**Knowledge Production Modes:**

**The Relevance and Feasibility of  
‘Reproducibility’**

Sven Ulpts1\*, Jesper Wiborg Schneider1

1. Danish Center for Studies in Research and Research Policy, Department of Political Science, Aarhus University

\* Corresponding author. Email: [su@ps.au.dk](mailto:su@ps.au.dk)

**Introduction**

In the wake of the proclaimed ‘reproducibility crisis’ in science (Baker, 2016), there have been growing and impactful discussions about the supposed paramount importance of ‘reproducibility’. These claims about a so called ‘crisis’ largely originate from challenges and an apparent lack of reproducibility of empirical studies in some behavioral and biomedical fields (Fidler & Wilcox, 2021). However, there is some dispute about where and whether such challenges surmount to a ‘crisis of reproducibility’ (Fanelli, 2018; Guttinger, 2020). Nevertheless, the discussions have culminated in strong advocacy for normative actions to address issues of ‘reproducibility’ (Guttinger, 2020; Penders et al., 2019). These actions are sometimes referred to as a ‘replication drive’ with policy changes aimed at fostering or even mandating ‘reproducibility’ of all research (Bissell, 2013; Drummond, 2019; Penders, 2019).

The notion of ‘reproducibility’ seems to emanate from experimental sciences and perceived ideals of ‘the scientific method’ and ‘self-correction’ of science (e.g., Derksen, 2019; Earp & Trafimow, 2015; Gundersen, 2021; Peterson and Panofsky, 2021; Haig, 2022). Hence, what is often neglected is that reproducibility is rooted in specific epistemic traditions. The claim of the universal appropriateness and necessity of ‘reproducibility’ is not without its critics. For example, qualitative researchers in the social sciences have debated the issue for some time (Bennett, 2021; Büthe & Jacobs, 2015; Pownall, 2022; Pratt et al., 2020; Tuval-Mashiach, 2021). And when Peels and Bouter (2018) advocated for the desirability of ‘reproducibility’ in the ‘humanities’, an important debate emerged about how epistemic diversity in research affects the suitability of ‘reproducibility’ (e.g., de Rijcke & Penders, 2018; Leonelli, 2018; Devezer et al., 2019; Holbrook et al., 2019; Peels, 2019; Penders et al., 2019,2020; Guttinger, 2020; Sikorski & Andreoletti, 2023). Epistemic diversity refers to the circumstance that there is not just one way to acquire and justify knowledge, but rather a diversity of approaches to produce and justify knowledge. Furthermore, Guttinger (2020) in terming this debate the ‘new localism’ highlighted how the appropriateness of ‘reproducibility’ depends on the unique conditions of the specific research situation (Leonelli, 2022). Therefore, a universal ’replication drive’ is not unproblematic. Penders et al. (2019, 2020) warn about the potential to commit ‘epistemic injustice’(Fricker, 2007), if reproducibility is incentivized or mandated where it is inappropriate. Likewise, Malich and Rehmann-Sutter (2023) warn about the ‘homogenizing tendency’ of the science reform movement in psychology due to their narrow focus on hypothetico-deductivism and inferential statistics (also see Haig, 2022).

