

A conceptual review of uses and meanings of reproducibility and replication.

Sven Ulpts^{1*} and Jesper W. Schneider¹

¹Danish Centre for Studies in Research and Research Policy, Department of Political Science,
Aarhus University, Aarhus, Denmark

*Corresponding author: su@ps.au.dk

Summary

The present conceptual review, a component of work task 3.1 of the TIER2 project, examines the confusion surrounding terms like reproducibility and replication across the vast research landscape. This review is the foundation for a proposed conceptual framework outlined in Ulpts and Schneider (2024). In a time when people widely promote notions of reproducibility, this inclusive framework accommodates epistemic diversity. It focuses on two key issues: to what extent reproducibility is relevant and feasible across diverse research contexts.

Our conceptual review, distinct from numerous others, takes a unique stance. Rather than striving to establish new authoritative definitions or typologies, we contend that such efforts often remain entrenched, favouring one term over the other and introducing an expanding set of idiosyncratic qualifiers. Instead, we offer a practical, more straightforward solution to the terminological confusion applicable to real-world research while acknowledging the intricate complexities involved.

The conceptual review aims to provide a comprehensive yet condensed analysis of the confusion in the literature surrounding definitions, uses, and meanings of reproducibility, replication, and their variants. This analysis is based on existing reviews and a meticulous sampling of 422 definitions from primary documents across scientific fields. To supplement the study, we first outline generic linguistic issues with the seemingly related nouns reproducibility and replication. The main findings of this review are as follows.

First, our analysis of the sampled reviews and definitions suggests that the terminological and conceptual confusions surrounding reproducibility and replication are more extensive than previous literature indicates. There is no consensus at the aggregate conceptual level. The literature within and across disciplines contains considerable confusion regarding the

definitions of reproducibility, replicability, related terms such as repeatability, and their derivatives. There is also confusion at the disaggregate level of qualifiers. Qualifiers like conceptual, direct, exact, partial, and computational do not clarify matters, as no unified understanding exists. Different terms can mean the same, and the same terms can mean completely different things.

Second, although reproducibility and replication are used interchangeably as descriptive terms and implicitly refer to some broader, scholarly concepts, their definitions remain ambiguous. Despite this lack of clarity, these concepts are deeply rooted in a specific philosophy of science, aligning with the general but vague methodological notion of the 'scientific method' and its common epistemic principles of empiricism, fallibilism, and objectivity. This is well-known but also important to note because it underlies some of the challenges encountered when notions of reproducibility and replication are normatively promoted to be epistemic criteria and norms for all kinds of research.

Third, in their dictionary noun forms, reproduction and replication refer to an act or process of producing the same thing or something again. Thus, replication and reproduction both refer to a "redoing". Ambiguity sets in when we ask what should be redone: the same, similar, or varied.

Fourth, types of reproducibility or replication are defined either on the aggregate or disaggregate level by compound terms (e.g., qualifier plus noun). Unfortunately, this only gives a vague indication of what parts of a study are to be kept the same and similar, what is to be varied, and what epistemic function should be fulfilled by the practice. Defining types of reproducibility or replication in this enumerative way is lacking, and the general usefulness of proposed, seemingly exhaustive, enumerative typologies is limited. They do not have a unifying potential as the terminology is contested. They also suffer from a generic limitation. Their top-down approach, where a fixed universe is postulated a priori, is often too simplistic in its subject characterisation and inflexible, as new compound terms may arise that can be hard to accommodate into the existing scheme, if at all. We, therefore, argue that a bottom-up approach is more precise and flexible and thus more suitable for characterising types of reproducibility or replication.

Fifth, the noun forms of replication and reproducibility in the research literature represent vaguely defined or tacitly shared ideal types. These ideals or prototypes are embedded in specific research communities' values, conditions, and practices. They constitute combinations of what practices should ideally be achieved and which purposes should be fulfilled. However, there are inevitably deviations from these ideal types in research practice. The literature acknowledges these deviations, with some proponents of replication and reproducibility, for example, admitting that exact or direct replication is impossible to achieve.

Sixth, our conceptual review identifies two crucial underlying meanings about the verb forms reproduce and replicate: 1) they both refer to practices involved in redoing and 2) the epistemic functions intended to be achieved by these practices. Practices are determined by what parts of a study are to be redone, what should be kept the same or similar and what should be varied. Such parts are manifold and most likely impossible to enumerate in a fixed typology. Our review indicates that dominating discourses often narrowly link epistemic functions of reproducibility or replication to issues such as reliability or validity. However, our review also suggests that such discourses fail to acknowledge the vast number of functions identifiable in the literature. Complexity or diversity is set aside, probably because such taxonomies have normative aims rather than descriptive ones. Consequently, like

practices, epistemic functions of reproducibility or replication are likewise manifold and most likely impossible to enumerate in a fixed typology.

Seventh, based on these findings, we propose a practical approach to understanding types of redoing by juxtaposing epistemic functions and practices (parts) while indicating what essential parts will be varied and what parts will remain the same or similar. Essentially, this is a two-by-two matrix with open dimensions. This approach can handle the wide variability in practices and functions, allowing for a specific description of redoing in the concrete case without relying on ambiguous terminology at aggregate or disaggregate levels.

Eighth, in their adjective forms, replicable and reproducible relate to the feasibility of these acts or processes. A more encompassing term is the verb enable, which conveys the idea of making something possible or facilitating an action. Enable also encompasses the strongly associated aspects of transparency and sharing. Transparency refers to the quality or state of the reporting, where actions, intentions, or details should be clear, visible, and readily understood. Sharing refers to making material available to others in a usable way. The practices can make a specific kind of redoing feasible for others or enable others to trace and understand the how and why of research practices and claims. Hence, in the case of enable redoing, the specific form of enabling depends on the specific type of redoing intended. While enabling more general mutual accountability among researchers, the required form of enabling depends on what is necessary and sufficient to understand the how and why of the research practices involved.

We propose a practical solution by 1) substituting the acts associated with *replication* and *reproducibility* (and related terms) with **redoing** and 2) *replicable* and *reproducible* with **enabling**. We further acknowledge the intricate complexities in describing these practices and their diverse epistemic functions. We suggest that practices and functions are mapped for individual cases while specifying the intended sameness and variation of the key parts. This is a situated definition of a specific kind of redoing or enabling (see Figure 1). The distinction between redoing and enabling clarifies what is to be done (the practice) and for what purpose (the function), facilitating researchers or stakeholders in navigating this conceptually confused territory.

