UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Methodology and Research Practice

A Roadmap to Large-Scale Multi-Country Replications in Psychology

Hannes Jarke[1,2] [a], Shaakya Anand-Vembar[2,3], Shilaan Alzahawi[4], Thomas Lind Andersen[2,5], Lana Bojanić[6], Alexandra Carstensen[7], Gilad Feldman[8], Eduardo Garcia-Garzon[2,9], Hansika Kapoor[10,11], Savannah Lewis[12,13], Anna Louise Todsen[2,14], Bojana Većkalov[2,15], Janis H. Zickfeld[16,17], Sandra J. Geiger[2,18]

[1] Centre for Business Research, Judge Business School, University of Cambridge, Cambridge, UK, [2] Junior Researcher Programme, Cambridge, UK, [3] Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland, [4] Graduate School of Business, Stanford University, Stanford, CA, US, [5] Child and Adolescent Mental Health Center, Copenhagen University Hospital, Mental Health Services Copenhagen, Copenhagen, Denmark, [6] Department of Psychology and Mental Health, University of Manchester, Manchester, UK, [7] Psychology Department, University of California, San Diego, La Jolla, CA, US, [8] Psychology Department, University of Hong Kong, [9] School of Health, Camilo José Cela University, Villafranca del Castillo, Madrid, Spain, [10] University of Connecticut, Storrs, CT, USA, [11] Department of Psychology, Monk Prayogshala, Mumbai, India, [12] Psychological Science Accelerator, [13] International Collaboration Research Center, Ashland University, Ashland, OH, US, [14] School of Psychology and Neuroscience, Faculty of Arts, University of St Andrews, St Andrews, UK, [15] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands, [16] Centre for Integrative Business Psychology, Aarhus University, Aarhus, Denmark, [17] Department of Management, Aarhus University, Aarhus, Denmark, [18] Environmental Psychology, Department of Cognition, Emotion, and Methods, Faculty of Psychology, University of Vienna, Vienna, Austria

Collabra: Psychology

Classic findings from psychology and the behavioural sciences are increasingly being revisited. Methodological and technological advances provide opportunities to replicate studies across a wide range of countries and settings to investigate whether these findings are universally applicable, limited to specific countries, or vary in magnitude depending on settings. Researchers from around the world connect to revisit such findings collaboratively, adapt the original design to the Zeitgeist, integrate new knowledge to improve statistical analyses, and broaden the scope by testing effects globally – or at least in as many countries, as budget and feasibility allow. We currently observe multiple international consortia conducting large-scale multi-country replications. How do such collaborations form and how do they approach these complex investigations? This paper brings together researchers from different initiatives that conduct replications on an international scale to outline approaches and summarises what we have learned in applying them: Junior Researcher Programme (JRP), Psychological Science Accelerator (PSA), ManyBabies, Collaborative Open-science REsearch (CORE), and International Study of Metanorms (ISMN). We describe different ways for study selection, methodological approaches, statistical analyses, ethical issues, and most importantly, how the different collaborations formed and how team communication worked. We look in detail at challenges of including typically underrepresented countries in psychological science, not only in terms of data collection but also in making it possible for local researchers to contribute. This paper provides a structured insight into how different collaborations work and issues to consider for anyone who seeks to conduct a multi-country replication in psychology, or looking for additional perspectives to their existing plan. We close the article with a checklist built as a helpful tool for colleagues putting together their study protocols for such efforts – and invite them to collaboratively expand it in the future.

Introduction

The replication and reproducibility crisis (Pashler & Wagenmakers, 2012; Wiggins & Christopherson, 2019) has eroded trust in classic findings from psychology and behavioural sciences (Anvari & Lakens, 2018; Wingen et al., 2020). Revisiting such findings and investigating whether widely cited and applied concepts can be replicated and generalised across time, context, and countries is more and more visible and encouraged by researchers and journals.

---

a Correspondence should be addressed to Hannes Jarke, hj318@cam.ac.uk

Methodological and technological advances, as well as improvements in data collection and management practices can now be utilised to update our knowledge about theory and findings and expand further to reach populations around the globe. These advances provide new opportunities for initiatives that are sometimes labelled *Big Team Science*: collaborations "involving a relatively large number of collaborators who may be dispersed across labs, institutions, disciplines, cultures, and continents" (Forscher et al., 2020, p. 2). For example, to test their 1979 Nobel prize-winning *Prospect Theory,* Daniel Kahneman and Amos Tversky relied on a small random sample of students from the US, Sweden, and Israel. Using collaborative online tools, Ruggeri and colleagues (2019) recently replicated patterns of prospect theory across 19 countries and in 13 languages with over 4,000 participants. This project was facilitated with the help of the *Junior Researcher Programme* (JRP), an initiative that provides opportunities for early-career researchers in psychology (jrp.pscholars.org; Jarke, 2021; Ruggeri, 2020). Multi-country replication initiatives that started earlier include the *Open Science Collaboration* (2012), *ManyLabs* (e.g., Ebersole et al., 2014; Klein et al., 2014), or the *Collaborative Replications and Education Project* (CREP; Grahe et al., 2013). Another approach to such multi-country replications is facilitated by the *Psychological Science Accelerator* (PSA; Beshears et al., 2020; Moshontz et al., 2018), an organisation focused on producing psychological science that is generalizable and reliable by utilising their distributed network of labs.

In this paper, we aim to synthesise insights gained from several multi-country collaborations. In two online panel discussions (Anand-Vembar et al., 2021; Jarke et al., 2021), we brought together researchers who participated in and organised multi-country replication efforts to discuss key aspects of such undertakings (see Table 1 for an overview).

Based on these discussions, we summarise structured insights and experiences with different approaches on key topics, including study selection, replication criteria, translation, selection of potential collaborators, team communication, data collection methods and privacy issues, gaining ethical approval for multiple countries, and statistical approaches. We also address the inclusion of underrepresented countries and outline hurdles in including colleagues from these areas. We conceptualised this paper as a roadmap to support researchers planning a multi-country replication so that they can incorporate what has been learned in previous attempts and tailor successful approaches to their lab's situation. Some consortia already have compiled information on their approaches in detail. These are included in the accompanying reading list to this paper (Appendix 1). We discuss our insights chronologically, starting at the beginning of a study until publication and beyond. In addition, we compiled a checklist of lessons learned (Appendix 2). This checklist is not a comprehensive "must-have" but rather a tool that summarises considerations for preparing a multi-country replication. Researchers already involved in such efforts may also find suggestions and insights that can strengthen their existing approaches or add a perspective they had not previously considered. While this paper primarily focuses on replications including large numbers of labs and researchers, many insights are also applicable to replications within fewer countries or settings.

Building a Consortium: Identifying Collaborators and Team Communication

Perhaps the easiest way to build a consortium[1] of collaborators is to draw on existing networks. Uhlmann et al. (2019) describe multi-lab collaborations along two axes, one being inclusiveness versus selectivity (i.e., who can participate and make decisions), the other being communication (i.e., is there constant collaboration, or does the PI "collect" and synthesise the work of otherwise mostly independent units). We outline our own experiences below.

The JRP draws on its network of programme alumni and current team members, with current programme interns usually leading a country team in collaboration with students participating in GLOBES - a programme at Columbia University, which connects undergraduate students in behavioural sciences to researchers. The majority of collaborators are early-career researchers (ECRs; here: students and postdocs who make up small project teams each year but often go on to participate in larger collaborations), whereas the project itself is led by an established academic. JRP interns and alumni can be integrated into projects without further screening, as a working relationship and trust are already established through them having participated in a summer school and a one-year research project supported by the programme at that point. Participating researchers are briefed usually five to six months before activities commence and learn about their responsibilities, the project timeline, and how communication will be facilitated (Ruggeri, 2020). Shortly before and during data collection, communication is facilitated via daily briefing emails with updates, key developments, and reminders. These are supported by group calls and a slack channel for discussion between country teams as well as an FAQ page for providing important information across country teams. While the first JRP-organised multi-country replication (Ruggeri et al., 2020) was conducted with most contributors on-site in one place, the two following projects have been conducted online (Ruggeri et al., 2021, 2022) due to the COVID-19 pandemic. Although online work clearly allowed for more contributors, it also constitutes a much higher workload for the project leaders and administrators. Further, with the JRP's contributors being mostly ECRs, the in-person approach also provided a unique and motivating experience

---

1 We refer to a consortium in this context as any formal or informal collaboration between multiple labs

Table 1. Contributors who Participated in the Panel on Multi-Country Replication Studies.