Unfortunately, another aspect further complexifying the discussions about ‘reproducibility’ is that the terms ‘reproducibility’ and ‘replication’ are not clear either. In fact, considerable conceptual confusion exists across, but also within, fields and research approaches (see e.g., Barba, 2018; Gomez et al., 2014; Gundersen, 2021; Matarese, 2022; Plesser, 2018)[[1]](#footnote-1). Several attempts have been made to provide uniform typologies, often by use of qualifiers, and primarily from a statistical and (post)-positivist perspective (see e.g., Goodman et al., 2016). As part of the debate around concerns regarding ‘reproducibility’ in the realm of epistemic diversity, a few attempts, have been made to provide a more localized perspective on reproducibility (Guttinger, 2020; Leonelli, 2018; 2022; Penders et al., 2019, 2020; Tuval-Mashiach, 2021). Attempts at framing reproducibility according to types of research based on methods or fields are welcome and inspiring. However, that focus is problematic as they seem to neglect that the relevance of ‘reproducibility’, regardless of how it is defined, is first and foremost linked to epistemic traditions that can differ within fields and methods. For instance, Issues of ‘reproducibility’ are ceteris paribus not irrelevant for qualitative research conducted from a post(positivist) position, although it may be, if the research is conducted from a constructivist or interpretivist position (Tuval-Mashiach, 2021). Likewise, individual disciplines can address the same subject matter but from different epistemic positions and often without properly engaging with each other (e.g., Robins, Gosling & Craik, 1999). It therefore seems that neither methods nor disciplines or fields are suitable entities when it comes to assessing to what extent ‘reproducibility’ may be relevant to a particular kind of research in a given context. In the present paper, we therefore suggest knowledge production modes (KPM) as a more suitable entity that captures crucial aspects of epistemic diversity that influence the appropriateness of ‘reproducibility’ for diverse kinds of research. KPM encapsulate both the epistemic and social aspects of knowledge production and are local in the sense that they are organized around a subject matter, an epistemic position, and preferred methodologies within a specific research situation. Moreover, KPMs can form parts of research specialties.

A further issue which seems underdeveloped in current attempts to address reproducibility is the actual feasibility. While reproducibility may be relevant, it may still be unfeasible. Since the feasibility may vary considerably according to many dimensions on which the actual practices of knowledge production depend. Therefore, in our KPM framework we distinguish between the relevance and feasibility of ‘reproducibility’. Our aim is to extend the discourse around ‘reproducibility’, not just by stressing the subtle complexities, but also by providing an analytical framework that can support notions of epistemic diversity.

We first review existing attempts at assessing the appropriateness of ‘reproducibility’ in the realm of epistemic diversity. We selected Leonelli’s (2018), Penders et al’s. (2019), Guttinger’s (2020), and Tuval-Mashiach’s (2021) work to present the current situation and motivate the need for our framework. Due to the immense variation in 'reproducibility' practices and the aforementioned conceptual confusion, before outlining our KPM framework we present an approach to identifying characteristics of types of 'reproducibility' that are necessary to assess the appropriateness of 'reproducibility, Then we will present our knowledge production modes (KPM) framework for the assessment of the relevance and feasibility of ‘reproducibility’ across the diversity of research. Lastly, we will provide some questions to ask for a case-by-case assessment of the appropriateness of ‘reproducibility’ for individual studies.

**Review**

Leonelli (2018), questioning the appropriateness of ‘reproducibility’ as a universal epistemic criterion developed a six-part spectrum for different kinds of ‘reproducibility’ with different interpretations, practices, and functions.[[2]](#footnote-2) In her taxonomy, the meanings, depend on the degree of control, standardization, and reliance on statistical inference. Furthermore, she allocates these different kinds of reproducibility to different types of research that, according to her framework, have or allow for specific combinations of control, standardization, and reliance on statistics. Leonelli (2018), starts with computational reproducibility: “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 2016, p. 12)” (Leonelli, 2018, p. 7). Leonelli (2018) ascribes this to ‘software development’, an activity she classifies as a type of research characterized by (allows for) total control over the environment and high reliance on statistical inference. She locates this type of research in the fields of software engineering and informatics. Then follows direct reproducibility, obtaining the same results by applying the same methods again. Leonelli (2018) deems direct reproducibility appropriate for so-called ‘standardized experiments’ in high energy particle physics and clinical drug trials. The claim is that experiments in these fields supposedly have similarly high degree of control and standardization, and especially drug trials rely heavily on statistical inference (Leonelli, 2018). Next up are three kinds of reproducibility linked to semi-standardized experiments where conditions are presumably less tightly controlled with a more variable reliance on statistics. These are (i) scoping reproducibility, reproducing to explore sources of variation potentially important for interpreting the outcome; (ii) indirect reproducibility; and (iii) hypothetical reproducibility (see Romero, 2017). Followed by reproducible expertise in non-standardized experiments and research on rare materials where the same result should be achievable by researchers who possess the necessary skills or expertise. This kind of reproducibility is applicable to, for instance, history and paleontology (Leonelli, 2018). Finally, Leonelli (2018) identifies irreproducible research for participant observations, in for instance anthropology, due to the inherent subjectivity and context dependent nature of this research type, as well as the idiosyncratic features and ‘situatedness’ of findings. It seems her framework suggests that with increasing reliance on statistical inference, control, and standardization, kinds of reproducibility that demand more uniformity become more and more appropriate.