Introduction

Since claims of an alleged so-called “reproducibility crisis” in the early 2010s, there has been an increased (re)focus on the practices and roles of replication and reproducibility in research (see, e.g., OSC, 2015; Romero, 2019). Similarly, following this supposed “crisis”, there have been reforms with policy changes in line with a replication drive (Bissell, 2013; Cole et al., 2024; Drummond, 2019; Guttinger, 2020; Penders et al., 2019; Ulpts et al., 2025) motivated by a conviction of the general importance and necessity of replication and reproducibility as criteria and practices to ensure the quality and trustworthiness of research (findings) (see e.g., Peels and Bouter, 2023). Therefore, part of recent science reform movements is attempting to increase the credibility of science by not just increasing the number of replication studies and the reproducibility of research but spreading replication and reproducibility throughout the research landscape (see, e.g., Peels & Bouter, 2018; Penders et al., 2020; Peterson & Panofsky, 2023; Ulpts et al., 2025). Considering the heightened (re)focus on reproducibility, the sheer scale of these calls to increase reproducibility, and the critical role that seems to be attributed to it as an indicator for the quality of research and proper science (see e.g., Haig, 2022; Penders et al., 2019; Rubin, 2023), one has to ask what these terms mean. Previous reviews attempting to capture the meanings of reproducibility indicate that recently there is an immense terminological proliferation and fundamental conceptual confusion surrounding reproducibility across the research landscape (see, e.g., Barba 2018; Gomez et al., 2014; Gundersen, 2021; Liberman, 2015; Matarese, 2022). As Renee Borges put it after comparing, among others, the definitions for reproducibility in a report from the Netherlands Academy of Science (KNAW, 2018) with the definitions in the US National Academy of Science, Engineering and Medicine report (2019): “There is therefore confusion in the definition of the terms themselves, although everyone believes that they know what is being said.” (Borges, 2022, p. 1).

Most likely prompted by the sense of urgency, concerning issues around reproducibility, carried by the crisis narrative behind the Open Science and metascience reform movement (Bartscherer & Reinhart, 2025; Peterson & Panofsky, 2023), in recent years terms like reproducibility have become quite widespread in the literature throughout the research landscape. However, these terms originate from some form of experimental positivist orientation of scientific investigation (Penders et al., 2019; Ulpts & Schneider, 2023). Currently, the various uses and meanings of the terms seem to be determined by how much the types of reproducibility carry these notions of experimental positivist research, which are often linked to the notions of ‘the scientific method’ as well as the ‘self-correction’ mechanism of science, and to what degree they are embedded in the specific conditions, values, and practices of the research communities these terms are used in. Usually, reviews on this conceptual confusion or suggestions for capturing reproducibility across all kinds of research and contexts also provide their own definitions and taxonomies. This often adds to the confusion rather than resolving it. For example, Matarese (2023) attempts a general definition for replication, while Leonelli (2018) provides definitions for specific forms of reproducibility, illustrating how the notion of reproducibility relates differently to six different types of research.

Therefore, in this paper we present a short review of the literature on the meanings and usages of the terms reproducibility and replication as well as related terms like repeatability, and reanalysis. To that end we will first investigate how these terms are defined in general English dictionaries and use the distinction between nouns, verbs, and adjectives as guidance throughout the rest of the paper. Next, we will review existing reviews on the reproducibility and replication terminology, with special focus on two of the most prominent

ones, Barba (2018) and Matarese (2022). Afterwards we present our own framework to distinguish different types of reproducibility and replication independent of the confused reproducibility and replication terminology. To that end we differentiate between practices of redoing and enabling as well as the varying epistemic functions ascribed to those practices. Lastly, we will end with a short outlook and conclusion.

Methods

In an approach similar to Albertoni and colleagues' (2023) intriguing review on reproducibility in machine learning, but more extensive regarding how much literature was considered, we analysed a corpus of literature, starting with some existing reviews about the different meanings and usages of replication, replicability, reproduction, and reproducibility (see, e.g., Barba, 2018; Gomez et al., 2014; Gundersen, 2021; Köhler & Cortina, 2021; Lieberman, 2015; Matarese, 2022; Plessner, 2018). This broad set of eminent reviews was used as a seed for a subsequent snowball sampling of primary documents containing definitions across various scientific fields until a saturation point was reached. Sampling resulted in 422 definitions from 120 sources in the context of diverse disciplines as well as approaches over the last 55 years for kinds of replication, replicability, reproducibility, reproduction, and repeatability, as well as some related terms such as triangulation and reanalysis (<https://osf.io/6exzt>) to further, if not understand then at least reveal the immense conceptual confusion about reproducibility. Saturation means a sufficient pool of definitions was reached to reveal and analyse the underlying dimensions of these terms. Furthermore, for practical reasons, we had to stop at some point because there will inevitably be ever-new definitions for these terms due to the current climate surrounding reproducibility and replication and the diversity in types of research and research contexts.

Linguistic analysis of reproducibility and replication

Philosopher of science Nancy Cartwright (1991) provides an epistemological perspective on the meaning(s) of reproducibility and replication in reaction to Harry Collins' (1991) sociological perspective. Cartwright (1991) claims that terms like replication, reproduction and repetition denote distinct epistemological concepts, while many other authors seemingly use these terms interchangeably without any strict conceptual reference or define them more as sociological terms of practice within a research community (see, e.g., Barba, 2018; Collins, 1991). As is well known, neither terminological nor conceptual unity exists across or within disciplines. Interestingly, similar ambiguities can also be seen in definitions of these terms in everyday and specialised language dictionaries such as the Cambridge English Dictionary and Oxford Learner's Dictionary of Academic English. See the definitions in Table 1.

Table 1: Reproduction and Replication definitions in the Cambridge English Dictionary and The Oxford Learner's Dictionary of Academic English

	Cambridge English Dictionary	The Oxford's Learner's Dictionary of Academic English
Term	Definition	Definition
To reproduce [verb]	To produce a <u>copy</u> of <u>something</u> or to be copied in a production process. To produce a <u>copy</u> of <u>something</u> or to show or do something again.	To produce <u>something</u> again. To make a <u>copy</u> of a picture, piece of text, etc.
Reproducible [adjective]	<u>Able to</u> be shown, done, or made again.	That <u>can</u> be produced again; that can be copied.
Reproduction [noun]	A <u>copy</u> of <u>something</u> , especially a painting, or the process of copying something such as a book, document, or photograph.	The act or process of producing <u>something</u> again or of making something happen again in the same way; the act or process of producing copies of something
Replicate [verb]	To make or do <u>something</u> again in exactly the <u>same way</u> .	To produce the <u>same thing</u> again
Replicable [adjective]	that can be done in exactly the same way as before or <u>produced</u> again to be <u>exactly</u> the same as before:	That can be copied exactly
Replication [noun]	The act of making or doing <u>something</u> again in exactly the <u>same way</u> , or <u>something</u> that is made or done in this way	The act or process of producing the <u>same thing</u> again

In the case above, a noun represents entities or ideas, the act or process of producing something or the same again. Interestingly, all the nouns describe the acts described in their

verb counterparts. However, the Cambridge English Dictionary also includes the outcomes of those actions as parts of the definitions of the nouns. For instance, reproduction is also defined as a copy of something instead of just the act of copying and replication is also defined as something that is made or done again. Curiously, the noun forms reproducibility and replicability of the adjectives are not represented in either of the dictionaries.

A verb expresses an action, something that is done. In the case above, to reproduce means to produce something again or to produce or copy something. In the case of replicate, it means making, doing, or producing something again in (exactly) the same way. The adjective form reproducible is captured in both dictionaries as the ability that something can be done and made again or copied; for replicable, both dictionaries also add that it has to be exactly the same. Furthermore, the Cambridge Dictionary also includes the aspect of showing something as part of being reproducible.

As can be seen with the dictionary definitions, the challenge with terms like reproduce and replicate is that their meanings are barely distinguishable outside the research context.