Consortium	Authors	Relevant examples of work related to this topic
Psychological Science Accelerator (PSA)	Shilaan Alzahawi, Savannah Lewis, Janis Zickfeld	Moshontz et al. (2018), Beshears et al. (2020)
Junior Researcher Programme (JRP)	Hannes Jarke, Lana Bojanić, Eduardo Garcia-Garzon, Bojana Većkalov, Sandra J. Geiger	Ruggeri et al. (2020), Ruggeri et al. (2021), Ruggeri et al. (2022)
ManyBabies	Alexandra Carstensen	ManyBabies Consortium (2020), Byers-Heinlein et al. (2020), Byers-Heinlein et al. (2021), Tsui et al. (2021); see also https://manybabies.github.io/
Collaborative Open-science REsearch (CORE)	Gilad Feldman	Efendić et al. (2022), Ziano, et al. (2021), Chandrashekar, et al. (2021), Chen et al. (2021), Brick et al. (2021), Anvari et al. (2021), see also Collaborative Open-science REsearch (2022)
International Study of Metanorms (ISMN)	Hansika Kapoor	Eriksson et al. (2021)

and ECRs benefit more by forming stronger relationships with international colleagues. CORE works with a different model in which the Principal Investigator (PI) conducts all replications with students as part of the coursework in a one-semester course, publicly sharing each stage of replication, and inviting international involvement and collaborations throughout. Once projects are finalised at the end of the semester, ECRs are publicly invited to take the lead over the completed projects, verifying and extending the available outputs, and then helping bring these to publication in journals. The PI works with each ECR on their led projects, and over time, with gained experience, the ECRs support each other on specific issues or initiatives, and join forces in tackling new projects.

The PSA represents a globally distributed network of labs located on all populated continents (see https://psysci-acc.org/map/ for an updated map of all labs) organising data collection of democratically chosen projects (Moshontz et al., 2018). The network includes one director and four associate directors that are elected by all members. In addition, the network is coordinated by several committees focusing on different aspects of the study process such as study selection, ethics, translation and cultural diversity, community building, project monitoring, data and methods, training, and funding. The network regularly issues calls for project submissions from both inside and outside the network, which are first screened by the selection committee, then sent out for review by at least three experts, and finally voted on by all PSA members. Once a project has been successfully selected, PSA members are able to sign up for contributing to the project by collecting data, translating questionnaires or coordinating aspects of the project. All researchers can sign up to become a PSA member. Therefore, the PSA focuses both on inclusiveness and a mode of constant collaboration.

ManyBabies is a network of developmental psychologists distributed across a range of large-scale replication pro-

jects. New members opt in to the network and join individual projects through the organization's website. The governing board reviews new project proposals and advertises approved projects to the general listserv, where collaborators can volunteer to contribute to any and all parts of a project, including study design, data collection, analysis, and writing. The organization also has several dedicated positions, including an executive director, office assistant, and postdocs who support the range of ongoing studies. ManyBabies is committed to open science and uses a consensus-based approach for all project decisions to promote transparency, inclusivity, diversity, and collective governance.

Authorship and credit

No matter how a consortium is established, it is advantageous if everyone's responsibilities and the consequences of inaction are clear from the beginning. These responsibilities should ideally be listed in the final output (e.g., as a Contributor Role Taxonomy, or short CRediT; Allen et al., 2019; Holcombe et al., 2020, adopted also by the PSA) to transparently report what each author has contributed. To set expectations, it is also advisable to decide on the order of authors (or principles for determining author order) as early as possible and be clear on who will be responsible for article processing fees. For instance, the PSA has specific roles for each project (e.g., data management, translation coordinator, ethics coordinator) and these roles typically qualify for different authorship tiers; but also adapts criteria on a case-by-case basis beforehand. ManyBabies has a guiding example for determining authorship (https://manybabies.github.io/authorship) but the leadership team for each new project is responsible for defining and communicating authorship guidelines. Similarly, CORE employs a detailed guide for determining authorship (https://mgto.org/joinmassreplication).

We suggest that authorship be granted to all collaborators who contribute substantially to any crucial parts of the study. A collaboration agreement that transparently outlines the criteria for co-authorship as well as how authorship order is determined can ensure expectations on this topic are aligned. Authorship may act as a strong individual, and career incentive to academic collaborators, increases the public representation of researchers from marginalised backgrounds, and shows that contributions are credited and valued. In terms of administration, it should also be kept in mind that typing in all author names and data for submission can take many hours of work. Therefore, we urge journals to implement systems that allow for easier ways of providing such data, such as the option of filling in template spreadsheets that can be read and copied by the submission system. Big Team Science projects may run into other unexpected issues regarding authorship: For example, in a recent submission, the PSA Self-Determination Theory Collaboration (2022) could not list all 500 authors on the byline because this would have exceeded the journal's maximum page count. Instead, consortium authorship was adapted, which especially disadvantaged researchers from low-and middle-income countries and early career researchers: Institutional policies often do not recognize this type of authorship (for an extensive discussion about disadvantages resulting from group authorship and a call for change, see PSA's Open Letter[2]).

Study Selection

The process of deciding which study to replicate can vary substantially from lab to lab and consortium to consortium. Approaches can take various forms from where the PI or their lab decide which study to replicate, how to do it, and whom to include (*top-down*) to consortia forming through loose ideas and developing goals and structure from the *bottom-up*.

Within the PSA, several labs or PIs submit a study proposal, the study selection committee checks its feasibility, the network votes on the blinded proposal, then the lead team of the accepted studies pre-registers the experiment, and publicly opens it up for potential collaborators. There are also more bottom-up approaches to study selection. For instance, the ManyBabies Consortium accepts study proposals from any researcher, which are then evaluated by the ManyBabies Governing Board. Accepted proposals are advertised to ManyBabies Consortium members and through related organisations, enabling interested researchers and labs to join new projects. Similarly, CREP starts projects through collaborators expressing interest first. The study Director looks for top-cited studies in a range of sub-disciplines of psychology. After being checked for feasibility by established researchers, feasible studies are sent to students and collaborators to rate on different aspects, such as feasibility and interest. The CREP administrator then takes

those ratings and identifies the next studies for collaboration. After a study-specific admin team is put together for each study, labs are allowed to sign on to the project. On the other hand, replications conducted by the JRP have started with a small team around the PI deciding on the topic and detailed study plan first, before recruiting collaborators.

Feasibility and durability: Is this topic suitable for replication?

The most important criterion to decide on a topic is feasibility: whether a project is doable for the PI's team (planning and managing) and the partnered labs (execution). Key considerations include how easily the research can be executed across environments by all participating labs and whether all equipment is available across sites and whether necessary training has already been attained or can be provided. All parties need to carefully consider whether there are sufficient resources to coordinate such a large-scale effort, which can require substantial amounts of time-consuming administrative work. For example, the ManyBabies consortium has multiple staff members dedicated exclusively to non-research tasks. Basic considerations must include budget (for ethics review, study materials, personnel, etc.), whether research goals are attainable (e.g., study materials are clear and available), and whether the expected impact is worth the large-scale investment.

Replication targets can also be selected based on feasibility in terms of the replicators' qualifications, and the intended methods and target sample. In CORE, replication studies are designed by undergraduate students as part of a one-semester course with data collection designed using simplified survey platforms such as Qualtrics to be collected online using labour markets such as Amazon Mechanical Turk (with CloudResearch) and Prolific, mostly with US American and British participants. Therefore, in CORE the selected target articles are typically the highest impact classics that can be easily adjusted to an online design and with methods and statistics that are doable for undergraduate student level.

Isager et al. (2020) propose another approach, based on a formalised model that can help guide the decision on choosing a topic, aiming to calculate an *expected utility gain*, which they label *replication value*. For this model, the authors consider the costs of the study, uncertainty about the claim before replication, the value of the scientific claim, and the expected utility of the claim before the replication as determinants for the expected utility gain. However, they also state that these variables are not necessarily exclusive and further, unexpected variables can influence the utility gain.