We identify several critical issues with Leonelli’s (2018) framework for our purpose. According to Penders et al. (2019) the framework neglects the issue of epistemology. Most types of research she refers to can be conducted with the same or different epistemological stances which are paramount for determining the appropriateness of ‘reproducibility’. Furthermore, there are issues with the criteria standardization and control. For example, Leonelli (2018) does not consider the so-called ‘standardization fallacy’ (Guttinger, 2020; Voelkl & Würbel, 2016; Würbel, 2000), which states that the belief that with ever more standardization, reproducibility would always be improved, is false. For instance, in the biological sciences, obtaining consistent findings can often be achieved by introducing (systematic) variation or ‘heterogenization’ instead of standardization, because increasing standardization and control usually simultaneously decreases external validity (Voelkl & Würbel, 2021). Such an unreflective focus on standardization does neglect the influence of the nature of the subject of investigation. This seems problematic because the possible amount of control and standardization are not just determined by understanding or the amount of accumulated knowledge, but also by, how stable or context- and time-sensitive the subject is. Moreover, standardization and control are not epistemic practices or criteria that are relevant across the research landscape but seem to have been chosen with experimentation in mind (e.g., to increase internal validity). Making a framework that attempts to capture epistemic diversity across vastly different domains, is questionable if the main criterion is alien to some research types. This focus on standardization and control might originate from an underlying uniformity assumption regarding ‘reproducibility’ in experimental research. However, this assumption might be problematic when the subject of investigation is not uniform or the type of research does not seek uniformity (see e.g., Guttinger, 2020; Schmidt, 2009).

Like several others, Leonelli (2018) also introduces a set of idiosyncratic definitions of ‘reproducibility’ adding to the already existing conceptual confusion surrounding these contested terms (Barba, 2018; Borges, 2022). Moreover, it is difficult to see how ‘software development’ could be a uniform research type allowing for the possibility of total control considering concerns about ‘reproducibility’ in this type of research e.g., issues related to the dependency on software and computing environments (see e.g., Peng & Hicks, 2020). Additionally, Leonelli (2018) states that direct reproducibility is about the sameness of results based on the sameness of methods and views it as appropriate for clinical drug trials. Although, then she confusingly in the same paragraph states that sameness is not to be expected in clinical science. In addition, while the idea of ‘reproducible expertise’ (and ‘reproducible observations’) might be useful for qualitative approaches with positivistic and nomothetic leanings, it seems questionable for idiographic interpretivist or constructivist qualitative investigations. Here it is usually less about the ‘redoing’ of some investigation, or obtaining similar results based on some kind of expertise, but more about enabling the larger community to understand and trace where a specific interpretation or conclusion is coming from. Büthe and Jacobs (2015) call this replication-in-thought. It is also unclear what this expertise actually is. For example, is it understanding of the subject of investigation, of the methods, or of both?

Essentially, Leonelli’s framework does not assess the relevance of different kinds of reproducibility for different practices and purposes of knowledge production but focuses mostly on the feasibility provided by research characteristics, such as, control, standardization, and reliance on statistical inference linked to six vaguely defined types of research. Control and standardization are aspects of the research practices that can work towards certain kinds of ‘reproducibility’ because they provide some sorts of uniformity, but they do not provide indication regarding their relevance for specific kinds of research or their goals. Lastly, Leonelli’s framework is unsuitable for our purpose because it is enumerative focusing on vaguely defined entities and ‘types of research’ which does not allow for an empirical case-by-case assessment of individual studies.