Reviewing reviews

Numerous scholars have already tried to capture the meanings and confusions around the replication and reproducibility terminology across disciplines and over time (see, e.g., Barba, 2018; Gomez et al., 2014; Gundersen, 2021; Köhler & Cortina, 2021; Lieberman, 2015; Matarese, 2022; Plessner, 2018). Hence, there are already some reviews of the reproducibility terminology, with some trying to find the origin of the confusion and map a typology onto disciplines like Barba (2018), and others like Gundersen (2021), Köhler and Cortina (2021), and Matarese (2022, 2023) coming up with their own definitions after reviewing the literature. For brevity reasons here we mainly focus on the reviews from Barba (2018) and

Matarese (2022). In the reviews, some speak of the Claerbout/Donoho/Peng convention (Claerbout & Karrenbach, 1992; Donoho et al., 2009; Peng et al., 2006; Peng, 2011) originating from computational sciences as one of the most widely used terminologies: “‘replicability’ applies to a study whose conclusions have been confirmed by another study using different data and different analyses, while ‘reproducibility’ applies to those studies that once duplicated in their every single aspect, obtain the same result” (Matarese, 2022, p. 653). Matarese (2022) distinguishes between three main taxonomies:

- 1) methods, result, and inferential reproducibility (MRIR) from Goodman et al. (2016),
- 2) exact, direct, and conceptual replicability (EDCR: eminent in some social sciences, especially psychology), and
- 3) repeatability, replicability, and reproducibility (RRR).

Matarese (2022) states that all of these definitions, besides methods reproducibility, imply some replication test, while methods reproducibility is about providing conditions that enable replication tests. Replication test here means the actual conduct of a study (e.g., analysis, investigation, experiment) again according to some degree of sameness or variation regarding specific parts relevant to the kind of evaluation or assessment intended compared to an original study. (Matarese, 2022). Matarese (2022) further classifies these different concepts according to how they relate to the variation and sameness of the parts of a study. According to Matarese (2022), exact replicability and repeatability keep everything invariant to obtain the same result; result reproducibility, direct replicability and replicability change irrelevant parts but keep relevant one’s invariant to obtain the same result, while inferential reproducibility, conceptual replicability and reproducibility vary methodological aspects but obtain the same conclusion or inference.

Lastly, in her functional approach, she classifies them according to their epistemic function. Epistemic functions are “the roles they play in checking the scientific qualities of an experiment or study. Why is each kind of replicability important? What kind of scientific standard does it certify?” (Matarese, 2022, pp. 5660-5661). Accordingly, replications that keep everything the same address reliability and random errors without sampling errors; replications that vary irrelevant aspects address reliability and sampling errors; and finally, replications that change methodological aspects address validity and systematic errors.

In contrast, based on a sample of publications in different fields, Barba (2018) classifies the confused uses of the terms reproduce and replicate according to a decision tree:

- A) authors make no distinction between the terms and demonstrate an ignorant or careless use where differences in meaning or understanding are not acknowledged,
- B) authors use them distinctly. However, this group is further divided into two subgroups, B1 and B2. Barba (2018) argues that on a “spectrum that starts at a minimum standard of same data + same methods = same results” going to “new data and/or new methods in an independent study = same findings”, group B1 calls the former “reproduce” and the latter “replicate”, while group B2 calls them in order “replicate” and “reproduce” (p.6).

Although a common understanding seems to be assumed, the use of the terms is swapped between the groups, contributing further to the terminological confusion. Barba (2018) then categorises different disciplines’ uses of the terminology according to the three groups. For example, the use in political science and economics is categorised as A, microbiology, immunology (FASEB), and computer science (ACM) as B2. However, according to Barba

(2018), the reproduce-replicate terminology in B1 is most widely used and seemingly dominant in signal processing, scientific computing, econometry, epidemiology, clinical studies, internal medicine, neurophysiology, computational biology, biomedical research, and statistics. Hence, according to Barba (2018), B1 (the Claerbout/Donoho/Peng convention) is the most widely used, and the other disciplines should also adopt B1 while the confusion and deviations from it can, according to her, be traced back to Drummond (2009). Interestingly, just looking at these two reviews, we find a crucial contradiction in the meanings of the term reproduce or reproducibility. Matarese (2022) describes reproducibility as similar to result and inferential reproducibility with irrelevant aspects and relevant procedures and outcomes not being held constant, which is in apparent contradiction to Barba's (2018) most widely identified usage of B1 (Claerbout/Donoho/Peng convention). Interestingly, Matarese openly acknowledges this contradiction (confusion) and states: "This means that every time we find the terms 'replicability' and 'reproducibility' in a paper, it is always important to check how the terms are used." (Matarese, 2022, p. 653).

Reviewing meanings over disciplines and time

As a preamble, it is essential to remember that these conversations about types of replication and reproducibility often stem from an experimental positivist context or are formulated in reference to some form of experimentation. After looking at the existing reviews, we explored the 422 definitions that resulted from the snowball sampling with the existing reviews as the starting seed. Box 1 below provides an example of conceptual confusion, in this case, the one surrounding the compound term computational reproducibility.

Box 1. Case of confusion surrounding computational reproducibility

With the example of **computational reproducibility**, we can find extensive conceptual confusion. AlNoamani and Borghi (2018) define it:

“computational reproducibility generally refers to the description and sharing of software tools and data in such a manner as to enable their use and evaluation by others (Stodden, Guo & Ma, 2013).”

This could be seen as the computational version of the first half of Goodman et al.’s. (2016) definition of **methods reproducibility**:

“Methods reproducibility refers to the provision of enough detail about study procedures and data so the same procedures could, in theory or in actuality, be exactly repeated. Operationally, this can mean different things in different sciences. Methods reproducibility is meant to capture the original meaning of reproducibility, that is, the ability to implement, as exactly as possible, the experimental and computational procedures, with the same data and tools, to obtain the same results.”

Methods reproducibility even includes confusion within its definition. As one can see, even within the short definition of methods reproducibility in the beginning, no redoing is required. At the same time, in the second part, where the intention behind its formulation is outlined, it is included. Methods reproducibility (at least according to the first half of the definition) and AlNoamani and Borghi’s (2018) computational reproducibility are also very similar to what Romero (2017) calls **hypothetical replicability**, Nosek et al. (2022) **process reproducibility** and Büthe and Jacobs (2015) call **replication-in-thought**, while others again sometimes just call it **transparency** (Pratt et al., 2020). Shapin (1984) called this **virtual witnessing** and defined it as *“the technology of virtual witnessing involves the production in a reader’s mind of such an image of an experimental scene as obviates the necessity for either its direct witness or its replication.”* (Shapin, 1984, p. 491).

Interestingly, the definition from AlNoamani & Borghi (2018) and the similar ones listed above do not imply redoing to conduct what Matarese (2022) calls a replication test but are focused on potentially enabling redoing or understanding and traceability of research practices in some form. However, Kitzes (2017) definition, which, for instance, Leonelli (2018) uses in her six-part typology, does include actual redoing to conduct a replication test:

“A research project is computationally reproducible if a second investigator [...] can recreate the final reported results of the project, including key quantitative findings, tables, and figures, given only a set of files and written instructions” (Kitzes, 2017, p. 12). ”

This also aligns with what Hardwicke and colleagues (2021) call **analytic reproducibility**. Hence, even for computational reproducibility, we have the same term that refers to different things (including or excluding actual redoing/ a replication test) and different terms that refer to the same thing.