---

2 https://docs.google.com/document/d/1OLb1VSkHLBo4z8XTEsqPUOK5PEYltbjAkpP2LyxNDSM/

Different types of replications

Direct replications[3] are usually defined as the repetition of a procedure, whereas *conceptual replications* test hypotheses or results of existing research using different methods or participant populations (Open Science Collaboration, 2015; Schmidt, 2009). In principle, both forms of replications can add valuable knowledge to the literature. A third possibility combining the two types is to run a direct replication with a conceptual replication extension added in a way that would have minimal impact on the direct replication yet allow for a comparison of the two types of replication. For example, Korbmacher et al. (2022) ran a direct replication of Kruger's (1999) above and below average effects which manipulated difficulty in a within-subject design, with an added extension on top of that introducing a manipulation in a between-subject design, resulting in a mixed design demonstrating the strength, robustness, and generalizability of the phenomenon. Chen et al. (2021) report a combination of two direct replications (Studies 1 and 2) and one additional conceptual replication (Study 3 building on the same design as in Study 1) within the same article.

A replication and extension design is especially helpful in disentangling the cause for failure in finding support for a novel contribution, whether the failure has to do with the novelty added or the replicability of the underlying phenomenon. Extensions can also add nuances to what is known, address potential weaknesses in the original design, or explore new directions that would advance the literature beyond the original's. For example, when Jakob et al. (2019) replicated an investigation of the relationship between membership in the fictional, fraternity-like Hogwarts houses from the Harry Potter franchise and psychometric measures of personality, the authors identified the concept of basic human values to be more in line with the description of these houses found in the books, and added this measurement to their questionnaire.

In addition to planning these technical aspects, it is also helpful to take a step back and consider the implications of the study design and its potential outcomes. Let's assume an effect replicates in the country where the original study was conducted, in a neighbouring country as well, but not in a third country, or across a specific set of states. The study would then look into not just whether an effect replicates, but also its *generalisability*. The two competing hypotheses here would be consistency in the effect vs. variation. This variability in specific countries can add nuance and allow for comparison of populations, even without adding extensions. For example, Hornsey et al. (2018) provide country-specific insights regarding political ideology and scepticism about climate change, highlighting how these relationships were stronger in the USA compared to the rest of the world. In a similar fashion, other studies can reveal how country or cultural differences might mediate the effect of interventions, as shown for social discounting (Tiokhin et al., 2019). As such, multi-country replications can be both replications and investigations into generalisability, showing either variation or robustness of a phenomenon across countries, or even means of data collection (e.g. near-identical results across platforms Mturk and Prolific, see Brick et al., 2021; Chandrashekar et al., 2022; Efendić et al., 2022).

Security and safety of collaborators and participants

Sometimes, a research topic may be incredibly valuable and relevant to both researchers and participants, but also too taboo in a country to study safely. This could involve research eliciting people's opinions on sensitive political, religious, health, social/moral issues, or sexuality, and can reach from these topics being frowned upon to being outright illegal to discuss in public. Even in situations where governing bodies or the laws of a country are not of concern, the community of a participant or collaborator may exact punishment if sensitive information is somehow made accessible. Consulting local collaborators in an informal manner (if possible) before deciding to go ahead with the replication is of utmost importance here, and if safety concerns are found, the best approach is to *not* conduct the study in that particular area. The security of participants and local researchers always comes before anything else.

Including Underrepresented Populations

Psychological research often lacks the inclusion of historically underrepresented populations (Thalmayer et al., 2021). However, their inclusion is key when trying to investigate phenomena globally. While researchers appear to be generally motivated to recruit collaborators and participants from outside more traditionally represented populations (e.g. those often summarised as western, educated, industrialised, rich and democratic [WEIRD], see also Arnett, 2008), they face a number of obstacles that limit the feasibility and perceived benefits of doing so. Below are some prevalent examples of such obstacles, along with suggestions for addressing them.

Funding

Reaching out to potential collaborators from different geographic areas often incurs costs that researchers are unprepared to bear. These may include costs for hiring and training local research assistants, high-quality translation of materials, travel associated with training and meeting collaborators, and incentivising participants. When multi-country efforts are not funded with an explicitly cross-cul-

---

3 Note that even a "direct" replication is not exactly identical to the original investigation. They are usually as close as possible, with ideally only inevitable differences present (e.g. a new sample). Brandt et al. (2014) argue that "close" replications would be a more fitting term.

tural approach in mind, it can be difficult for researchers to globalise their samples at later stages. In some cases, even well-funded studies designed to document cross-cultural phenomena can be difficult to implement because of restrictions from funding agencies requiring that purchases be made from companies based in the home country of the funder. Based on discussions in our panels and some literature (Azouaghe et al., 2020; Silan et al., 2021), funding appears to be the largest obstacle to achieving more globally representative samples in multi-country replications. Where the collaborations included here have made efforts to address these, they are summarised in Table 2 (When consulting this table, please note that the PI for JRP projects is not in the current authorship. Out of respect for the personal and professional sacrifices that the individual made toward the work, and that there will be an additional publication with extensive detail from the PI, we provide only brief information here). Consider also that the use of expensive software may provide a barrier to participating in a collaborative project for institutions that do not have necessary licenses. It is worth exploring if institutions are able to share licenses or provide project-based access, or considering open-source alternatives with bespoke edits made through expert collaborators.

Data vs. collaborators from underrepresented groups

Where funding is not a limiting factor, it is beneficial to approach cross-country replications with the mindset of not simply obtaining underrepresented or non-WEIRD *data*, but rather involving collaborators (researchers, clinicians, trainees, etc.) from underrepresented groups and regions in fundamental stages of the research process, ideally starting with the planning phase. This approach can help to broaden perspectives and mitigate potential power imbalances that may otherwise exist when lead researchers from high-resource institutions seek out collaborators from lower-resourced ones (Silan et al., 2021).

Part of making a study valuable to any group of people is approaching the topic of replication from all angles, including the local perspectives of researchers from all collaborating regions. Trying to conduct a replication in a vastly different culture than that of the original study may result in weaker replication findings, not due to issues in the methods, false positive findings, or a challenge to the central theory/phenomenon, but rather because that theory/phenomenon works differently in a different context in which the replication is conducted. For example, while conspiracy beliefs have been found to mediate the link between political ideology and risk perception in Americans, the effects seem to be much weaker in India (Puthillam & Kapoor, 2021). Conversely, a successful replication in a novel population may or may not mean that the studied phenomenon occurs in the same way across countries-it might also be the result of non-representative sampling.

Involving local collaborators from the outset of the project not only demonstrates consideration and respect but also benefits the research by ensuring that design decisions are cross-culturally informed. One option is to establish an "outreach committee" within the research group or consortium, that specialises in finding and contacting potential collaborators for specific tasks and studies. The PSA uses their community building and network expansion committee to find ways to engage with all members of the network (i.e. social events, conferences/seminars, slack threads, etc.) which in turn allows the committee to connect on a natural basis. For conceptual replications, inspiration could also be taken from the STRATEQ-2 project (Dujols & IJzerman, 2021), where researchers started developing an international questionnaire by first asking researchers from different countries around the globe to generate items relevant to the phenomenon at hand to gather what fits the respective cultural context.

Considering study value for participants and collaborators

The ability to recruit participants and collaborators will likely depend on how valuable a study's topic is to the local population it aims to sample (Silan et al., 2021). When considering a study replication, researchers are advised to consider (and ideally consult local collaborators about) how relevant the topic might be in different populations, as this might heavily influence participation rates or interest from potential collaborators. For instance, replicating a study about financial choices relating to health insurance costs might be of interest to US citizens but relatively inconsequential to someone in a country with universal healthcare provided by the state. Researchers can then either adapt the replication to better suit their target populations or, with the help of local collaborators, find ways to convey or augment the value and benefits (direct or indirect) of the research to potential participants. In the case of large studies requiring community investment or designed to impact communities through intervention, it may be most appropriate to involve community leaders and local stakeholders in ongoing conversations about both costs and benefits of the study. In determining adequate compensation, it is advisable to consult researchers and personnel with local knowledge, especially if a population may be vulnerable for economic or social reasons.