Penders et al. (2019) warn about the potential danger of epistemic injustices perpetrated if universal ‘replication drives’ are applied to kinds of research for which reproducibility might be inappropriate. For example, if funders or publishers incentivize or stipulate general ‘reproducibility’ requirements, certain kinds of research for which those demands are not appropriate will be unjustifiably disadvantaged and the researchers’ epistemic authority will be restricted. Practically such research and their researchers will therefore be excluded by refusing funding and publication space. For instance, in Nosek’s (2019) culture change pyramid the higher up you go, from making it possible through the infrastructure to making it required by policy, the more potentially harmful it can be to misunderstood or neglected types of research if potential differences in kinds of research that affect the appropriateness of ‘reproducibility’ are not accounted for.[[3]](#footnote-3) To evaluate the appropriateness of ‘reproducibility’ across the diversity of research Penders et al. (2019) extend Leonelli’s (2018) framework referring to Knorr-Cetina’s (1999) concept of epistemic cultures. In their view, Leonelli (2018) did not acknowledge that hermeneutical social sciences and humanities research belong in her last three ’types of research’ (‘non-standard experiments & research based on rare, unique, perishable, inaccessible materials’, ‘non-experimental case description’ and ‘participant observation’). Penders et al. (2019) also separate ‘non-standard experiments’ and ‘research based on rare, perishable, and inaccessible materials’ because those relate differently to ‘reproducibility’. Furthermore, in reference to Collins (2016, 2019) they add the research type of ‘immersion’ to account for differences in ‘reproducibility’ between qualitative fieldwork and Leonelli’s (2018) ‘participant observations’. More specifically, this allows for a consideration of epistemological differences in qualitative research approaches (Penders et al., 2019). Penders and colleagues (2019) also do not put clinical drug trials and experiments in high energy physics in the same category, as the epistemic culture of biomedicine is classified as part of the life sciences with the research type of ‘semi-controlled experiments’, while ‘controlled experiments’ in physics are seen as part of the exact sciences. Nevertheless, they maintain the criteria of control and reliance on statistics, but also add the criterion of expected uniformity to their framework, claiming that increased expectancy of uniformity in specific epistemic cultures makes ‘reproducibility’ more possible or appropriate.

Although Penders et al. (2019) are more nuanced in capturing the influence of epistemic diversity on the appropriateness of ‘reproducibility’ and highlight the significant role epistemology plays, they inherit the problems of Leonelli’s framework. Due to their focus on major fields and approaches (as epistemic cultures) they also end up with an enumerative framework that does not allow for a case-by-case assessment of the appropriateness of ‘reproducibility’ for individual studies in the realm of epistemic diversity. However, we acknowledge that Penders and colleagues (2019) are open about the inherent and inevitable limitations of any taxonomy that can be proposed for ‘reproducibility’, as they state: “Of course, epistemic cultures grow, change and evolve; they diverge, converge, merge or spawn new approaches. Epistemic cultures can be delegitimized and legitimized (e.g., the delegitimation of phrenology and the legitimation of gender studies). Distinctions of entities, as in other taxonomies, are in part, arbitrary and require consensus. Therefore, any taxonomy of replicability is difficult, situated, contestable and controversial—as it should be.” (Penders et al., 2019, p. 7).

Guttinger’s (2020) seeks a local approach of situating the appropriateness of ‘reproducibility’ in the actual research situation. Guttinger (2020) makes a plea for a case-by-case assessment (in semi- and non-standardized experiments) of the suitability of ‘reproducibility’ as an epistemic norm based on the kind of research question(s), the experimental (research) setup, and the nature of the subject of investigation. Importantly, he rejects existing distinctions for the nature of the subject such as living versus non-living (see e.g., Nadin, 2018), or indifferent versus interactive (see e.g., Penders et al., 2019), and suggests to characterize the nature of each subject of investigation as well as its complexity based on its plasticity and historicity (also see Schmidt, 2009). Historicity being the time-dependency of the subject of investigation and how much it changes over time and plasticity being the context sensitivity and interaction of the subject with its environment (Guttinger, 2020).