Based on these definitions we identified two crucial components of the meanings of the reproducibility and replication terminology. On the one hand, there are the practices involved, and on the other hand, there are the epistemic functions intended to be achieved by these practices. In line with the linguistic analysis, for the different practices we distinguish

between nouns, verbs and adjectives of replication and reproducibility and locate the distinction between practices of enabling and redoing within this linguistic typology of replication and reproducibility. Lastly, we present the multitude of possible epistemic functions of these various practices.

The Ideals of Replication and Reproducibility

Looking at the literature, the noun forms of replication and reproducibility describe ideals or prototypes instead of actual practices in everyday research. This is in accordance with Collins' and others' claims that the *rule of replicability* does not provide rules for its application (Chen, 1994; Collins, 1991). These ideals appear to be quite vague or tacitly shared types of replication and reproducibility that are largely dependent on the circumstances of the research provided by the values, conditions, and practices of the specific research communities they are used in (see e.g., Chen, 1994). These usually unattainable prototypes are combinations of certain practices that should ideally be implemented and the functions that these practices should fulfil. However, research reality and the literature indicate that in actual research practice, when replication or reproducibility are implemented into reality, there are inevitably deviations from these ideal types of reproducibility and replication in some way or another (see Chen, 1994). These deviations can, for instance, be related to the nature of the subject of investigation, forms of uncertainty and limited availability of resources like funding and time (Guttinger, 2020; Penders et al. 2019; Peterson & Panofsky, 2021; Schickore & Hangel, 2019). Furthermore, admissions concerning these deviations can be found in the literature, for instance, when even promoters of replication or reproducibility in the current Open Science reform movement state that exact replication is impossible in the reality of research (see, e.g., Nosek & Errington, 2017, 2020).

The practices of redoing and enabling

To distinguish different reproducibility and replication practices, the first question to ask is whether the practices involve an actual redoing of a previous study or specific parts of it or whether it is about enabling others to trace and understand what was done and why so that research is generally intersubjectively accountable by sharing material and providing enough detail and information (Matarese, 2022; Peels & Bouter, 2018). Here, we distinguish between verbs like to replicate or reproduce and adjectives like being reproducible or replicable. To circumvent the conceptual confusion about replication and reproducibility and not use the terms we are trying to capture themselves, we refer here to redoing and enabling. Our use of the term redoing was inspired by Schickore's (2011) "re-doing" in her historical analysis of experimental redoing. The decoupling of enabling from redoing was inspired, among others, by Pratt et al. (2020), who proposed decoupling of transparency from replication in qualitative research, and Matarese's use of the term enable when describing Goodman et al.'s (2016) methods reproducibility. The verbs to replicate and to reproduce refer to actual acts of redoing something. The success or failure of redoing is often evaluated based on some degree of sameness or variation of outcomes in comparison to a previous study. The outcome depends on what is being redone, the focus of evaluation, and the intended epistemic function. Although the definitions of replication or reproducibility are often formulated as successful replications (Matarese, 2022, 2023), there is considerable controversy in the literature regarding what actually counts as a successful replication (see, e.g., Nosek et al., 2022; Plant & Hanisch, 2020). Another distinction often mentioned is between factors that might vary and are relevant or irrelevant to a specific finding or understanding (Atmanspacher et al., 2014; Matarese, 2022; Nosek & Errington, 2020). Hence, here, the role of understanding, background knowledge, and theory determines what counts as a successful redoing and which features are necessary for such redoing (also see

Norton, 2015). Actual redoing is or should be done for a specific purpose; for that purpose, specific parts or the whole of a study are redone with certain degrees of variation or similarity. Because of the sheer number and diversity of definitions, practices and areas of application for reproducibility and replication, every list of addressed parts will inevitably be incomplete. Consequently, one can characterise types of replicating and reproducing that involve redoing parts or whole studies similarly, the same or with variation by identifying such parts and indicating what is supposed to be the same, similar or vary (see Table 2).

Table 2 Practices: Examples of redoing parts of original studies in the same, similar, or varied way.

	Same	Similar	Varied (not addressed or intentionally changed)
Researchers			
Theory			
Hypotheses			
Methods			
Software			
Data			
Instruments			
Operationalisations			
Ancillary assumptions			
Analysis			
Samples			
Environments			
Results			
Inferences			

...			
-----	--	--	--

Hence, if it is unclear what practices are involved, then redoing must be qualified by what parts of a study are to be kept the same, similar or vary. Albertoni et al. (2023) distinguish between two components: the team and parts of the workflow of a study. Workflow is what we call parts of a study, i.e. the practices, and we see the team as one such part of a study (also see Gundersen 2021).

The other type of practice often referred to in the reproducibility and replicability terminology is what we call enabling. Enabling practices are captured by the adjective forms of something being reproducible or replicable. The adjectives replicable and reproducible describe an ability, and the practice that allows for that ability is the specific enabling practice (see Peels and Bouters, 2018, for a definition of replication and replicability as a between replication studies and the ability to be replicated). Enabling is the provision of conditions that allow for “Nachprüfbarkeit”.

The general principle underlying enabling practices is what in German is called “Nachprüfbarkeit”. Similar to the issues with translating the German “Wissenschaft” into the English “science”, there are problems with direct translations into English. “Nachprüfbarkeit” is often translated into “verifiability”. However, like with the related German “Nachvollziehbarkeit” that does not really capture the meaning. It rather refers to an intricate combination of being able to trace what was done in a study and understand how as well as why. Hence, while “Nachprüfbarkeit” is a precondition for intersubjective accountability of research, reproducibility and replicability are just one possible instantiation of the more general “Nachprüfbarkeit”, but sometimes other practices are more relevant and feasible, or reproducibility and replication are simply redundant, considering that other practices already achieve it independent of reproducibility or replication (Ulpts & Schneider, 2023).

Enabling also encompasses practices that are often associated with the terms transparency and sharing. Transparency here is about the reporting and presentation of the research in a way that makes practices, details, and motives readily visible and understandable to readers. Sharing is concerned with making available parts of the research, like the data and material, in a usable way. This distinction between transparency and sharing is somewhat similar to Haven and colleagues' (2022) distinction between open science and transparency. On the one hand, the practice of enabling can make redoing a whole or specific combination of parts of a study feasible for others. On the other hand, these practices can enable others to trace and understand the how and why of the research practices. Hence, in the case of enabling redoing, the specific practices of enabling depend on the specific combination of study parts that are supposed to be redone the same, similarly, or with a variation. While enabling more general mutual (intersubjective) accountability ("Nachprüfbarkeit"), the required form of enabling depends on what is necessary and sufficient to trace and understand the how and why of the research practices and research decisions involved.