Strengths-based vs. deficit-based framing of cross-population comparisons

An important aspect of taking local perspectives into account is to frame research questions in a way that will adequately contextualise the study's results and allow room for the exploration of phenomena that may vary from those seen in the original study. One way of accomplishing this would be to use a strengths-based (as opposed to a deficit-based) model when comparing the original population with those from multi-country replications. Instead of viewing divergences in results as an indication of one population *performing better* at a task or *better understanding* a theory than the other population, it would be more meaningful to investigate how differing skill sets, values, goals, and incentives play into how different populations interact with the concepts being studied. As discussed in the previous

section, certain concepts or relationships may not be statistically significant in some populations; take for instance conspiratorial blame and political ideology as laid out by Puthillam and Kapoor (2021) – a researcher could either look at this phenomenon through a deficit-based lens (e.g. the comparison population has less of a grasp on conspiracy theories than the original sample, or they're less 'clued into'/exposed to international media), or they could use a strengths-based approach to explore *why* the comparison population is not incentivized or motivated to link conspiratorial blame to their political ideologies. What aspects of their context (daily life challenges, governmental structure, etc.) might counterbalance their (potential) inclination to engage with certain conspiracy theories?

Methodological Approaches

Even for conceptual replications, the general methodology is often at least partially based on the methodology of the original study. However, there are several methodological aspects that require extra attention and adaptation due to the cross-cultural nature of multi-country replications, and these come with some challenges to consider before embarking on such a study. Also, note that multi-country collaborations focussing on existing research and materials can take various forms, including the validation of metrics in different countries and cultures, or a revalidation to see if an instrument measures the same construct in different populations (see also Ruggeri et al., 2019).

Pre-registering your project

Pre-registration is always advisable in confirmatory research (Munafò et al., 2017) but especially important in multi-country replication studies, where adaptations are likely needed for different countries. Where possible, this may take the form of a registered report: a publishing model in which a journal first approves a study protocol and later the full report with results and conclusions, including potential deviations from the original plan. While registered reports are no deus ex machina to solve issues with reproducibility, initial evidence supports that they promote reproducibility, transparency, and self-correction across disciplines as intended (Chambers & Tzavella, 2021; Scheel et al., 2021) and that they are perceived as having greater research quality and rigour compared to the standard publishing model (Soderberg et al., 2021). However, consider also that publishing a registered report can impact the timeline of your project, as review at this early stage can take a long time and may delay the start of your study. If this a concern—or where materials need to remain private before the conclusion of data collection—a time-stamped pre-registration (private or public) on platforms such as the OSF can ensure that any deviations from the planned methodology are traceable upon publication.

Ethics

Applying for ethical approval can be challenging when operating in multiple countries. In general, there are two options: a centralised approval from the PI's Institutional Review Board (IRB) or approval from the PI's institution and at least one IRB per country. It is also possible that institutions or journals require approval from each institution's IRB. Having only one approval is obviously the easier approach but may not always be possible. IRBs may refuse such a wide-ranging approval because their members are not able to review materials in all languages involved or cannot assess data privacy matters for all countries. In our experience, contributors should at least sign a letter confirming their commitment to upholding procedures exactly as outlined in the procedural plans submitted to the IRB. Where approvals from multiple IRBs are needed, timelines should account for the potential of considerable delays so that data collection can commence in all countries around the same time, as global events may otherwise interfere with the validity of data collected at different time points. In the experience of PSA, approvals can take different forms and may take any time between just one week and nine months to be finalised, which might even lead to abandoning plans for a country. However, having obtained approval from one institution can often speed up the process as it may provide another IRB with extra confidence where other colleagues have already approved it. This should also be considered when making the choice of a private pre-registration versus registered report, as a journal may require all ethical approvals before consideration, which can cause delays in the project timeline. Further, you can find yourself in a situation where one IRB requires a change in approach (e.g., dropping a culturally sensitive question). This impact will then have to be reconsidered from a methodological perspective.

In addition to these administrative aspects, there are specific ethical considerations to keep in mind when planning research in multiple countries. Primarily, this means discussing cultural aspects with the local collaborators in detail and also involving local stakeholders for their approval. Especially when the research requires intense time commitment, the consortium should consider what to provide in return. For example, linguists commonly create dictionaries based on their work and share them with participants. Further, the age of consent for participation may differ between countries and IRBs may require a different age minimum, an aspect which makes sense to be checked with local collaborators early.

Translation and adaptation of materials

Another time-consuming aspect of multi-country studies is the translation of study materials. It may take multiple iterations and checks to confirm that materials are not only translated, but also that the translated text is perceived as intended in the base version of the materials. The forward- and back-translation protocol (Brislin, 1970) seems to be the most commonly utilised method in psychological research and requires at least one bilingual, as well as additional native speakers of the target language. Ideally, the forward- and back-translation should be complemented by methods that can test the conceptual (i.e., construct has a comparable meaning in all countries/cul-

Table 2. Actions regarding funding and including under-represented populations by different consortia so far.

	JRP	PSA	ManyBabies	CORE	ISM
Funding	Studies are conducted online; PI provides access to survey; centralised budget is provided to pay data collection services if not all data can be obtained by collaborators; collaborators work on their own time	Funding has been limited but does come up with a few of our projects for populations that are underrepresented or for translators. Surveys are typically run with an online link and funding for participants are most likely funded by the lab's funds or resources.	Project funding varies across the MB initiatives, but several have received funding through scientific agencies or through internal grants from a PI's home institution. Funds have been awarded to promote large collaborative projects and support capacity-building by supplying necessary training and equipment (e.g., laptops, speakers, software, ethical review, and a training workshop in the case of MB1-A, for labs that had not previously run infant looking-time studies). In many cases, however, collaborating labs opt in to projects that are consistent with existing grants and methods in the lab, and do not require additional funding.	Funding has been very limited so far. Single PI used his own seed and internal funding to conduct all data collections online using online labour markets.	Survey-based online research; PI provides access to survey; funds are available to offset data collection costs and participant incentives
Collaborators from underrepresented groups	Existing network is explored for trusted colleagues from as many countries as possible; the more the network grows, the more grows the outreach and colleagues from more countries can be included (First study: 32 collaborators and data from 19 countries; latest: 169 authors from 61 countries included)	The PSA is a Distributed Laboratory Network which means we have and recruit collaborators from any and all Geopolitical regions. Our network continues to grow and at the moment we have over 1300 researchers across 83 geopolitical regions.	Any researcher or lab can join an MB project; across the 16 MB studies to date, there are 490 collaborators from 45 countries on 6 continents, including 191 in the Americas, 230 in Europe, 21 in Africa, 34 in Asia, and 15 in Oceania.	The student body has been based at Hong Kong, extending to collaborations with early career researchers from around the world.	Researchers and participants from 57 countries, including 7 African countries, 10 American countries, 18 Asian countries, 21 European countries, and Australia.
Value for participants and collaborators	Participants: Surveys are short (5 minutes or less), minimal time investment – results are often shared via the same channels participants were recruited; Collaborators: Access to network of collaborators (option of further connection), opportunity to re-use data for additional research (e.g. country specific sub-analyses, or connecting data to new research question); for ECR – training (adaptation of materials, work with Qualtrics, data collection)	Participants: Value for participants varies for each project. Value can be monetary, knowledge based, and credit. Participants are also shared results through ways in which are understandable and public. Collaborators: The PSA allows for any career level to join and get involved in an aspect of the study to gain authorship, experience, and new skill sets. Collaborators also gain connections and	The families of infant participants are often compensated financially or given small gifts (e.g., books, toys) for participation. Collaborators benefit from training in experimental and analytic methods, open access to materials and data, authorship on publications, funding to attend professional conferences, and developing relationships with a large international network of colleagues.	For replication and extension projects early career researchers are invited to join as lead authors over projects started with students in course and thesis work. For the developed collaborative resources, any meaningful contribution is acknowledged and likely to lead to coauthorship.	Participants: Value for participants could be monetary or knowledge-based. Collaborators: Access to network of collaborators in the same area of research, opportunity to re-use data for additional research (e.g. country specific sub-analyses, or

	and support from country team members in conducting the study	access to researchers across the world. Other people: The PSA uses open material practices and therefore allows everyone to have access to our research methods and data. Value to the field is shown by contributing large and generalizable data sets for studies.			connecting data to new research question), co-authorship on all subsequent papers emerging from the main dataset.
Strengths-based approach	JRP encourages country-specific analyses, examples can be found in the appendix of Ruggeri et al. (2020)	N/A	By involving scientists from a diverse range of research sites in all project stages, MB provides a platform for perspectives that are often underrepresented within scientific discourse. MB projects have generally interpreted differential performance across participants and populations as evidence for the many paths that typical cognitive development may take.	N/A	ISM encourages country-level analyses and contextualisation of meta-norms in contemporary times (e.g., a follow-up study on COVID-19 related meta-norms)

tures), instrument (i.e., construct is operationalized in a similar way across countries/cultures), item (i.e., construct can be measured with the same instrument across countries/cultures), and scalar (i.e., construct can be assessed on the same metric) equivalence of an instrument to meaningfully compare different cultures/countries (Hui & Triandis, 1985).

