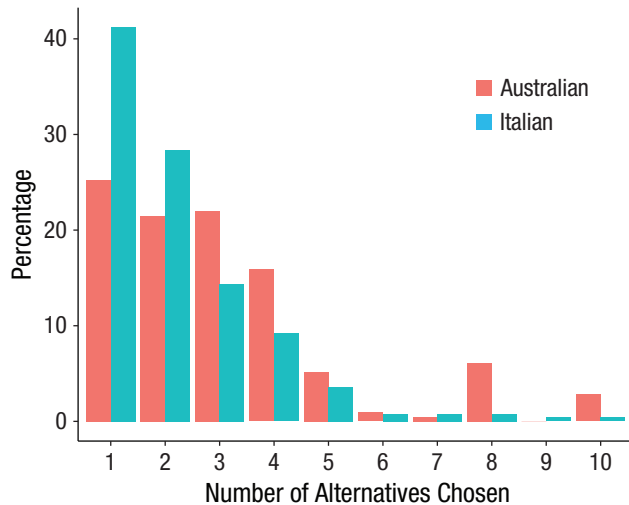


Fig. 1. Distribution of the number of descriptions selected (Australian $N = 214$, Italian $N = 250$).

hypotheses” and “Clarity of results presentation” both fall within the higher level category “Clarity and completeness”). After these categories were defined, two native English-speaking analysts independently sorted all 193 Australian text entries, and two native Italian-speaking analysts independently sorted the 193 Italian text entries into one or more of these 55 detailed categories. There were seven Australian responses and 11 Italian responses that the analysts could not categorize because they were incomplete, incoherent, or nonresponsive statements (e.g., “I don’t know”), which left 186 Australian and 182 Italian analyzed responses. The supplemental material available online includes all the text responses and the higher level codes assigned by both analysts for both countries. Cohen’s κ and observed

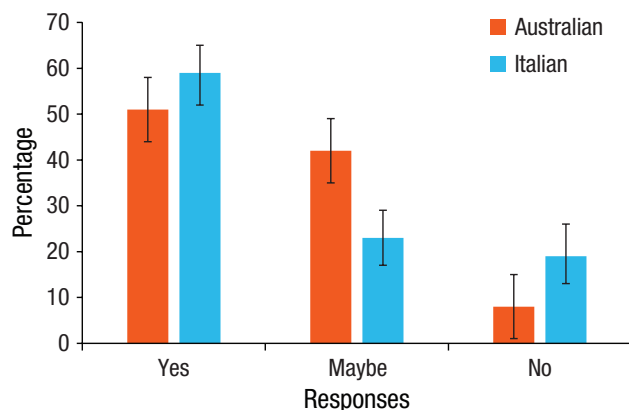


Fig. 2. Percentages and 95% confidence intervals of responses to the question, “Does an effect or phenomenon need to be successfully replicated before you believe or trust it?” (Australian $N = 214$, Italian $N = 253$). Confidence intervals were calculated using the method for multinomial proportions of Sison and Glaz (1995).

percentage interrater agreement were computed to assess interrater reliability; interrater agreement was 91% or greater for the Australian ratings of all categories and 89% or greater for the Italian ratings of all categories. Table 2 lists the nine categories, the number of instances of each category, and the two measures of interrater reliability.

Note that the nine categories are simply topics addressed in the text responses and do not capture what, specifically, the respondents wrote about them. We illustrate the ideas that were expressed with some representative examples.

Australian psychologists’ statements about information they seek to determine believability or trustworthiness most frequently fell into four categories: research design and method (52%), statistical analysis (47%), theory-data linkage (44%), and sample and population (35%). In the category of research design and method, one respondent wrote, “If a method is sound I am willing to believe the data until I have evidence otherwise.” In the statistical analysis category, a respondent wrote, “Appropriate statistical analysis,” which is similar to many of the statements included in this category. In the theory-data linkage category, one respondent wrote, “hypotheses founded on established theory, fit of method with data and fit of method with hypotheses.” In the sample and population category, most respondents simply noted “sample size.”

Surprisingly, few Australians responded that they look for replications (15%). Those who did, however, expressed a range of opinions about the role of replication (whether direct or conceptual). For example, one respondent wrote,

The more unlikely the finding, the stronger the evidence in favour of it would need to be convincing. I would actually find a so-called “conceptual replication” more convincing and more valuable than a direct replication. Direct replication should in theory demonstrate that the effect is not a Type 1 error, but add nothing else. . . . Whereas a conceptual replication, where a lot of things about the study have changed, including the entire paradigm even, are far more valuable and convincing, than a direct replication.