Epistemic Functions: The Why of Reproducibility and Replication

When claims are being made about replication or reproducibility being a crucial part of "the scientific method" or fundamental to "the nature of science" or a "cornerstone of science" (see, e.g., Gundersen, 2021; Haig, 2022; Jasny et al., 2011; Nosek et al., 2012; Simons, 2014)¹, one has to acknowledge that behind these terms, hides a plethora of different practices and ask what the functions of the various practices actually are (i.e. what we call epistemic functions, Albertoni et al., 2023 calls reason). The intended functions of replication

¹ Mark Rubin (2023), in his critical metascience piece on questionable metascience practices, he provides further references for these kinds of statements about the role of replication in science.

and reproducibility are linked to and often even depend on the specific underlying epistemology of the research. The epistemic function is the specific utility that the practice in question has for the research at hand and the researcher (Gläser, 2024). Matarese (2022), states that the epistemic function tells us why replication or reproducibility should be important based on the specific “roles” the practices “play in checking the scientific qualities” of a study and which specific “standards” they “certify” (pp. 5660-5661). Furthermore, one might be tempted to categorise or build a reproducibility and replication typology based on epistemic functions like, for example, Matarese’s (2022) or Schmidt’s (2009) functional approaches. However, our review indicates that the confusion about the epistemic functions is as immense as the confusion about the meanings uses concerning the practices these terms refer to. Just as a glimpse into this confusion, here is a non-extensive list of functions for the various types of reproducibility and replication we identified in the literature:

- Enable replication (redoing) identify findings likely to be wrong, determine effect sizes, detect, rule out and understand bias, increase certainty, assess and establish the integrity of scientific knowledge or science, empirically justify or qualify results, test theory, ensure reproducibility, generalizability, robustness, reliability, falsification, confirmation, verification, corroboration, various kinds of validity, check scientific merit, demonstrate flexibility in methods and variability, trustworthiness or increase trust, rigour, credibility, estimate, ruling out or reducing errors of various kinds, theory development and refinement, improve understanding or knowledge, learning or training, confidence in findings, understanding or conclusions, likelihood of findings being true/ facts (getting closer to “the” truth), prevent waste, calibration, traceability or transparency, knowledge accumulation, fraud detection or elimination, identify and reduce questionable research practices , establish stability/ consistency, objectivity,

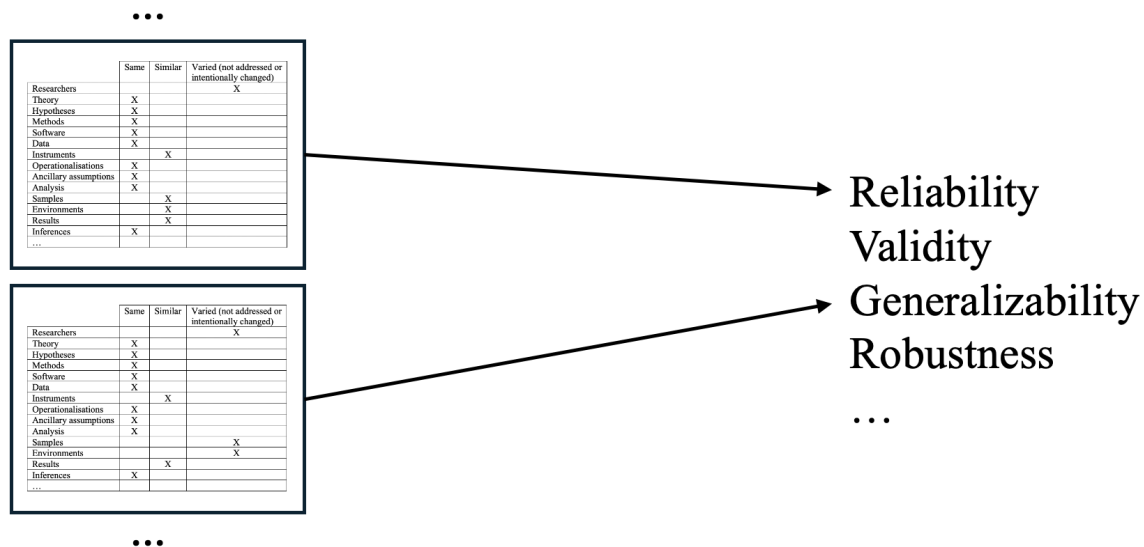
establish existence, improve sensitivity of analysis or measurement, improve design/methodology, accuracy, alleviate different kinds of underdetermination, reveal uncertainties, scientific progress, and precision.

The literature also indicates that epistemic functions are not directly linked to the actual practices of redoing and enabling but are part of the ideals of replication and reproducibility. Due to inevitable deviations, these ideals are only approximated in practice. Figure 1 shows two examples of mapping practices of redoing onto intended epistemic functions. The first combination of practices is an attempt by independent researchers to redo parts of a study as close as possible to the original; the purpose of such a redoing is often to illustrate or test reliability, a typical epistemic function for research with a positivist underpinning. In psychology, for instance, this is often referred to as direct or exact replication. The execution by an independent research team also qualifies this redoing as policing replication (Ankel-Peters et al., 2023). The other case below is an independent researcher's attempt at generalising findings by redoing the parts of an original study as close as possible to the original but changing the sample and environment compared to the original study. Again, the function is linked to notions of positivism.

Figure 1. Possible mapping of functions onto practices of redoing.

Practices of Redoing

Epistemic Functions



Then, there are ongoing discussions and disagreements about which epistemic functions are achievable or even appropriate with which practices. For instance, there is no agreement about what must be redone to achieve a specific epistemic function, nor is there consensus about what enabling practices serve which function. Hence, it is important to stipulate that there is no agreement about what combination of variation and similarity of which parts are necessary and/ or sufficient for a specific epistemic function, such as types of reliability, validity, or generalizability. Therefore, the many identified functions further complicate the conceptual confusion of reproducibility and replication.

Notably, most, if not all, of these suggested functions of replication and reproducibility are not unique to replication or reproducibility. Hence, in actual research, researchers will often have to choose whether to use practices of replication and reproducibility or to use other practices aiming to fulfil similar or the same epistemic functions. As indicated earlier, there are practices under different names, like triangulation or multiple determination, that fulfil similar or even the same functions as certain types of reproducibility and replication.

Given the sheer diversity and number of functions, meanings, disagreements, and typologies, it might be possible and appropriate to provide a reproducibility or replication typology for a specific kind of research under pre-specified conditions. However, it seems unfeasible if the aim is to construct a general typology for all kinds of research and situations. The literature indicates that the epistemic functions linked to replication and reproducibility are not directly linked to actual practices of redoing and enabling, but are instead part of the ideals of replication and reproducibility. Therefore, due to inevitable deviations from these ideals, in practice the stated epistemic functions are only approximated.