Despite the prevalence of the forward- and back-translation method, evidence that this method yields the best results is lacking (Epstein et al., 2015). While some researchers consider back-translation indispensable, others argue that it does not necessarily guarantee equivalence and linguistic appropriateness of an instrument (Behr, 2017) and therefore advise against it (Epstein et al., 2015). In addition, there is no standardised method of translating research material in psychology (Cha et al., 2007; Epstein et al., 2015). This poses an issue as the back-translation process requires several people with different backgrounds/ expertise levels, making it difficult to achieve a truly valid translation where the target language is very rare or in regions where psychological research may not have a large presence. In these cases, it may be worth considering utilising a combination of translation techniques, depending on the resources and personnel available.

The PSA uses a variation of back-translation as its official method and also appoints language coordinators for each study.

The PSA method uses two translators, and goes as follows:

1. First bilingual translator translates from original language into target language.
2. Second bilingual translator back-translates the first translator's work into the original language.
3. The two versions are discussed, and a third version is made with changes.
4. External non-academic readers fluent in target language read the third version and give feedback.
5. Final version is created based on cultural feedback from external readers.

When following the back-translation method, focusing on the feedback of non-academic speakers of the target language is crucial, as they are ensuring that materials are jargon-free, make practical/social (not just grammatical) sense in their cultural contexts, and are easy to follow for everyone. To put it bluntly: When your colleague tells you your questionnaire is fine, but your grandma tells you it does not make sense, your grandma is right.[4] Sometimes such feedback may require changes to very fundamental parts of the materials, such as items in a scale that do not apply or make sense in a particular culture. Another issue to consider is adapting materials to local dialects or versions rather than leaving them in the 'standard' version of the language (such as Austrian German vs. Standard High German) - an extra measure to ensure cultural validity.

Whether such an adaptation is adequate should be evaluated by a local contributor. Additionally, in countries with multiple major languages where people may be fluent in more than one, discussions to decide which languages the materials will be translated to are helpful.

Given the culturally dependent and, at times, subjective nature of translation, transparency in the process is paramount. One of the easiest ways to maintain transparency is to keep track of changes made to materials through the translation process, and comment on the rationale behind those changes. This can be done in an informal way at first (such as commenting on a Word document or using the Track Changes feature), and later refined into an easily digestible format (such as a table with every major change to each material; see Supplementary materials in Ruggeri et al., 2020, 2021).

Especially when replicating older studies, adaptations to the base materials are sometimes needed. For example, studies including financial decisions need to be adapted to the local currency, but also to inflation, and anchored to income levels of the country (for an example, see the appendix of Ruggeri et al., 2020). Similarly, the Zeitgeist might make adaptation of materials necessary, as encountered in Wagenmakers and colleagues' (2016) attempt at replicating experiments underlying the facial feedback hypothesis, where the pictures used in the original study to evoke laughs were deemed not funny anymore.

Data Collection

The process of collecting and storing data from multiple countries differs across consortia, with some collecting data centrally (i.e., the PI saves data from multiple countries into their own database) and others collecting data through individual researchers or labs (i.e., collaborators store the data they collected in their countries, and then pass it on to the PIs). Both, national data protection laws and ethics committees, determine some aspects of how data is supposed to be collected and stored. While these determinants provide guidance on establishing a data collection protocol, the multitude of regulations also limits how data can be collected and shared and may require differences in how data is collected and stored from country to country, which can cause methodological inconsistencies.

Data Privacy

Data privacy should be considered in parallel to ethical issues before applying to an IRB. IRB applications can be complicated when multiple countries with different data protection laws are involved. It is advisable to think about data protection from a general ethical perspective first and try to adhere to this set of standards throughout the project, even if the law in some countries may not require it. In most aspects, the EU's General Data Protection Regulation

---

4 The first author's grandmother agrees with this assessment.

(GDPR) appears to be the strictest law in that sense. For those coming in contact with the GDPR for the first time, it is important to know that it only applies if personal data are collected, but is not applicable if individuals are not directly identifiable from the answers in a research endeavour (i.e. the data is pseudonymised, see Recital 26). When collecting data online, researchers should pay special attention not to collect IP addresses or geolocation of respondents (as these may make participants identifiable). GDPR also usually does not apply to data collection in research contexts (Article 89), so the applicability should be checked in advance – where possible with the help of a legal advisor or data privacy expert, such as the institution's *Data Protection Officer*. As GDPR is sometimes poorly understood and interpreted differently by IRBs, both outside and inside Europe, it is advisable to plan additional time to clarify these issues. Independent of which data laws apply, collecting non-personal and anonymised data is always the least likely to cause problems and avoids having to add complicated and long notices that can scare people off. If personal identifiable data are involved, researchers may also need to set up *Data Sharing Agreements* between the consortium partners (these regulate how the collected data are shared between partners and processed), or even *Data Privacy Impact Assessments* (a process to identify and minimise data protection risks which is necessary when processing sensitive data on a large scale). As of 2022, this also largely still applies to the UK, as the *Data Protection Act 2018* is the UK's adaptation of GDPR. Note that, depending on the category of data concerned, countries will often have specific laws pertaining solely to those categories (e.g., HIPAA or FERPA in the US). Based on GDPR categories and the authors' experiences, the following categories of data always warrant special attention to details and protection and will therefore be more complicated to handle:

- Racial or ethnic origin
- Health or genetic data
- Financial data
- Philosophical, political, or religious beliefs
- Sexual orientation
- Trade union membership
- Educational data / IQ scores
- Data from underaged individuals, especially children
- Data from closed networks, where the combination of answers may reveal identities

Statistical Approach

If you ask multiple data analysts what analysis to choose for your research project, chances are high you will find yourself with many different replies (Menkveld et al., 2021; Silberzahn et al., 2018). However, there are multiple aspects that should always be considered when choosing the statistical approach to a multi-country replication. For direct replications, studies should be powered to detect an effect at least equal to the one found in the original study, ideally even smaller – if there is no existing cross-cultural comparison, it is possible that the effect you are investigating may be smaller in some of the target countries. More generally, due to the combination of publication bias and small sample sizes, reported effect sizes in the literature tend to be inflated (Gelman & Carlin, 2014; Lane & Dunlap, 1978). Due to this "winner's curse", it is advisable to power replications to a smaller effect size than the one reported in the literature (e.g., using the small telescopes approach; Simonsohn, 2015). Where resources are constrained, sequential analyses can also be considered: This approach, commonly used in medical trials (where stakes for participants are high), uses interim analyses to observe if a sufficient sample has been reached while controlling for Type-1 error rate (Lakens, 2014).

In addition, researchers should take into consideration whether it is advantageous and feasible to power the study for individual regions or groups. One option is to replicate the effect in one country only before opting for a multi-country replication. That is specifically useful when the original evidence may be old and conducted on small samples. It is also an option to repeat the original analysis (often frequentist) and additionally analyse the data with an equivalent Bayesian analysis. In any case, the statistical approach(es) should be part of the pre-registration, and any additional, exploratory analyses should be labelled as such. To estimate the necessary sample size per country it should be considered whether the approach allows for a multi-level structure. If so, the number of countries and the variation of the effect at that level should be taken into consideration.