Similarly, Tuval-Mashiach (2021), suggests four research aspects to assess the appropriateness of kinds of ‘reproducibility’ for qualitative research: (i) the type of ‘reproducibility’, (ii) the researchers’ epistemological stance, (iii) the function or purpose of ‘reproducibility’, and (iv) the nature of the study. In the context of the conceptual confusions, the immense plurality, as well as the diversity of meanings and functions of ‘reproducibility’, an important aspect of Tuval-Mashiach’s (2019) work is the explicit consideration of the type of ‘reproducibility’. It is meaningless to ask about or speak of the relevance, feasibility, desirability, appropriateness, importance, or applicability of ‘reproducibility’, without specifying what kind of ‘reproducibility’ one is referring to. Moreover, considering Penders et al.’s warnings of epistemic injustice, what seems to be needed is a more fine-grained assessment that does not rely on some simplistic notion of a ‘hierarchy of science’ (see e.g., Peterson & Panofsky, 2021) or even perpetuates a hegemony of a ‘hierarchy of science’. Hence, what is still missing is a workable framework capturing the characteristics of this diversity in knowledge production that are influencing the relevance and feasibility of different kinds of ‘reproducibility’ for different kinds of research conducted in individual studies.

Therefore, in this paper we propose the knowledge production modes (KPM) framework that allows navigation through the aforementioned conceptual confusion and enables an assessment of the relevance and feasibility of ‘reproducibility’ for individual studies. The aim is not to enumerate an exhaustive list of KPMs within or across research fields, but to present a manageable framework that enables identification of characteristics in a study to examine the relevance and feasibility of ‘reproducibility’.

**Kinds of ‘reproducibility’**

‘Reproducibility’ is not just one practice, nor is it intended for just one function. There is an immense variation in practices and functions (see e.g., Peterson & Panofksy, 2021).[[4]](#footnote-4) Furthermore, there seems to be a general conceptual confusion and lack of nomenclature about the ‘reproducibility’ terminology within as well as across disciplines and research approaches (Borges, 2022). Different (re-)terms refer to the same practices and functions while at the same time the same (re-)terms refer to completely different practices and functions (see e.g., Barba, 2018; Matarese, 2022). Hence, without specifying what is actually meant with ‘reproducibility’ it is practically meaningless to speak of the appropriateness or importance of ‘reproducibility’ (see e.g., Tuval-Mashiach, 2021). Therefore, to use our framework it is necessary to clarify critical aspects of what is meant with ‘reproducibility’. Specifically, people do not have to formulate their own definitions. However, we propose two crucial components of the meaning of ‘reproducibility’ that have to be explicated to be able to assess the relevance and feasibility. To evaluate the feasibility of different kinds of ‘reproducibility’, there needs to be information about whether it is referring to potentially “enabling” of redoing and traceability of what was done or the “actual redoing” of an original study. This is somewhat reminiscent of Matarese’s (2022) distinction between enabling and a replication test, and Peels’ and Bouter’s (2018) distinction between a replication (study) and replicability.

One aspect of “Enabling” are the different practices that provide understanding and traceability of what has been done or where an interpretation, meaning or conclusion comes from in an original study. Such practices of enabling could in principle enable someone to redo a study (this is sometimes called (methods, computational) reproducibility, replicability, (production & analytical) transparency, replication-in-thought, virtual witnessing, thick (rich) description or even, reflexivity). Another aspect of “enabling” can be the provision and availability of the resources or skills that are necessary to conduct a specific (kind of) investigation (e.g., code, data, software, instruments, technology, material, skills, etc.,). Depending on the context and the characteristics of a study, enabling through sharing has of course numerous, legal, practical, epistemological, and ethical barriers which have to be considered when assessing the feasibility (Field et al., 2022). It is inappropriate to claim that with increasing degree of enabling (e.g., transparency and openness) the probability of successful redoing also increases. The amount and specific nature of enabling that is necessary depends on the specific kind of redoing (or traceability one is after). Hence, the feasibility of redoing has a case-by-case link to enabling and not some general rule of open science and ‘reproducibility’ being two side of the same coin (Baumgaertner et al., 2018; Buzbas et al., 2023). For redoing we need to know what kind of practices are actually involved, what is expected to be similar or the same and what (can) vary to determine their feasibility. See Table 1 for a color scheme for an example of different combinations of components that might be redone either identical, similarly, or with variation.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Investigators**  Are the investigators supposed to be the same? | **Theory**  Is the same or similar theory tested? | **Hypothesis**  Is the same or similar hypothesis tested? | **Methods**  Are the same or similar, methods redone? | **Operationalization**  Are the same or similar operationalizations and measurements used? | **Ancillary**  Are the underlying assumptions similar or the same? | **Data**  Is new data generated or the same being used? | **Result**  Is the outcome supposed to be the same or similar? | **Interpretation**  Are the conclusions and interpretations supposed to be similar or the same? | **Analysis**  Is analysis the same or different? | **Environment**  Is the investigation carried out in a similar or the same lab, culture, country? |
| 1 |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |

*Table 1- color scheme of different forms of redoing based on example combinations of similar same or varying aspects of the research setup in comparison to the original study. black indicates keeping something the same, gray indicates that something should be similar while white indicates variation (not redone). Inspired by Gundersen (2021).*

To assess the relevance of different kinds of ‘reproducibility’ one has to be, similar to Matarese’s (2022), as well as Schmidt’s (2009), functional approach, clear about the function or purpose of ‘reproducibility’. Hence one has to link the specific combination of sameness, similarity, and variation of the aspects of a study to the intended function. This specification could allow to assess whether ‘reproducibility’ is actually relevant for a specific study. For instance, is the purpose to establish the existence of a phenomenon, stability, various kinds of validity, reliability, generalizability, or reduce errors? Furthermore, it is not just important to clarify what the actual purpose of a specific instance of redoing or enabling is, but also what part of the study this goal applies to, the instruments, a specific method, the whole study, a specific knowledge claim, or a numerical result?

We opt for this approach, letting users select their understanding based on these two crucial components as well as allow for the freedom of new definitions, due to the elusive nature and conceptual confusion in the ‘reproducibility’ terminology (Barba, 2018; Borges, 2022; Matarese, 2022; Penders et al., 2019). Furthermore, looking at the literature new definitions and understandings of ‘reproducibility’ will inevitably emerge (see e.g., Huijnen et al., 2022; Schöch, 2023; Talkad Sukumar & Metoyer, 2019). This approach leaves space for new directions and developments open instead of restricting the evaluation on specific understandings of ‘reproducibility’. Hence, any proposed new general typology aiming to capture ‘reproducibility’ for diverse research types will probably quite quickly become obsolete. Our approach also allows researchers, if wanted, to come up with their own understanding that can be more adapted to the specific kind of research, context, and the intended purpose. Moreover, this inherent flexibility and focus of our approach on practices and functions might even allow researchers from different backgrounds to recognize practices and criteria that are similar but more appropriate. These similar but different practices and criteria are often referred to by different terms like triangulation and crystallization or transferability and dependability (Pownall, 2022; Talkad Sukumar & Metoyer, 2019). Consequently, whether research quality is already demonstrated can be hidden behind differences in language regarding epistemic criteria and practices (Pownall, 2022). Hence our approach of explicating practices and functions might allow people to recognize already established potentially more appropriate practices like triangulation and crystallization that fulfill similar, but not necessarily the same functions in qualitative research as ‘reproducibility’ in many quantitative and positivist approaches. Instead of pushing such practices aside or putting a redundant workload on some researchers in form of a universal replication drive, we aim to enable people to recognize whether research quality has already been established by different means.

In other words, our approach might prevent criteria and practices like ‘reproducibility’ from being pushed into research approaches where they do not belong and replacing criteria and practices that are more appropriate or adapted for ensuring accountability or trust (Guttinger, 2020; Penders et al., 2019). Hence, when used with care our approach could counteract the danger of epistemic injustices due to an overly general, mindless, and performative ‘replication drive’. Instead of providing definitions of ‘reproducibility’ for different kinds of research like some previous approaches (Leonelli, 2018; Penders et al., 2019), our framework is applied to existing definitions or can be applied to new ones. Then based on the components of the KPM-framework targeted questions for the assessment of relevance and feasibility can be posed.