Other Australian respondents questioned the value of replication in psychology, as in this quote,

It seems to me that replication in the psychology research genre is of questionable value. It has some but quite limited value. For example, in question no. one, several items refer to changing the population but what does that mean? What is mean

Table 2. Categories of Text Responses to the Question, "Please Describe What Else You Look for to Determine Believability or Trustworthiness?" (Australian $N = 186$, Italian $N = 182$)

Category	Australian				Italian			
	n	Percentage	κ	IA (%)	n	Percentage	κ	IA (%)
Clarity and completeness	9	5	.52	95	11	6	.86	98
Transparency	27	15	.71	95	5	3	.33	98
Research design and method	97	52	.82	91	81	45	.81	91
Sample and population	66	35	.89	96	70	38	.93	97
Statistical analysis	87	47	.85	93	77	42	.88	94
Theory-data linkage	82	44	.80	91	59	32	.74	89
Replication	28	15	.65	94	14	8	.68	96
Reputation	39	21	.91	98	13	7	.92	99
Preregistration and open access	19	10	.91	99	7	4	.93	99

Note: IA = interrater agreement.

by the SAME population? If one is doing a psychological study of the acculturation of Indian immigrants, no population could be exactly the same given the diversity of the Indian Australian population.

In striking contrast to this opinion, another respondent wrote,

I should clarify that I don't think an effect should be replicated once but multiple times. I think an effect should first be replicated with the same methods and populations and, ideally, then also replicated across populations and with modifications to the methods that are consistent with the theoretical mechanism hypothesized to underpin the effect. Over multiple such replications, an effect size may be then estimated so it is possible to assert whether it is present and what its magnitude is.

Many Australian psychologists (21%) indicated that the reputation of the researcher and the prestige of the journal play a role in their decision about the believability of an effect. For example, one respondent wrote, "The status of the investigators and the journal in which the study was published." Statements such as this example can be viewed as problematic. The exaggerated value attributed to publications in journals with a high impact factor may encourage submitting incorrect or fraudulent results (Craig et al., 2020). Indeed, journals with high impact factors in biomedical research have had elevated retraction rates (Fang & Casadevall, 2011). Some respondents, however, expressed an opposite view, such as, "If it is in PNAS or Psych Science with p values $<.005$, I assume it is not believable."

Note that only 15% of Australians mentioned transparency and that only 10% mentioned preregistration and open access of materials, data, and statistical analyses.

Like Australian psychologists, the statements of Italian psychologists most frequently fell into the same four categories: research design and method (45%), statistical analysis (42%), sample and population (38%), and theory-data linkage (32%). Most of the responses in the research design and method category simply cited these topics (e.g., "the experimental paradigm"). Responses about statistical analysis were primarily about effect size, power, or simply statistical analysis in general. Responses about the sample and population were almost all about sample size. Responses about the theory-data linkage were mostly about validity and the theory.

Relatively few Italian respondents (8%) indicated that they seek information about replications when deciding whether to believe or trust an effect. When mentioned, respondents generally referred to the need for conceptual replications. For example, one respondent wrote, "demonstration of the effect in diverse versions of the same paradigm [translated]." One participant noticed that replications are more difficult to do in neuroscience compared with psychology (because of costs), but when they were done, the results with functional MRI did not replicate earlier findings (seemingly suggesting fraud on the part of the researchers).

Few Italian psychologists (7%) reported seeking information about the reputation of researchers or prestige of the journal, and even fewer (4%) reported an interest in preregistration or open access. These results (and others presented later) may suggest that discussions about positive responses for dealing with the crisis in psychology have not been as widespread or as prominent among Italian academic psychologists as among Australians.

Table 3. Responses to the Questions, “If an Effect or Phenomenon Seems Plausible/Implausible to You, How Often Do You Check if It Has Been Replicated?” (Australian Plausible $N = 208$, Italian Plausible $N = 246$, Australian Implausible $N = 209$, Italian Implausible $N = 244$)

Response	Plausible effect				Implausible effect			
	Australian		Italian		Australian		Italian	
	<i>n</i>	Percentage (95% CI)	<i>n</i>	Percentage (95% CI)	<i>n</i>	Percentage (95% CI)	<i>n</i>	Percentage (95% CI)
Almost always	43	21 (14, 28)	72	29 (23, 36)	80	38 (31, 46)	93	38 (32, 45)
Most of the time	62	30 (23, 37)	65	26 (20, 33)	69	33 (26, 40)	60	25 (18, 32)
Sometimes	94	45 (39, 53)	87	35 (29, 42)	52	25 (18, 32)	74	30 (24, 37)
Never	9	4 (0, 12)	22	9 (2, 16)	8	4 (0, 11)	17	7 (0, 14)

Note: CI = confidence intervals.

Do you check for replications? In two separate questions, we asked how often participants checked whether a plausible effect or phenomenon had been replicated and whether an implausible effect or phenomenon had been replicated. Table 3 shows the frequencies and percentages of participants from both countries who chose almost always, most of the time, sometimes, and never for both questions. Australian and Italian responses were similar; the most frequent response for both populations was sometimes when an effect was plausible and almost always when an effect or phenomenon was implausible.