For conceptual replications, assessing sample size is more complicated. Assumptions on effect sizes should be as conservative as possible, yet might yield unrealistically large minimum sample sizes. Lakens (2021) provides detailed discussions on different approaches and highlights that, depending on the justification for a sample size, it should be considered "1) what the smallest effect size of interest is, 2) which minimal effect size will be statistically significant, 3) which effect sizes they expect (and what they base these expectations on), 4) which effect sizes would be rejected based on a confidence interval around the effect size, 5) which ranges of effects a study has sufficient power to detect based on a sensitivity power analysis, and 6) which effect sizes are plausible in a specific research area" (p. 1). When powering your study, you should also consider the practical implications of the effect size you are aiming for: while an extremely small but significant effect may provide insight, its practical relevance may be nil (Götz et al., 2021; see also reply by Primbs et al., 2021). While we would clearly not advise against research for the sole purpose of gaining knowledge in general, practical relevance is surely a key consideration for such large-scale endeavours where resources could be used to investigate potentially more beneficial matters. Anvari and colleagues (2021) provide an insightful discussion on this matter.

Reporting and Disseminating Findings

Once the study is completed, it is time for the exciting part of sharing the results with the world. The unique advantage in communicating results of a multi-country replication lies in the opportunity to 1) share results in every

author's native language and 2) highlight country-specific findings that might fall short of attention in the shadow of the overall findings of the paper. It can be advantageous to put a dissemination plan in writing early, to avoid scattered communication around results. All collaborators may inform their universities and institutions or faculty, who often provide the opportunity to write a blog post or news article. Sharing results before peer review has become fairly common but also raises concerns about research being shared uncritically, as most people may not be able to differentiate between peer-reviewed research and preprints. This concern was confirmed by Wingen et al. (2022), who experimentally showed that a brief explanation can help clarify this issue. As such, if you opt for a preprint publication, we recommend both clearly marking your preprint as such, as well as adding the explanation developed by Wingen and colleagues. In addition, it is good practice to ensure that all materials are accessible. Collecting information and conducting studies is costly – it is important to make this data available to achieve the maximum benefit of your work and resources, so that others may answer additional research questions. The PSA has even incentivised secondary analysis, by challenging researchers to work with one of their large datasets, offering monetary rewards (Forscher et al., 2019).

Concluding Remarks

Large-scale multi-country replications are not the most straightforward or easiest research endeavours. Yet they come with large benefits such as comparable data from different countries, and datasets which—if curated well—can be the source of future insights. While there is clearly no one-size-fits-all approach, we hope that the lessons we learned and summarised in this paper will be helpful at all steps of planning future multi-country replications. However, when approaching your multi-country replication, keep in mind to plan the protocol and responsibilities of collaborators well in advance and listen closely to your collaborators' insights of their own countries, identify potential pitfalls, and make sure everyone's safety is guaranteed. Likewise, share your knowledge about your own region or country. Your study also does not need to solve every question there is – a simple effect or theory is much more realistic to test at scale and makes it easier to provide tangible insights from the observations. Likewise, extensions should be equally straightforward.

Working on a large scale with colleagues from many countries can be a challenging but enriching experience and provides collaborators with research expertise and insights into how scientists operate in other parts of the world. These differences allow for additional perspectives and a more holistic view of phenomena, but they also require clear guidelines on communication channels and responsibilities (ideally provided in a way that is easily accessible to all collaborators.) This includes using communication tools that are accessible to every partner in the consortium. Simple and prevalent communication is key to the success of every project employing Big Team Science. However, it should also be considered that communication tools may not always take into account their usability for researchers with disabilities. In order to ensure that colleagues with disabilities can contribute without additional hurdles, such factors should be explored before the start of a project and not be made the responsibility of potential contributors.

Lastly, it is our impression that despite the differences in approaches, all large-scale multi-country endeavours we have been part of have one thing in common: the motivation to conduct such studies mostly stems from a drive to produce solid research that can help improve people's lives. Multi-lab approaches may facilitate a mindset-shift in how research is conducted, where instead of operating in silos and potentially competing for publication spots in journals, a collaborative approach allowing for different perspectives *within* one research project allows for not only more productive discussions where everyone has the same goal, but also leads to more nuanced insights. Including students in such initiatives can further help to support developments towards more collaborations of this kind and provide valuable early research experience and network opportunities. On the other hand, journals will also have to consider how to best provide impartial reviews as experts for such collaborations will become more and more connected to each other and it will become increasingly difficult to find editors who are knowledgeable in multi-country replications but without ties to author consortia that often include more than a hundred researchers. While we hope that our insights and checklist are helpful in conducting your multi-country replication, we also encourage researchers to build meaningful and lasting relationships with their project partners, moving towards a methodologically sound, more collaborative, and inclusive field. We would also like to encourage researchers who have created, or know of, additional materials for approaching this topic to get in touch and add resources or links to the accompanying online repository (https://osf.io/xrv5p/).

....

Contributions

HJ, SAV, and SJG have conceptualised and organised the panel discussions that form the basis of this paper. SA, LB, AC, GF, EGG, HK, SL, BV, and JZ were speakers at these panels (moderated by HJ and SAV) and contributed their insights. ALT and TLA compiled notes at these sessions that informed the writing of this manuscript. HJ structured the insights gained and led the writing of the first draft, supported by SAV and SJG. HJ supervised the administrative aspects of this project. All authors contributed to the writing and revision of this manuscript. SAV is overseeing the updating of the accompanying OSF repository.

Funding Information

Competing Interests

The authors declare no competing interests.

References

Allen, L., O'Connell, A., & Kiermer, V. (2019). How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*, *32*(1), 71–74. https://doi.org/10.1002/leap.1210

Anand-Vembar, S. (chair), Alzahawi, S., Carstensen, A., Garcia-Garzon, E., Kapoor, H., Lewis, S., Geiger, S. J., & Jarke, H. (2021, August 22). *Approaches to Multi-country Replications in Psychology and Behavioural Sciences – Session Two (of Two)* [Panel discussion]. 10th Junior Researcher Programme Conference, Online, UK. https://youtu.be/lGFi2Dp_Hp4

Anvari, F., Kievit, R., Lakens, D., Pennington, C. R., Przybylski, A. K., Tiokhin, L., Wiernik, B. M., & Orben, A. (2021). *Evaluating the practical relevance and significance of observed effect sizes in psychological research*. https://doi.org/10.31234/osf.io/g3vtr

Anvari, F., & Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, *3*(3), 266–286. https://doi.org/10.1080/23743603.2019.1684822

Anvari, F., Olsen, J., Hung, W., & Feldman, G. (2021). Misprediction of affective outcomes due to different evaluation modes: Replication and extension of two distinction bias experiments by Hsee and Zhang (2004). *Journal of Experimental Social Psychology*, *92*, 104052. https://doi.org/10.1016/j.jesp.2020.104052

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*(7), 602–614. https://doi.org/10.1037/0003-066x.63.7.602

Azouaghe, S., Adetula, A., Forscher, P. S., Basnight-Brown, D., Ouherrou, N., Charyate, A., & IJzerman, H. (2020). *Psychology and open science in Africa: Why is it needed and how can we implement it?* https://doi.org/10.31730/osf.io/ke7ub

Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology*, *20*(6), 573–584. https://doi.org/10.1080/13645579.2016.1252188

Beshears, J., Gjoneska, B., Schmidt, K., Pfuhl, G., Saari, T., McAuliffe, W. H. B., Steltenpohl, C. N., Onie, S., Chartier, C. R., & Moshontz, H. (2020). *Psychological Science Accelerator: A promising resource for clinical psychological science*. https://doi.org/10.31234/osf.io/qwy4k

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*(1), 217–224. https://doi.org/10.1016/j.jesp.2013.10.005

Brick, C., Fillon, A., Yeung, S., Wang, M., Lyu, H., Ho, J., Wong, S., & Feldman, G. (2021). Self-interest is overestimated: Two successful pre-registered replications of Miller and Ratner (1998). *Collabra: Psychology*, *7*(1), 23443. https://doi.org/10.1525/collabra.23443