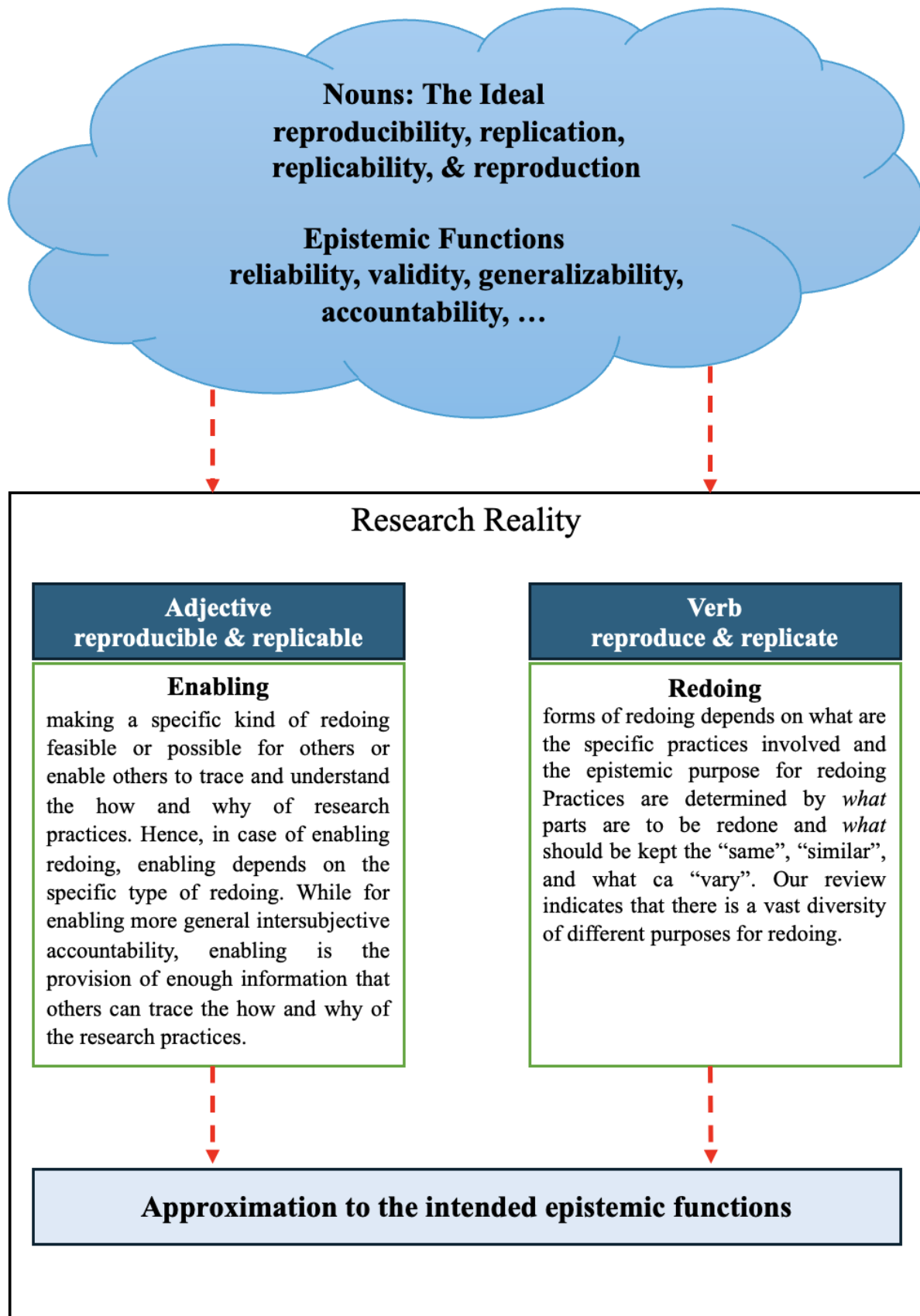
Synthesis: From ideal practices and functions to research reality

After scrutinising the varied components of reproducibility and replication practices and their intended epistemic functions, we will now briefly summarise and bring it all together. The nouns replication, replicability, reproduction and reproducibility are ideal forms of redoing and enabling with idealised epistemic functions rooted in basic epistemologies. When these abstract ideals are translated into reality, something is either, on the one hand, replicated or reproduced and, on the other hand, made replicable or reproducible. Hence, in actual research, these are practices of redoing and enabling. Redoing is about specific practices that prescribe a specific combination of parts of a study that are supposed to be redone the same, similarly or with variation. Enabling practices, such as detailed reporting and sharing specific parts of a study, are about making it possible for someone to redo something or trace and understand what was done and why.

Our review further indicates that the epistemic functions are attached to the ideal types of replication and reproducibility. No direct link exists between the practices of replicating or reproducing something in the form of redoing. Because actual practices of

redoing and enabling always deviate more or less from the ideals during actual research conduct, the indented functions captured by the abstract ideals are usually approximated in practice but not fulfilled in a definite way. See Figure 2 for a synthesis of our categorisation of essential aspects of reproducibility and replication.

Figure 2. From the Ideals of Replication and Reproducibility with their epistemic functions to the research reality of enabling and redoing practices approximating these functions.



Conclusion

Our review is, as is any, inevitably incomplete and purpose-driven. However, it indicates that the conceptual confusion surrounding reproducibility and replication is more severe than the previous literature indicates. There is conceptual confusion in multiple ways. There are diverse and contradicting meanings and uses of these terms. More precisely, we found the same term with multiple meanings referring to diverse combinations of practices and epistemic functions and multiple different terms with the same meanings across and within disciplines and approaches over the last 55 years. There is conceptual confusion in multiple ways. There are higher-level distinctions, with considerable variations in meanings and implications, between replicability, reproducibility, replication, reproduction, and repeatability (see e.g., Barba, 2018; Plessner, 2018), but there are also confusions at lower levels between different kinds of replication and reproducibility that are supposed to be distinguished by qualifiers like exact, conceptual, direct, computational, and operational, leading to compound terms such as computational reproducibility, inferential reproducibility, direct and conceptual or exact and inexact replication, and so forth. Worsening this confusion is the circumstance that ever-new terms and definitions are introduced in the literature (e.g., Matarese, 2023, for a recent attempt at a general definition of replication). One of the reasons why new definitions continue to emerge in the context of a prevailing replication drive and why there is redundancy and overlap in terms (Albtertoni et al., 2023), is that diverse kinds of research and research situations relate quite differently to different types of reproducibility and replication (see, e.g., Leonelli, 2018; Penders et al., 2019; Ulpts & Schneider, 2023). Likewise, recently, numerous scholars have attempted to provide definitions of reproducibility or replication that are appropriate for the nature of their research in which replication or reproducibility is traditionally often alien concepts (see, for example, Huijnen et al., 2022; Schöch, 2023; Talkad Sukumar & Metoyer, 2019).

Moreover, some philosophers of science continue to suggest clarifications about what the concepts of replication mean (see, e.g., Fletcher, 2021; Hudson, 2023). More positively, one might also argue that each community or, in a more extreme case, each researcher has their own understanding of these terms according to their own local, epistemological, ontological, disciplinary, methodological, theoretical, and institutional context. Furthermore, one comprehensive general typology or definition might not just be inappropriate but maybe even impossible, considering the diversity of functions, contexts, methods, subjects of investigation, epistemologies, ontologies, disciplinary standards, practices and much more across the research landscape (Guttinger, 2020; Leonelli, 2018; Penders et al., 2019; Ulpts & Schneider, 2023). Moreover, any new terminology or typology might be short-lived and inevitably controversial and contestable (see, e.g., Penders et al., 2019). Therefore, any new typology based on the replication or reproducibility terminology might, instead of being helpful, exacerbate the confusion and make the situation even less practical. As Borges (2022) states, despite the conceptual confusion, everyone seems to assume to know what reproducibility means.