Consider replications only with the same population, methods, and statistical analyses

In the preceding questions, participants were given no guidance about what constitutes a replication or what kind of replication to consider when responding. For the remaining questions in the survey, participants were instructed,

Please just think about the type of replication study where an experiment is repeated with the same population, the same methods (e.g., stimuli, measures, and procedure) and the same statistical analyses, regardless of whether or not it finds the same result as the original study.

How important is replication in psychology?

Participants were asked to rate the importance of replication in psychology. Figure 3 shows the percentages of respondents who chose very important, somewhat important, not important, and no opinion. The results indicate that participants recognize the importance of replications in psychology. The majority of Australians

(76%) and Italians (59%) chose very important, and only 3% of Australians and 5% of Italians chose not important or no opinion.

What percentage of published studies are replications? We asked participants to estimate the percentage of published studies in psychology that are replications, which they did by manipulating a slider (initially set at the value 100) to select any value between 0 and 100. The mean and median of 205 Australian estimates were 18% (95% confidence interval [CI] = [16%, 21%]) and 13%, respectively. The mean and median of 236 Italian estimates were 29% (95% CI = [26%, 32%]) and 20%, respectively. As the differences in these estimates of central tendency suggest, the distributions of estimates (shown in Fig. 4) are strongly skewed to the right (positively skewed). Both Australian and Italian psychologists substantially overestimated the 1.07% prevalence of replication studies reported by Makel et al. (2012) and the 5% prevalence found by Hardwicke et al. (2021). The Italian estimates were even further than the Australian estimates from the replication rates found in the literature.

Is there enough replication in psychology? Participants were asked to judge whether enough replication takes place in psychology using a 5-point scale. The results shown in Table 4 indicate that nearly all participants believe that more or much more replications should be done.

We examined the relationship between judgments about the need for more replications and estimates (shown in Fig. 4) of the percentages of published studies that are replications. Because relatively few people responded that there should be less or much less replication in psychology, we considered only the first three response options in Table 4. The mean estimates of the percentages of replications for participants who said there should be much more replication was 13% for Australians and 22% for Italians. These mean estimates

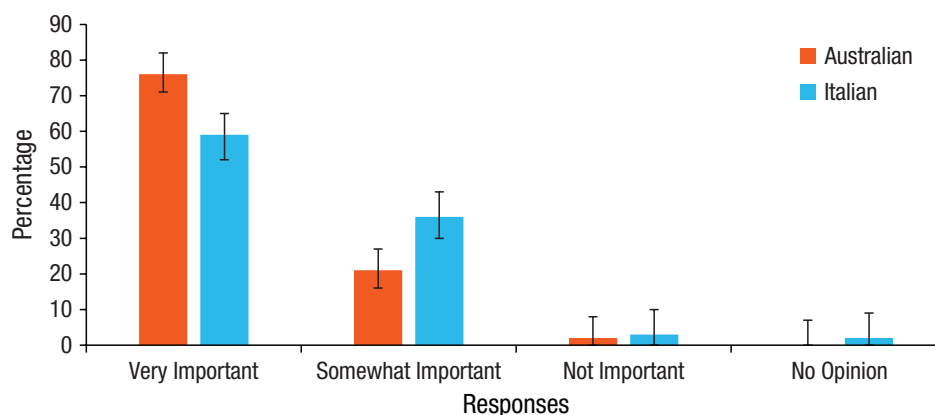


Fig. 3. Percentages and 95% confidence intervals of responses to the question, "How important would you say replication is in psychology?" (Australian $N = 210$, Italian $N = 241$).

increased for participants who said there should be more replications to 22% for Australians and 28% for Italians. The mean estimates increased again for participants who said about the right amount of replication takes place to 29% for Australians and 44% for Italians. In both nations, the greater the need for more replications, the lower the estimates of replications in the literature.

Are replications a good use of resources in psychology? We asked participants whether replication studies are a good use of resources in psychology. A majority of both Australian (67%) and Italian (51%) psychologists considered replications to be crucial, as shown in Table 5. Note that a much larger percentage of Italian psychologists (19%) responded, "I don't know." This response could

indicate that they are not well informed about the need for resources to be allocated to replication studies or that they are undecided among the response options offered.

How often should replication studies be published?

We asked participants how often replication studies should be published in psychology. A large majority of both Australian and Italian psychologists responded that replication studies should be published more often in all journals, as shown in Table 6.