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., Mastroberardino, M., Singh, L., Waddell, C. P. G., Zettersten, M., & Soderstrom, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology / Psychologie canadienne*, *61*(4), 349–363. https://doi.org/10.1037/cap0000216

Byers-Heinlein, K., Tsui, A. S. M., Bergmann, C., Black, A. K., Brown, A., Carbajal, M. J., Durrant, S., Fennell, C. T., Fiévet, A.-C., Frank, M. C., Gampe, A., Gervain, J., Gonzalez-Gomez, N., Hamlin, J. K., Havron, N., Hernik, M., Kerr, S., Killam, H., Klassen, K., … Wermelinger, S. (2021). A multi-lab study of bilingual infants: Exploring the preference for infant-directed speech. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1–30. https://doi.org/10.1177/2515245920974622

Cha, E. S., Kim, K. H., & Erlen, J. A. (2007). Translation of scales in cross-cultural research: Issues and techniques. *Journal of Advanced Nursing*, *58*(4), 386–395. https://doi.org/10.1111/j.1365-2648.2007.04242.x

Chambers, C. D., & Tzavella, L. (2021). The past, present and future of Registered Reports. *Nature Human Behaviour*, *6*(1), 29–42. https://doi.org/10.1038/s41562-021-01193-7

Chandrashekar, S., Adelina, N., Zeng, S., Chiu, Y., Leung, Y., Henne, P., Cheng, B., & Feldman, G. (2022). *Defaults versus framing: Revisiting Default Effect and Framing Effect with replications and extensions of Johnson and Goldstein (2003) and Johnson, Bellman, and Lohse (2002)*. Pre-print. https://osf.io/krsf2/

Chandrashekar, S., Yeung, S., Yau, K., Cheung, C., Agarwal, T. K., Wong, C., Pillai, T., Thirlwell, T. N., Leung, W., Li, Y., Tse, C., Cheng, B., Chan, H., & Feldman, G. (2021). Agency and self-other asymmetries in perceived bias and shortcomings: Replications of the Bias Blind Spot and extensions linking to free will beliefs. *Judgment and Decision Making*, *16*(6), 1392–1413.

Chen, J., Kwan, L. C., Ma, L. Y., Choi, H. Y., Lo, Y. C., Au, S. Y., Tsang, C. H., Cheng, B. L., & Feldman, G. (2021). Retrospective and prospective hindsight bias: Replications and extensions of Fischhoff (1975) and Slovic and Fischhoff (1977). *Journal of Experimental Social Psychology*, *96*, 104154. https://doi.org/10.1016/j.jesp.2021.104154

Collaborative Open-science REsearch. (2022). *Replications and extensions of classic findings in Social Psychology and Judgment and Decision Making.* http://osf.io/5z4a8

Dujols, O., & IJzerman, H. (2021, January 8). *Creating a psychological tool applicable to individuals around the world (STRAEQ-2 item generation and selection phase* [Blog post]. https://corelab.blog/creating-a-psychological-tool-applicable-to-individuals-around-the-world-straeq-2-items-generation-and-selection-phase/%26sa=D%26source=docs%26ust=1644954470444371%26usg=AOvVaw09PqgKtM_lPV84MLT2NHl3

Ebersole, C. R., Klein, R. A., & Atherton, O. E. (2014). *The Many Lab.* Online Project Repository. https://osf.io/89vqh/

Efendić, E., Chandrashekar, S., Cheong, S., Yeung, L., Kim, M., Lee, C., … Feldman, G. (2022). Risky therefore not beneficial: Replication and extension of Finucane et al. (2000)'s Affect Heuristic experiment. *Social Psychological and Personality Science*. https://doi.org/10.1177/19485506211056761

Epstein, J., Osborne, R. H., Elsworth, G. R., Beaton, D. E., & Guillemin, F. (2015). Cross-cultural adaptation of the Health Education Impact Questionnaire: Experimental study showed expert committee, not back-translation, added value. *Journal of Clinical Epidemiology*, *68*(4), 360–369. https://doi.org/10.1016/j.jclinepi.2013.07.013

Eriksson, K., Strimling, P., Gelfand, M., Wu, J., Abernathy, J., Akotia, C. S., Aldashev, A., Andersson, P. A., Andrighetto, G., Anum, A., Arikan, G., Aycan, Z., Bagherian, F., Barrera, D., Basnight-Brown, D., Batkeyev, B., Belaus, A., Berezina, E., Björnstjerna, M., … Van Lange, P. A. M. (2021). Perceptions of the appropriate response to norm violation in 57 societies. *Nature Communications*, *12*(1), 1481. https://doi.org/10.1038/s41467-021-21602-9

Forscher, P. S., DeBruine, L. M., Jones, B. C., Flake, J. K., Coles, N. A., & Chartier, C. R. (2019, September 1). *Introducing the PSA001 Secondary Analysis Challenge*. https://psysciacc.org/2019/09/01/introducing-the-psa001-secondary-analysis-challenge/

Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A. A., Dutra, N. B., Basnight-Brown, D., & IJzerman, H. (2020). *The Benefits, Barriers, and Risks of Big Team Science*. Preprint. https://doi.org/10.31234/osf.io/2mdxh

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2021). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, *17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Grahe, J., Brandt, M., Wagge, J., Legate, N., Wiggins, B., Christopherson, C., Weisberg, Y., Corker, K., Chartier, C., Fallon, M., Hildebrandt, L., Hurst, M., Lazarevic, L., Levitan, C., McFall, J., McLaughlin, H., Pazda, A., IJzerman, H., Nosek, B., … Adetula, A. (2013). Collaborative Replications and Education Project (CREP). *Open Science Framework*. https://doi.org/10.17605/OSF.IO/WFC6U

Holcombe, A. O., Kovacs, M., Aust, F., & Aczel, B. (2020). Documenting contributions to scholarly articles using CRediT and tenzing. *PLoS ONE*, *15*(12), 1–11. https://doi.org/10.1371/journal.pone.0244611

Hornsey, M. J., Harris, E. A., & Fielding, K. S. (2018). Relationships among conspiratorial beliefs, conservatism and climate scepticism across nations. *Nature Climate Change*, *8*(7), 614–620. https://doi.org/10.1038/s41558-018-0157-2

Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, *16*(2), 131–152. https://doi.org/10.1177/0022002185016002001

Isager, P. M., van Aert, R. C. M., Bahník, Š., Brandt, M. J., DeSoto, K. A., Giner-Sorolla, R., Krueger, J., Perugini, M., Ropovik, I., van 't Veer, A. E., Vranka, M. A., & Lakens, D. (2020). *Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints.* Preprint. https://doi.org/10.31222/osf.io/2gurz

Jakob, L., Garcia-Garzon, E., Jarke, H., & Dablander, F. (2019). The Science Behind the Magic? The Relation of the Harry Potter "Sorting Hat Quiz" to Personality and Human Values. *Collabra: Psychology*, *5*(1), 31. https://doi.org/10.1525/collabra.240D

Jarke, H. (2021, June 24). The Junior Researcher Programme - An initiative providing opportunities to early career researchers in psychology [Lightning Talk]. *Annual Conference of the Society for the Improvement of Psychological Science (SIPS) 2021, Online*. https://osf.io/s82qp/

Jarke, H. (chair), Bojanić, L., Feldman, G., Većkalov, B., Zickfeld, J., Geiger, S. J., & Anand-Vembar, S. (2021, August 22). *Approaches to Multi-country Replications in Psychology and Behavioural Sciences – Session One (of Two)* [Panel discussion]. 10th Junior Researcher Programme Conference, Online, UK. https://youtu.be/CHlgMCFTj9c

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica: Journal of the Econometric Society*, *47*(3), 263–291. https://doi.org/10.1111/j.1536-7150.2011.00774.x

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Korbmacher, M., Kwan, C., & Feldman, G. (2022). *Both better and worse than others depending on difficulty: Replication and extensions of Kruger's (1999) above and below average effects*. https://osf.io/7yfkc/

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, *77*(2), 221–232. https://doi.org/10.1037/0022-3514.77.2.221

Lakens, D. (2014). Refinements and New Developments: Performing High-Powered Studies Efficiently With Sequential Analyses. *European Journal of Social Psychology*, *44*(7), 701–710. https://doi.org/10.1002/ejsp.2023