Due to the multiplicity in practices, functions and meanings and the prevailing conceptual confusion in some research communities, without stating or knowing what replication or reproducibility means, it is practically meaningless and potentially misleading to argue for their widespread relevance, importance or even necessity. As a constructive solution, we propose to use two generic terms, redoing and enabling, to indicate the fundamental practices underlying the reproducibility and replication terminology, thereby escaping the confused terminology. By using our distinction between enabling and redoing in addition to the associated functions, this boils down to detailing:

- What is to be done - what are the practices?
- And for what purpose - what is the epistemic function?

Researchers and other stakeholders will thus be able to navigate this conceptually confused territory. Unfortunately, the literature does not suggest any agreement about what practices of redoing or enabling precisely fulfil which function. By allowing researchers to navigate through the conceptual confusion in the literature, our distinctions between practices and epistemic functions as well as between practices of redoing and enabling might provide an opportunity to more accurately assess and compare the rates of replication by ensuring that usages of terms and associated operationalizations of replication and reproducibility are actually comparable (Bak-Coleman & Devezer, 2023) as well as to more precisely assess the appropriateness of different types of replication and reproducibility for different kinds of research in various research contexts (Guttinger, 2020; Leonelli, 2018; Penders et al., 2019, 2020; Ulpts & Schneider, 2023).

Acknowledgement: This research is supported by the Horizon Europe project TIER2 under grant agreement No 101094817

References

- Albertoni, R., Colantonio, S., Skrzypczyński, P., & Stefanowski, J. (2023). *Reproducibility of Machine Learning: Terminology, Recommendations and Open Issues* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2302.12691>
- AlNoamany, Y., & Borghi, J. A. (2018). Towards computational reproducibility: Researcher perspectives on the use and sharing of software. *PeerJ Computer Science*, 4, e163. <https://doi.org/10.7717/peerj-cs.163>
- Ankel-Peters, J., Fiala, N., & Neubauer, F. (2023). Do economists replicate? *Journal of Economic Behavior & Organization*, 212, 219–232. <https://doi.org/10.1016/j.jebo.2023.05.009>
- Atmanspacher, H., Bezzola Lambert, L., Folkers, G., & Schubiger, P. A. (2014). Relevance relations for the concept of reproducibility. *Journal of The Royal Society Interface*, 11(94), 20131030. <https://doi.org/10.1098/rsif.2013.1030>
- Bak-Coleman, J. B., & Devezzer, B. (2023). *Causal claims about scientific rigor require rigorous causal evidence*. <https://doi.org/10.31234/osf.io/5u3kj>
- Barba, L. A. (2018). *Terminologies for Reproducible Research*. <https://doi.org/10.48550/ARXIV.1802.03311>
- Bartscherer, S. F., & Reinhart, M. (2025). The (Non)Academic Community Forming around Replications: Mapping the International Open Science space via its Replication Initiatives. Zenodo. <https://doi.org/10.5281/zenodo.16752370>
- Bissell, M. (2013). Reproducibility: The risks of the replication drive. *Nature*, 503(7476), 333–334. <https://doi.org/10.1038/503333a>
- Borges, R. M. (2022). Reproducibility and replicability in science: A Sisyphean task. *Journal of Biosciences*, 47(1), 15, s12038-022-00259–6. <https://doi.org/10.1007/s12038-022-00259-6>

Büthe, T., & Jacobs, A. (2015). *Conclusion: Research Transparency For Diverse Discipline*.

<https://doi.org/10.5281/ZENODO.892983>

Cartwright, N. (1991). Replicability, Reproducibility, and Robustness: Comments on Harry Collins. *History of Political Economy*, 23(1), 143–155.

<https://doi.org/10.1215/00182702-23-1-143>

Chen, X. (1994). The rule of reproducibility and its applications in experiment appraisal.

Synthese, 99(1), 87–109. <https://doi.org/10.1007/BF01064532>

Claerbout, J. F., & Karrenbach, M. (1992). Electronic documents give reproducible research a new meaning. *SEG Technical Program Expanded Abstracts 1992*, 601–604.

<https://doi.org/10.1190/1.1822162>

Cole, N. L., Ulpts, S., Bochynska, A., Kormann, E., Good, M., Leitner, B., & Ross-Hellauer, T. (2024, December 23). Reproducibility and replicability of qualitative research: an integrative review of concepts, barriers and enablers.

https://doi.org/10.31222/osf.io/n5zkw_v1

Collins, H. M. (1991). The Meaning of Replication and the Science of Economics. *History of Political Economy*, 23(1), 123–142. <https://doi.org/10.1215/00182702-23-1-123>

Committee on Reproducibility and Replicability in Science, Board on Behavioral, Cognitive, and Sensory Sciences, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Nuclear and Radiation Studies Board, Division on Earth and Life Studies, Board on Mathematical Sciences and Analytics, Committee on Applied and Theoretical Statistics, Division on Engineering and Physical Sciences, Board on Research Data and Information, Committee on Science, Engineering, Medicine, and Public Policy, Policy and Global Affairs, & National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and Replicability in Science* (p. 25303). National Academies Press. <https://doi.org/10.17226/25303>

- Donoho, D. L., Maleki, A., Rahman, I. U., Shahram, M., & Stodden, V. (2009). Reproducible research in computational harmonic analysis. *Computing in Science & Engineering*, 11(1), 8–18. <https://doi.org/10.1109/MCSE.2009.15>
- Drummond, C. (2009). *Replicability is not reproducibility: Nor is it good science*. Proceedings of the Evaluation Methods for Machine Learning Workshop at the 26th ICML, Montreal, Canada, 2009, Collection / Collection : NRC Publications Archive / Archives des publications du CNRC.
- Drummond, C. (2019). Is the drive for reproducible science having a detrimental effect on what is published?: Influencing what is Published. *Learned Publishing*, 32(1), 63–69. <https://doi.org/10.1002/leap.1224>
- Fletcher, S. C. (2021). The role of replication in psychological science. *European Journal for Philosophy of Science*, 11(1), 23. <https://doi.org/10.1007/s13194-020-00329-2>
- Gläser, J. (2024). How can we make ‘research quality’ a theoretical concept? *Research Evaluation*, 33, rva038. <https://doi.org/10.1093/reseval/rvae038>
- Gómez, O. S., Juristo, N., & Vegas, S. (2014). Understanding replication of experiments in software engineering: A classification. *Information and Software Technology*, 56(8), 1033–1048. <https://doi.org/10.1016/j.infsof.2014.04.004>
- Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, 8(341). <https://doi.org/10.1126/scitranslmed.aaf5027>
- Gundersen, O. E. (2021). The fundamental principles of reproducibility. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2197), 20200210. <https://doi.org/10.1098/rsta.2020.0210>
- Guttinger, S. (2020). The limits of replicability. *European Journal for Philosophy of Science*, 10(2), 10. <https://doi.org/10.1007/s13194-019-0269-1>