**Knowledge Production Modes**

Inspired by earlier work, but in contrast to such approaches, our knowledge production mode (KPM) framework allows us to distinguish between relevance and feasibility of ‘reproducibility’ for individual studies[[5]](#footnote-5). To prevent the potential danger of unjustified negative consequences in the form of funders or journals demanding ‘reproducibility’ from researchers conducting research for which such demands are epistemically irrelevant or practically unfeasible we need a situated understanding of how knowledge production and justification relate to ‘reproducibility’. Therefore, we think there is a need to be able to assess on a case-by-case basis when ‘reproducibility’ is more or less relevant and feasible for specific studies in specific contexts (Guttinger, 2020; Leonelli, 2022; Penders et al., 2019; Steltenpohl et al., 2023).

The relevance is mostly determined by whether the functions or purposes of ‘reproducibility’, like types of validity, reliability, establishing stability, generalizability, or error reduction are appropriate for the KPM of a study. The components of a KPM used to assess the relevance of ‘reproducibility’ are the epistemology, system of justification, kind of knowledge claim, and the research goal behind a study. However, considering relevance alone is not sufficient. Even in cases where reproducibility might be relevant that does not necessarily imply that it is feasible to expect or demand reproducibility. Therefore, the feasibility of ‘reproducibility’ is assessed based on the nature of the subject of investigation, the research setup and its resource dependency/ availability as well as the uncertainty. See Figure 1 for the components of the KPM-framework.[[6]](#footnote-6)

A diagram of knowledge

Description automatically generated

Fig 1 – Illustrating the components of the KPM-framework to assess the relevance and feasibility of ‘reproducibility’.

Kuhn (1970) used the term ‘disciplinary matrix’, and one might be tempted to call our KPM framework a knowledge production matrix. A KPM is characterized by the interdependencies between epistemologies, ontologies, methods, practices, criteria, and conditions to create and evaluate knowledge claims within and/ or across fields with a prominent role being played by systems of justification as a crucial (gateway) node within that network of interdependencies.

Crucially, methods and practices are not just shaped by, but also often contain certain epistemological and ontological assumptions that are at least partially influenced by the nature of the subject of investigation due to the KPMs’ and its components (continued) reciprocal interaction with that subject. Hence, KPMs are to an extent structured around the subjects of investigation. The outcome of a KPM is a specific kind of knowledge claim, which is not just the result of a KPM, but also guides it in a certain way since the kind of knowledge claim one is after is in turn influenced by the epistemology and the research goal of a study.

KPMs unfold within and across research specialties and are embedded in dynamic institutional settings driving the social system of science (e.g., norms, publication, funding, communication, reward structures etc.). Hence, there is a mutual dependence between the epistemic and social aspects of knowledge production. Similar to Kuhnian paradigms (1962), KPMs with markedly different epistemologies are essentially incommensurable in the sense that aspects such as research goals or epistemic criteria and values, as well as kinds of knowledge claims produced, can be radically different. KPMs can be more or less typical for a specific field or discipline. Importantly, this organization is dependent on and interacts with the particular context they are situated in. Hence, KPMs are situated and do not exist in some kind of clean neutral vacuum, but specific dependencies have to be accounted for.

**How are Knowledge Claims Justified?**

**Epistemology**

Epistemology is the theory of knowledge and determines in what kind of system of knowledge one is practicing research.[[7]](#footnote-7) The epistemological stance one is taking determines what is supposed to be knowledge, what kinds of knowledge claims are possible, how to produce knowledge and how to justify knowledge claims. In other words, epistemology is functioning as a value system guiding knowledge production and justifying knowledge claims. There are some more or less established epistemological traditions and we list some of them here, for example, (post-)positivism, (social) constructivism/ interpretivism and pragmatism (see e.g., Guba & Lincoln, 1994; Moon & Blackman, 2014). We will not provide clear-cut definitions, because in actual research practice epistemologies are probably never as clean or explicit as defined in textbooks. Therefore, we will provide rough orientations regarding what each of these more common epistemological positions mean.