How often do you replicate studies? We asked participants how often they or the students they supervise replicate studies. As shown in Figure 5, about half the Australian (52%) and Italian (45%) psychologists reported that they

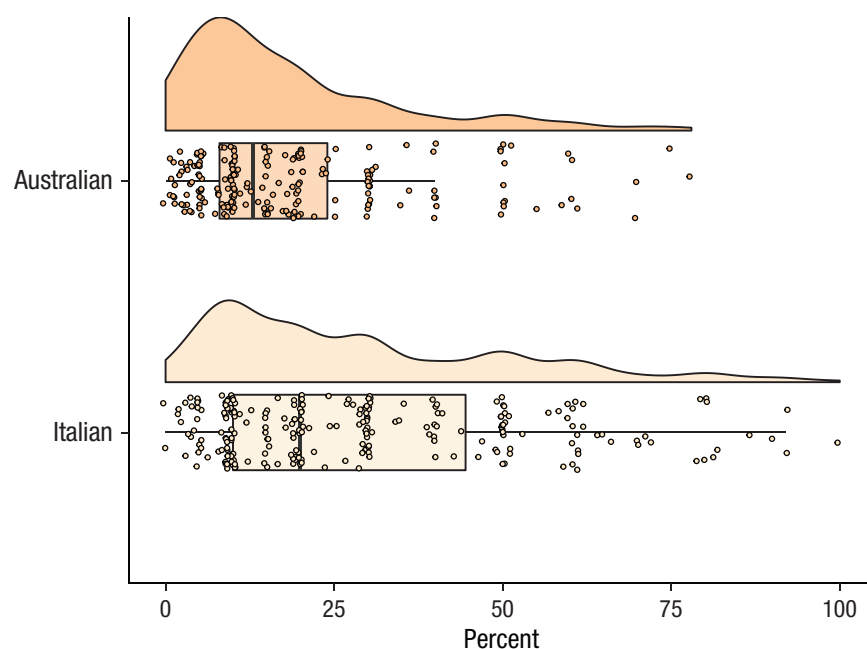


Fig. 4. Frequency distributions and box plots (with jitter) of estimated percentages of published studies that are replications (Australian $N = 205$, Italian $N = 236$).

Table 4. Responses to the Question, “Do You Think Enough Replication Takes Place in Psychology?” (Australian $N = 208$, Italian $N = 239$)

Response	Australian		Italian	
	<i>n</i>	Percentage (95% CI)	<i>n</i>	Percentage (95% CI)
There should be much more	96	46 (39, 54)	95	40 (34, 47)
There should be more	96	46 (39, 54)	108	45 (39, 52)
About the right amount takes place	11	5 (0, 13)	25	10 (4, 18)
There should be less	5	2 (0, 10)	7	3 (0, 10)
There should be much less	0	0 (0, 7)	4	2 (0, 9)

Note: CI = confidence intervals.

or their students sometimes replicate studies, and a smaller fraction of Australian (13%) and Italian (10%) psychologists reported that they or their students often replicate studies. It is likely that many of these replication studies were undergraduate student projects supervised by the respondents.

Have you had a replication study published recently?

When asked, “Have you had a replication study published in the past 5 years?” a large majority of Australians (77% of 208 respondents) and Italians (78% of 237 respondents) indicated that they had not published a replication study. Note that a majority of psychologists in both countries have often or sometimes replicated studies (see Fig. 5), but relatively few have published a replication in the past 5 years. Unfortunately, we do not know whether the low rate of published replications is due to researchers trying and failing to publish replications or deciding not to try to publish them.

What are the obstacles to replications? We asked participants to describe in a free-text format what they consider to be the main obstacles to replications; 194 Australian and 178 Italian participants responded to this question. The Australian responses ranged from one to 355 words in length, with a median length of 21 words. The Italians generally gave shorter responses, which ranged from one

to 177 words in length, with a median length of eight words.

As in the previous free-text-format question, we analyzed these texts using the open-coding method of grounded theory (Corbin & Strauss, 2015). A native English-speaking analyst and a native Italian-speaking analyst studied the responses of the Australian and Italian participants, respectively, and identified categories of content. The analysts compared and integrated the English- and Italian-language categories, which resulted in 26 detailed categories that fall into the seven higher level categories shown in Table 7. Two native English-speaking analysts sorted all the Australian responses into one or more categories, and two native Italian-speaking analysts sorted all the Italian responses. There were five Australian responses and nine Italian responses that the analysts could not categorize because they were incomplete or incoherent (e.g., “the illusion of replicability”), which left 189 Australian and 169 Italian analyzed responses. Table 7 shows the number of responses in each of the seven categories and two measures of interrater reliability.

The most frequently identified obstacle, reported in 66% of Australian responses and 67% of Italian responses, was the difficulty of publishing replications or other issues related to publishing (e.g., an Australian researcher wrote, “The lack of interests from journals”). In a similar

Table 5. Responses to the Question, “Do You Consider Replication Studies to Be a Good Use of Resources in Psychology?” (Australian $N = 209$, Italian $N = 237$)

Response	Australian		Italian	
	<i>n</i>	Percentage (95% CI)	<i>n</i>	Percentage (95% CI)
Replication studies are crucial, and more of the available resources should go toward funding them	139	67 (60, 73)	121	51 (45, 58)
Replication studies are good, but given limited available resources, we should continue to focus on novel research	61	29 (23, 36)	58	24 (18, 31)
Replication studies are a poor use of available resources	5	2 (0, 9)	13	6 (0, 12)
I don't know	4	2 (0, 9)	45	19 (13, 26)

Note: CI = confidence intervals.