Lakens, D. (2021). *Sample Size Justification*. https://doi.org/10.31234/osf.io/9d3yf

Lane, D. M., & Dunlap, W. P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 107–112. https://doi.org/10.1111/j.2044-8317.1978.tb00578.x

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johanneson, M., Kirchler, M., Razen, M., Weitzel, U., Abad, D., Abudy, M. (Meni), Adrian, T., Ait-Sahalia, Y., Akmansoy, O., Alcock, J., Alexeev, V., Aloosh, A., Amato, L., Amaya, D., Angel, J. J., … Bao, L. (2021). Non-Standard Errors. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3961574

Moshontz, H., Campbell, L., Ebersole, C. R., Ijzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., … Chartier, C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515. https://doi.org/10.1177/2515245918797607

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

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

Primbs, M., Pennington, C. R., Lakens, D., Silan, M. A. A., Lieck, D. S. N., Forscher, P. S., Buchanan, E. M., & Westwood, S. J. (2021). *Are Small Effects the Indispensable Foundation for a Cumulative Psychological Science? A Reply to Götz et al. (2022)*. https://doi.org/10.31234/osf.io/6s8bj

Psychological Science Accelerator Self-Determination Theory Collaboration, Legate, N., Nguyen, T., Weinstein, N., Moller, A., Legault, L., Vally, Z., Tajchman, Z., Zsido, A. N., Zrimsek, M., Chen, Z., Ziano, I., Gialitaki, Z., Ceary, C. D., Jang, Y., Lin, Y., Kunisato, Y., Yamada, Y., Xiao, Q., … Primbs, M. A. (2022). A global experiment on motivating social distancing during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *119*(22), 2111091119. https://doi.org/10.1073/pnas.2111091119

Puthillam, A., & Kapoor, H. (2021). *Does conspiratorial blame mediate the relationship between political ideology and risk perception? Evidence from India and the US*. https://doi.org/10.31234/osf.io/frgzq

Ruggeri, K. (2020, May 19). *The collaboration behind the replication* [Blog post]. https://socialsciences.nature.com/posts/the-collaboration-behind-the-replication

Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., Davison, C., Demić, E., Esteban-Serna, C., Friedemann, M., Gibson, S. P., Jarke, H., Karakasheva, R., Khorrami, P. R., Kveder, J., Andersen, T. L., Lofthus, I. S., McGill, L., Nieto, A. E., … Folke, T. (2020). Replicating patterns of Prospect Theory for decision under risk. *Nature Human Behaviour*, *4*(6), 622–633. https://doi.org/10.1038/s41562-020-0886-x

Ruggeri, K., Bojanić, L., van Bokhorst, L., Jarke, H., Mareva, S., Ojinaga-Alfageme, O., Mellor, D. T., & Norton, S. (2019). Editorial: Advancing methods for psychological assessment across borders. *Frontiers in Psychology*, *10*, 503. https://doi.org/10.3389/fpsyg.2019.00503

Ruggeri, K., Panin, A., Vdovic, M., Većkalov, B., Abdul-Salaam, N., Achterberg, J., Akil, C., Amatya, J., Amatya, K., Andersen, T. L., Aquino, S. D., Arunasalam, A., Ashcroft-Jones, S., Askelund, A. D., Ayacaxli, N., Sheshdeh, A. B., Bailey, A., Barea Arroyo, P., Mejía, G. B., & García-Garzon, E. (2022). The globalizability of temporal discounting. *Nature Human Behaviour*, *6*. https://doi.org/10.1038/s41562-022-01392-w

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., Barea-Arroyo, P., Berge, M. L., Bjørndal, L. D., Bursalıoğlu, A., Bühler, V., Čadek, M., Çetinçelik, M., Clay, G., Cortijos-Bernabeu, A., Damnjanović, K., Dugue, T. M., Esberg, M., Esteban-Serna, C., … Folke, T. (2021). The general fault in our fault lines. *Nature Human Behaviour*, *5*(10), 1369–1380. https://doi.org/10.1038/s41562-021-01092-x

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard Psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science*, *4*(2), 251524592110074. https://doi.org/10.1177/25152459211007467

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

Silan, M., Adetula, A., Basnight-Brown, D. M., Forscher, P. S., Dutra, N., & IJzerman, H. (2021, October 26). Psychological Science Needs the Entire Globe, Part 2 Let's Talk About the "C" Word: Colonialism and the Challenges of Psychological Science in the Developing World [Blog Post]. *Aps Observer*. https://www.psychologicalscience.org/observer/psychological-science-needs-the-entire-globe-part-2

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Rosa, A. D., Dam, L., Evans, M. H., Cervantes, I. F., … Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*(5), 559–569. https://doi.org/10.1177/0956797614567341

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., Esterling, K. M., & Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, *5*(8), 990–997. https://doi.org/10.1038/s41562-021-01142-4

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, *76*(1), 116–129. https://doi.org/10.1037/amp0000622

Tiokhin, L., Hackman, J., Munira, S., Jesmin, K., & Hruschka, D. (2019). Generalizability is not optional: Insights from a cross-cultural study of social discounting. *Royal Society Open Science*, *6*(2), 181386. https://doi.org/10.1098/rsos.181386

Tsui, A. S. M., Carstensen, A., Kachergis, G., Abubakar, A., Asnake, M., Barry, O., Basnight-Brown, D., Bentu, D., Bergmann, C., Binan Dami, E., Boll-Avetisyan, N., de Jongh, M., Diop, Y., Herrmann, E., Jang, C., Kizito, S., Lamba, T., Maliwichi-Senganimalunje, L., Marangu, J., … Frank, M. C. (2021). *Exploring variation in infants' preference for infant-directed speech: Evidence from a multi-site study in Africa*. Stage 1 RR accepted at Developmental Science.

Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., McCarthy, R. J., Riegelman, A., Silberzahn, R., & Nosek, B. A. (2019). Scientific utopia III: Crowdsourcing science. *Perspectives on Psychological Science*, *14*(5), 711–733. https://doi.org/10.1177/1745691619850561

Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., Jr., Albohn, D. N., Allard, E. S., Benning, S. D., Blouin-Hudon, E.-M., Bulnes, L. C., Caldwell, T. L., Calin-Jageman, R. J., Capaldi, C. A., Carfagno, N. S., Chasten, K. T., Cleeremans, A., Connell, L., DeCicco, J. M., … Zwaan, R. A. (2016). Registered Replication Report. *Perspectives on Psychological Science*, *11*(6), 917–928. https://doi.org/10.1177/1745691616674458

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

Wingen, T., Berkessel, J. B., & Dohle, S. (2022). Caution, preprint! Brief explanations allow non-scientists to differentiate between preprints and peer-reviewed journal articles. *Advances in Methods and Practices in Psychological Science*, *5*(1), 251524592110705. https://doi.org/10.1177/25152459211070559

Wingen, T., Berkessel, J. B., & Englich, B. (2020). No replication, no trust? How low replicability influences trust in psychology. *Social Psychological and Personality Science*, *11*(4), 454–463. https://doi.org/10.1177/1948550619877412

Ziano, I., Xiao, Q., Yeung, S. K., Wong, C. Y., Cheung, M. Y., Lo, C. Y. J., Yan, H. C., Narendra, G. I., Kwan, L. W., Chow, C. S., Man, C. Y., & Feldman, G. (2021). Numbing or Sensitization? Replications and Extensions of Fetherstonhaugh et al. (1997)'s "Insensitivity to the Value of Human Life." *Journal of Experimental Social Psychology*, *97*, 104222. https://doi.org/10.1016/j.jesp.2021.104222

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/57538-a-roadmap-to-large-scale-multi-country-replications-in-psychology/attachment/123028.docx?auth_token=KvOiqA6SMDFrubXeqHxa

Appendix 1

Download: https://collabra.scholasticahq.com/article/57538-a-roadmap-to-large-scale-multi-country-replications-in-psychology/attachment/123030.docx?auth_token=KvOiqA6SMDFrubXeqHxa

Appendix 2

Download: https://collabra.scholasticahq.com/article/57538-a-roadmap-to-large-scale-multi-country-replications-in-psychology/attachment/123193.docx?auth_token=KvOiqA6SMDFrubXeqHxa