- Haig, B. D. (2022). Understanding replication in a way that is true to science. *Review of General Psychology*, 26(2), 224–240. <https://doi.org/10.1177/10892680211046514>
- Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., deMayo, B. E., Long, B., Yoon, E. J., & Frank, M. C. (2021). Analytic reproducibility in articles receiving open data badges at the journal *Psychological Science*: An observational study. *Royal Society Open Science*, 8(1), 201494. <https://doi.org/10.1098/rsos.201494>
- Haven, T., Gopalakrishna, G., Tjldink, J., Van Der Schot, D., & Bouter, L. (2022). Promoting trust in research and researchers: How open science and research integrity are intertwined. *BMC Research Notes*, 15(1), 302. <https://doi.org/10.1186/s13104-022-06169-y>
- Hudson, R. (2023). Explicating Exact versus Conceptual Replication. *Erkenntnis*, 88(6), 2493–2514. <https://doi.org/10.1007/s10670-021-00464-z>
- Huijnen, P., & Huistra, P. (2022). *On the use of replications in history*. <https://doi.org/10.5281/ZENODO.7037401>
- Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Again, and Again, and Again *Science*, 334(6060), 1225–1225. <https://doi.org/10.1126/science.334.6060.1225>
- Kitzes, J. (Ed.). (2017). *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences*. University of California Press. <https://doi.org/10.1525/9780520967779>
- KNAW. (2018). *Replication studies: Improving reproducibility in the empirical sciences*. Royal Netherlands Academy of Arts and Sciences.
- Köhler, T., & Cortina, J. M. (2021). Play It Again, Sam! An Analysis of Constructive Replication in the Organizational Sciences. *Journal of Management*, 47(2), 488–518. <https://doi.org/10.1177/0149206319843985>

- Leonelli, S. (2018). Rethinking Reproducibility as a Criterion for Research Quality. In L. Fiorito, S. Scheall, & C. E. Suprinyak (Eds.), *Research in the History of Economic Thought and Methodology* (Vol. 36, pp. 129–146). Emerald Publishing Limited.
<https://doi.org/10.1108/S0743-41542018000036B009>
- Liberman, M. (2015, October 31). Language Log » Replicability vs. Reproducibility—Or is it the other way around? *Language Log*.
<https://languagelog.ldc.upenn.edu/nll/?p=21956>
- Matarese, V. (2022). Kinds of Replicability: Different Terms and Different Functions. *Axiomathes*, 32(S2), 647–670. <https://doi.org/10.1007/s10516-021-09610-2>
- Matarese, V. (2023). A new concept of replication. *Inquiry*, 1–26.
<https://doi.org/10.1080/0020174X.2023.2278032>
- Norton, J. D. (2015). Replicability of experiment. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 30(2), 229.
<https://doi.org/10.1387/theoria.12691>
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: Ii. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615–631. <https://doi.org/10.1177/1745691612459058>
- Nosek, B. A., & Errington, T. M. (2017). You are making sense of replications. *eLife*, 6, e23383. <https://doi.org/10.7554/eLife.23383>
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLOS Biology*, 18(3), e3000691. <https://doi.org/10.1371/journal.pbio.3000691>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, 73(1),

- 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, 7(6), 615–631. <https://doi.org/10.1177/1745691612459058>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Peels, R., & Bouter, L. (2018). The possibility and desirability of replication in the humanities. *Palgrave Communications*, 4(1), 95. <https://doi.org/10.1057/s41599-018-0149-x>
- Peels, R., & Bouter, L. (2023). Replication and trustworthiness. *Accountability in research*, 30(2), 77–87. <https://doi.org/10.1080/08989621.2021.1963708>
- Penders, B., de Rijcke, S., & Holbrook, J. B. (2020). Science’s moral economy of repair: Replication and the circulation of reference. *Accountability in Research*, 27(2), 107–113. <https://doi.org/10.1080/08989621.2020.1720659>
- Penders, Holbrook, & de Rijcke. (2019). Rinse and Repeat: Understanding the Value of Replication across Different Ways of Knowing. *Publications*, 7(3), 52. <https://doi.org/10.3390/publications7030052>
- Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, 334(6060), 1226–1227. <https://doi.org/10.1126/science.1213847>
- Peng, R. D., Dominici, F., & Zeger, S. L. (2006). Reproducible Epidemiologic Research. *American Journal of Epidemiology*, 163(9), 783–789. <https://doi.org/10.1093/aje/kwj093>
- Peterson, D., & Panofsky, A. (2021). Self-correction in science: The diagnostic and integrative motives for replication. *Social Studies of Science*, 51(4), 583–605. <https://doi.org/10.1177/03063127211005551>

- Peterson, D., & Panofsky, A. (2023). Metascience as a scientific social movement. *Minerva*, 61(2), 147–174. <https://doi.org/10.1007/s11024-023-09490-3>
- Plant, A. L., & Hanisch, R. J. (2020). Reproducibility in Science: A Metrology Perspective. *Harvard Data Science Review*, 2(4). <https://doi.org/10.1162/99608f92.eb6dde4>
- Plessner, H. E. (2018). Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Frontiers in Neuroinformatics*, 11, 76. <https://doi.org/10.3389/fninf.2017.00076>
- Pratt, M. G., Kaplan, S., & Whittington, R. (2020). Editorial Essay: The Tumult over Transparency: Decoupling Transparency from Replication in Establishing Trustworthy Qualitative Research. *Administrative Science Quarterly*, 65(1), 1–19. <https://doi.org/10.1177/0001839219887663>
- Romero, F. (2017). Novelty versus Replicability: Virtues and Vices in the Reward System of Science. *Philosophy of Science*, 84(5), 1031–1043. <https://doi.org/10.1086/694005>
- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, 14(11), e12633. <https://doi.org/10.1111/phc3.12633>
- Rubin, M. (2024). Questionable metascience practices. *Journal of Trial and Error*, 4(1), 5–20. <https://doi.org/10.36850/mr4>
- Schickore, J. (2011). The Significance of Re-Doing Experiments: A Contribution to Historically Informed Methodology. *Erkenntnis*, 75(3), 325–347. <https://doi.org/10.1007/s10670-011-9332-9>
- Schickore, J., & Hangel, N. (2019). “It might be this, it should be that...” uncertainty and doubt in day-to-day research practice. *European Journal for Philosophy of Science*, 9(2), 31. <https://doi.org/10.1007/s13194-019-0253-9>
- 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>

Schöch, C. (2023). Repetitive research: A conceptual space and terminology of replication, reproduction, revision, reanalysis, reinvestigation and reuse in digital humanities.

International Journal of Digital Humanities, 5(2–3), 373–403.

<https://doi.org/10.1007/s42803-023-00073-y>

Shapin, S. (1984). Pump and Circumstance: Robert Boyle's Literary Technology. *Social Studies of Science*, 14(4), 481–520. JSTOR.

Simons, D. J. (2014). The Value of Direct Replication. *Perspectives on Psychological Science*, 9(1), 76–80. <https://doi.org/10.1177/1745691613514755>

Stodden, V., Guo, P., & Ma, Z. (2013). Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. *PLoS ONE*, 8(6), e67111. <https://doi.org/10.1371/journal.pone.0067111>

Talkad Sukumar, P., & Metoyer, R. (2019). *Replication and Transparency of Qualitative Research from a Constructivist Perspective* [Preprint]. Open Science Framework. <https://doi.org/10.31219/osf.io/6efvp>

Ulpts, S., & Schneider, J. W. (2023). *Knowledge Production Modes: The Relevance and Feasibility of 'Reproducibility.'* <https://doi.org/10.31222/osf.io/ujnd9>

Ulpts, S., Bartscherer, S. F., Field, S. M., & Penders, B. (2025, March 13). The social replication of replication: Moving replication through epistemic communities. https://doi.org/10.31235/osf.io/pqc4v_v1