Table 6. Responses to the Question, “How Often Should Replication Studies Be Published in Psychology?” (Australian $N = 209$, Italian $N = 237$)

Response	Australian		Italian	
	<i>n</i>	Percentage (95% CI)	<i>n</i>	Percentage (95% CI)
More often, in all journals, including top-tier journals	143	68 (62, 75)	148	62 (57, 69)
More often, but only in special sections/issues/dedicated journals	53	25 (19, 32)	55	23 (17, 30)
They should only be published if they demonstrate different results to the original study	8	4 (0, 10)	22	9 (3, 16)
We should never publish replications	0	0 (0, 6)	2	1 (0, 7)
I don't know	5	2 (0, 9)	10	4 (0, 11)

Note: CI = confidence intervals.

vein, 28% of Australian responses and 17% of Italian responses said that replications were not highly valued. For example, an Italian researcher wrote, “The low value that a study of this type receives [translated].” In contrast, 35% of Australian responses and 24% of Italian responses stated that the high value placed on novelty was a main obstacle to replication (see Romero, 2017). For example, an Australian researcher wrote, “Research time is precious. I would like [to] contribute to advancing theory than replicate past research.” Somewhat more explicitly, an Italian researcher wrote, “They have low innovation value and therefore are difficult to place [translated].”

In line with the difficulty publishing and the low value attributed to replications, 46% of Australians noted the difficulty obtaining funding for replication studies (e.g., an Australian wrote, “Obtaining funding for replication when the push is for novel research”). Surprisingly, only 12% of Italian responses mentioned the challenge of obtaining funding for replications (e.g., “There is very little public funding [translated]”). This low Italian rate may be due to the difficulty of obtaining research funds for any studies; according to the OECD (2019), Australia has consistently had substantially higher spending on research

and development as a percentage of GDP since 1992, with a rate of 1.88% compared with 1.34% in Italy in 2015.

Some responses (Australian 12%, Italian 7%) also noted that conducting replications can carry some risk to the researchers, noting both opportunity loss (e.g., an Australian wrote, “Lack of prestige, cannot build a reputation, career, get funding by replication,” and an Italian wrote, “The attitude of the editors and the criteria for career advancement [translated]”) and the risk of upsetting senior researchers (e.g., an Australian wrote, “Boring at best. At worst, other scientists get defensive and mad at you”).

Some participants (11% of Australian and 21% of Italian responses) also noted that replication studies are difficult, and many cited the difficulty of recreating the elements of the original study (e.g., an Australian wrote, “replicating the experimenter and laboratory conditions, including treatment of the participants”).

Several Australian (16%) and a few Italian (4%) participants also suggested ways of overcoming the obstacles to replication, such as cooperating with the original authors when conducting a replication or conducting replications with extensions (e.g., an Australian wrote,

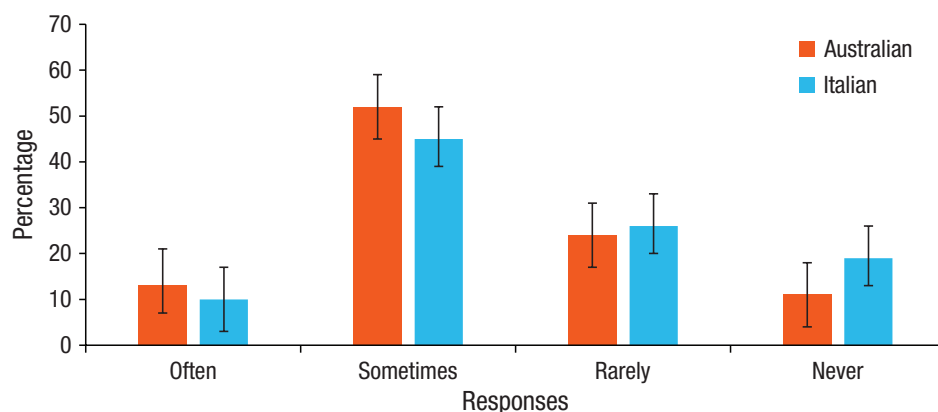
**Fig. 5.** Percentages and 95% confidence intervals of responses to the question, “How often do you, or the students you supervise, replicate studies?” (Australian $N = 208$, Italian $N = 235$).

Table 7. Categories of Text Responses to the Question “In Your Opinion What Are the Main Obstacles to Replication?” (Australian $N = 189$, Italian $N = 169$)

Category	Australian				Italian			
	n	Percentage	κ	IA (%)	n	Percentage	κ	IA (%)
Publishing	124	66	.75	87	113	67	.84	93
Resourcing	86	46	.75	89	20	12	.94	99
High value placed on novelty	66	35	.70	88	40	24	.77	92
Value of replications	52	28	.51	84	28	17	.52	88
Risks	23	12	.64	95	11	7	.77	98
Replication study difficulty	20	11	.78	97	36	21	.86	95
Solutions offered	30	16	.74	95	7	4	.60	97

Note: IA = interrater agreement.

“I prefer systematic replications that repeat the central effects but extend them by testing for moderation or mediation. This seems to advance theory/science as well as provide some confidence in the original effects”).

Discussion

A central question of this research is what participants consider to be replications. When asked what they consider to be a replication study, nearly all our survey participants said a replication was redoing an experiment keeping the same methods, the same statistical analyses, and either the same or a different population. Some participants accepted that a replication study might adopt a different statistical analysis. Few participants, however, considered a study that changes the experimental method to be a replication. Although a few participants wrote comments endorsing conceptual replications that modify the population and/or methods, this approach is apparently not widely accepted among Australian and Italian academic psychologists. Instead, they consider a replication to be an assessment of the reliability of the original findings, possibly an assessment of generalizability to other populations, but not an assessment of generalizability to other experimental methods.

Direct replication is commonly understood to be the repetition of an experimental procedure, whereas conceptual replication is a test of a finding from prior research using different methods (Schmidt, 2017). Recently, new ways of thinking about replications have proposed discarding the idea of conceptual replications while still recognizing that replications may modify attributes of a prior experiment. Nosek and Errington (2020) proposed considering any study a replication if its outcome would be considered diagnostic evidence about a claim from prior research, essentially using a new study to confront a prior claim with new evidence. A key requirement is that outcomes inconsistent with the prior claim must decrease confidence in the claim and

outcomes consistent with the prior claim must increase confidence in the claim. Studies may be conducted that explore whether a finding generalizes beyond the conditions of the original study, but these are considered generalizability tests, not conceptual replications, because an outcome inconsistent with the original finding may not decrease confidence in that finding.

Although methodologists view replications as essential in determining the reliability of research findings, only about half our participants responded that an effect must be replicated before they would believe it. Instead, most participants report that they judge reliability by considering the research design and method, statistical analyses, theory-data linkage, and the sample size and population. Few participants mentioned pre-registration or open access, which suggests that these practices have not yet had their intended effect of increasing trust in research (Washburn et al., 2018). This is not surprising given that Hardwicke et al. (2021) found that in a random sample of psychology studies published between 2014 and 2017, only 2% had shared data, and 3% were preregistered. A larger percentage responded that their assessment of the believability of research is influenced by the reputation of the journal and the investigators, which is worrisome considering the high rates of problems that have been found in prestigious journals (Nuijten et al., 2016; Veldkamp et al., 2014). Note that Alister et al. (2021) found that psychological scientists use a similar set of attributes to assess research replicability.

Both Australian and Italian psychologists report that they sometimes check for replications and that they are a little more likely to check if the result seems implausible. Checking for replications appears to be a standard practice for only about a third of the participants, who said they almost always check. Perhaps other participants are discouraged from searching for replications because they are not easy to find and not often published (Hardwicke et al., 2021; Makel et al., 2012).

For the remaining questions, we asked participants to consider only those replications that repeat the same population, methods, and statistical analyses. This is, indeed, the most frequently selected option when they were asked what they consider to be a replication study. We cannot be certain, of course, that participants remembered this instruction as they worked through the remaining questions.

Both Australian and Italian research psychologists substantially overestimated the percentage of published studies that are replications. Their estimates of the published replication rate were about 3 to 4 times the 5% rate observed by Hardwicke et al. (2021) in publications dating from 2014 to 2017. Although Australian and Italian psychologists overestimate the percentage of published replication studies, nearly all consider replication studies to be somewhat or very important, and nearly all think there should be more or much more replication. A majority of Australian and Italian psychologists also think that these studies are crucial and that more resources should be allocated to them.

These responses suggest that both Australian and Italian psychologists view replication studies as undervalued (Earp & Trafimow, 2015; Higginson & Munafò, 2016), and indeed nearly all psychologists from both nations responded that replication studies should be published more often, either in all journals or in special journals. A majority of psychologists responded that they or their students sometimes or often replicate studies themselves. They are not, however, successful at publishing these studies; about 77% reported that they have not published a replication study in the past 5 years.

In their own judgment, the main obstacles to conducting replication studies are the difficulty publishing them and the low value attributed to these studies. The difficulty publishing replications was the most frequently cited obstacle, coupled with the high value given to innovative or novel research and the low value given to replication studies. Romero (2017) observed that this value system is problematic because innovative results may not be valid (as evidenced by the low success rates of systematic replications), and we cannot know whether innovative results are valid if replication studies are not published. Although a majority of Australian and Italian psychologists asserted that replication studies are crucial and should be published more often, they believe that the publication process devalues them. Indeed, Martin and Clarke (2017) found that only 3% of psychology journals encourage submission of replications. Australian and Italian psychologists also noted that it is difficult to get funding for replication studies. This collection of obstacles reflects the implicit and explicit reward system of the scientific enterprise (e.g., De Boeck & Jeon, 2018).

Our results tell a story of researchers who value replication but who are not incentivized—through funding

or publishing opportunities—to engage in replication studies. And they perceived other obstacles. Our participants observed that failures to replicate may lead to conflict with the original investigators, which can be unpleasant and potentially harmful to a psychologist's career. They acknowledged that it may be difficult to obtain exactly the same population and/or exactly repeat the same methods and statistical analyses. An apparent failure to replicate may later be blamed on some unexpected detail of the method that had changed or the failure to follow strict procedures. This is an important recognition that the obstacles to replication are not only institutional (e.g., failure to fund or publish replication studies) but also cultural (e.g., not wanting to challenge or offend other researchers; see Nosek et al., 2012).

Where did the disconnect between the perceived value of replication and the incentives for its practice come from? Perhaps it is partly driven by a lack of understanding of the role different types of replication studies play in practice. Most participants in our sample had a working definition of replication that was procedural and fit a direct replication profile. Direct replications are, in our view, both extremely important (they control for sampling error, questionable research practices, mistakes, and fraud), *and* they have limited feasibility (there exist many situations in which they are impractical or impossible; see Nosek et al., 2012). Facing the reality of the difficulties in carrying out a direct replication may unwittingly impinge on the recognition of their importance. In other words, as researchers struggle with the fact that direct replications are not always possible, it becomes difficult to simultaneously hold the ideal that they have great epistemic value. This perhaps underlies participants' responses to how replication relates to believability and trustworthiness. Instead of relying on replication studies, they have developed substitutes for replication—both good and bad—to fulfill the perceived epistemic function of replication. Some of these substitutes may do part of the job of some types of replication, whereas others do very little and may even mislead.

Advocating for resources and rewards to support replication efforts—most notably, funding and publication of replication studies—remains very important. For researchers in our study, this was an almost universally endorsed refrain. Beyond this, we highlight the need for a cultural shift as well, in addition to an investment in infrastructure. In our study, obstacles to replication often included issues related to reputational damage (e.g., that one would be perceived as stepping on others' toes by engaging in replication studies and that career damage would ensue).

Efforts to encourage more replications studies may benefit from considering how Australian and Italian research psychologists think about replications and the obstacles that discourage conducting and publishing

them. Their dominant view is that a replication closely copies the original population, methods, and statistical analyses, which is consistent with the approach adopted by large systematic replication studies. They are discouraged from conducting and publishing replications because of the high value placed on novelty, the difficulty of adequately mimicking the conditions of the original study, and the systemic resistance to publishing replications. Some have tried to finesse this system by performing conceptual replications, but as Nosek and Errington (2020) explained, these studies should usually be viewed as generalizability tests, not as replications. Indeed, both generalizability tests and replications can provide useful information about the validity and limitations of research findings, and both should be encouraged.

We note that the findings we have discussed here are applicable to both the Australian and Italian participants. The similarities in their responses are much more striking than the differences, but there were some areas of difference. Italian psychologists appeared to have a more restrictive concept of replication, rejecting changes from the original study, but were also less likely to base their trust in an effect on the existence of a replication. Possibly the similarities in responses are due to the participation of researchers in both countries in the international English-language scientific publishing system and to similarities in the reward systems in both countries based on successful publishing in this system.

The results we report are from the 15% of Australian researchers and 26% of Italian researchers who answered at least some of the survey questions. Note that the research psychologists in our sample may not be representative of these populations. We would, however, expect that researchers who are more engaged in research activities would be more interested in research practices and more likely to participate in the survey.

Transparency

Action Editor: Brent Donnellan

Editor: Daniel J. Simons

Author Contributions

F. Agnoli, H. Fraser, and F. Fidler designed the survey. F. Agnoli, H. Fraser, and F. S. Thorn conducted the survey and analyzed the survey data. F. Agnoli wrote an initial draft of the manuscript. All four authors reviewed and revised the manuscript, and all four approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

F. Agnoli was supported by a Bilateral Agreement between the University of Padova and the University of Melbourne. F. Fidler was supported by Australian Research Council Future

Fellowship FT1501100297. The funders had no role in the decision to publish or preparation of the manuscript.

Open Practices

Open Data: <https://osf.io/bqm7s>

Open Materials: <https://osf.io/bqm7s>

Preregistration: not applicable

All data and materials have been made publicly available via OSF and can be accessed at <https://osf.io/bqm7s>. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Franca Agnoli  <https://orcid.org/0000-0001-7940-4080>

Hannah Fraser  <https://orcid.org/0000-0003-2443-4463>

References

- Alister, M., Vickers-Jones, R., Sewell, D. K., & Ballard, T. (2021). How do we choose our giants? Perceptions of replicability in psychological science. *Advances in Methods and Practices in Psychological Science*, 4(2). <https://doi.org/10.1177/25152459211018199>
- Buchanan, R. D. (2012). Australia. In D. B. Baker (Ed.), *The Oxford handbook of the history of psychology: Global perspectives* (pp. 18–33). Oxford University Press.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., & Wu, H. (2018). Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644.
- Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press.
- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). SAGE.
- Craig, R., Cox, A., Tourish, D., & Thorpe, A. (2020). Using retracted journal articles in psychology to understand research misconduct in the social sciences: What is to be done? *Research Policy*, 49(4), Article 103930. <https://doi.org/10.1016/j.respol.2020.103930>
- Crisp, R. J., Miles, E., & Husnu, S. (2014). Support for the replicability of imagined contact effects. *Social Psychology*, 45, 303–304. <https://doi.org/10.1027/1864-9335/a000202>
- De Boeck, P., & Jeon, M. (2018). Perceived crisis and reforms: Issues, explanations, and remedies. *Psychological Bulletin*, 144(7), 757–777.
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, Article 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.

- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLOS ONE*, 5(4), Article e10271. <https://doi.org/10.1371/journal.pone.0010271>
- Fang, F. C., & Casadevall, A. (2011). Retracted science and the retraction index. *Infection and Immunity*, 79, 3855–3859.
- Fraser, H., Barnett, A., Parker, T., & Fidler, F. (2020). The role of replication studies in ecology. *Ecology and Evolution*, 10(12), 5197–5207.
- 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*. Advance online publication. <https://doi.org/10.1177/1745691620979806>
- Higginson, A. D., & Munafò, M. R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLOS Biology*, 14(11), Article e2000995. <https://doi.org/10.1371/journal.pbio.2000995>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., & Sowden, W. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490.
- Lilienfeld, S. O., & Waldman, I. D. (Eds.). (2017). *Psychological science under scrutiny: Recent challenges and proposed solutions*. John Wiley & Sons.
- Machery, E. (2020). What is a replication? *Philosophy of Science*, 87. <https://doi.org/10.1086/709701>
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43, 304–316.
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7, 537–542.
- Marsman, M., Schoenbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., & Wagenmakers, E. J. (2017). A Bayesian bird's eye view of 'Replications of important results in social psychology.' *Royal Society Open Science*, 4(1), Article 160426. <https://doi.org/10.1098/rsos.160426>
- Martin, G. N., & Clarke, R. M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology*, 8, Article 523. <https://doi.org/10.3389/fpsyg.2017.00523>
- Mulkay, M., & Gilbert, N. (1991). Replication and mere replication (1986). In M. Mulkay (Ed.), *Sociology of science. A sociological pilgrimage* (pp. 154–166). Open University Press.
- Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, 17, 857–872. <https://doi.org/10.1002/1097-0258>
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLOS Biology*, 18, Article e3000691. <https://doi.org/10.1371/journal.pbio.3000691>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Almenberg, A. D., . . . Vazire, S. (2012). *Replicability, robustness, and reproducibility in psychological science*. PsyArXiv. <https://doi.org/10.31234/osf.io/ksfvq>
- Nosek, B. A., & Lakens, D. (2014). A method to increase the credibility of published results. *Social Psychology*, 45(3), 137–141.
- Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, 48, 1205–1226.
- OECD. (2019). *Gross domestic spending on R&D*. <https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm>
- The Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), Article aac4716. <https://doi.org/10.1126/science.aac4716>
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528–530.
- Romero, F. (2017). Novelty versus replicability: Virtues and vices in the reward system of science. *Philosophy of Science*, 84(5), 1031–1043.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (Vol. 2). McGraw-Hill.
- 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, 90–100.
- Schmidt, S. (2017). Replication. In M. C. Makel & J. A. Plucker (Eds.), *Toward a more perfect psychology: Improving trust, accuracy, and transparency in research* (pp. 233–253). American Psychological Association.
- Schwarz, N., & Strack, F. (2014). Does merely going through the same moves make for a "direct" replication? Concepts, contexts, and operationalizations. *Social Psychology*, 45, 305–306. <https://doi.org/10.1027/1864-9335/a000202>
- Sison, C. P., & Glaz, J. (1995). Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association*, 90, 366–369.
- Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project. *Psychological Science*, 30(5), 711–727.
- Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PLOS ONE*, 9(12), Article e114876. <https://doi.org/10.1371/journal.pone.0114876>
- Wallis, S. (2013). Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics*, 20, 178–208. <https://doi.org/10.1080/09296174.2013.799918>
- Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M., & Carsel, T. S. (2018). Why do some psychology researchers resist adopting proposed reforms to research practices? A description of researchers' rationales. *Advances in Methods and Practices in Psychological Science*, 1(2), 166–173.
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, Article e120. doi:10.1017/S0140525X17001972