* (Post-)positivism assumes that there is a singular mind and researcher independent truth in reality that can be discovered or at least approximated through logic and rigorous methods. Hence, there is a duality with the researcher and the researched being somewhat separate entities during the research processes. Research conducted under the positivistic positions tends to thrive for some kind of objectivity and the elimination of biases which can be introduced by the researcher.
* (Social) constructivism and interpretivism presuppose that knowledge is not discovered, but rather (co)constructed or even negotiated between the different agents that are involved in the research processes, such as the researcher and the researched. Consequently, in constructivism and interpretivism there is no clear separation between researcher and researched regarding the question who produced the knowledge. Relatedly, constructivist and interpretivist research usually does not thrive for objectivity but acknowledges the role of subjectivity during the process of knowledge construction. It is not about finding or getting closer to some underlying truth, but about understanding individual or collective meanings and lived experiences. Therefore, interpretivist and constructivist epistemologies focus on situating knowledge in the historical, social, and cultural contexts. Ideas such as consensus or unification are uncommon, and diversity or pluralism of meanings, experiences and interpretations are encouraged. Some argue that it is even questionable whether concepts such as bias make sense in interpretivist and constructivist research (Yanow & Schwartz-Shea, 2009).
* Pragmatism can be traced back to American scholars such as, William James, John Dewey and Charles Sanders Peirce. Pragmatism is not only about the practical implications and necessities of conducting research, but it is often also seen as a compromise where all the epistemological and ontological positions should be used that seem to work the best for a specific purpose or help to understand the problem under investigation (see e.g., Legg & Hookway, 2021).

The distinction between these epistemological traditions also illustrates that epistemologies are axiological, meaning that they are value systems for kinds of knowledge (see e.g., Carter & Little, 2007; Shan, 2022). It is crucial to appreciate the differences between epistemologies, because while some might not be mutually exclusive there are still certain differences regarding kinds of knowledge claims and epistemic values that can lead to incommensurabilities that have to be considered. This can be especially important for pragmatism because discussions about pragmatism are often held in the context of mixed methods research. Hence, here different epistemological stances could collide and neglecting those differences can impede the research quality by using practices or values that are inappropriate. As Penders et al. (2019) hint at, the epistemological orientation underlying a study not just critically influences whether ‘reproducibility’ is relevant, but also which *kind* of ‘reproducibility’ might be appropriate (see also Tuval-Mashiach, 2021). The different underlying epistemologies set different criteria and emphasize particular aspects of justification, but also determine how the systems of justification (see below) are applied.

Knowledge claims refer to statements or assertions made by scholars based on their research. Depending on the KPM and especially the epistemology. Importantly, there are different kinds of knowledge claims that can be made. Penders et al. (2020) distinguish between ‘matters of fact’ and ‘matters of concern’. Research with positivist leanings tends to aim for ‘matters of fact’ to find consensus and accumulating knowledge working towards the ‘truth’ while in more hermeneutic, constructivist and interpretivist research endeavours, the knowledge claims are more like ‘matters of concern’ sparking a discussion and fostering a plurality of diverse positions. Relatedly, de Rijcke and Penders (2018) also talk about the difference in kinds of knowledge claims concerning research aiming for meaning(s) or for truth(s). Hence, knowledge claims can be about consensus, truth, diversity, plurality, meanings, lived experiences, opinions, or style and each of those have different relations to ‘reproducibility’ (see also Guttinger, 2020). Importantly, two different KPMs can lead to similar or even identical knowledge claims, but they are arrived at through a different combination of the KPM components and can be assessed based on different epistemic criteria. The kinds of knowledge claims can be especially important to be able to identify the underlying epistemology of a study.

**Systems of justification**

Epistemologies are formal, abstract, and often implicit logics of knowledge production, but in actual practice these epistemologies transform into, as Carter and Little (2007) call it adapted from Kaplan (1964), logic-in-use. There is not just one way to justify knowledge claims. Instead, there is a variety of more or less traditional epistemic criteria (values & norms) which are used to evaluate the quality and trustworthiness of research thereby justifying the knowledge claims that are being produced. Research quality here means to what degree peers regard the knowledge claims as substantiated according to specific epistemic criteria and practices. Penders et al. (2019) call these constellations of practices and epistemic criteria politics/ technologies of accountability, while Guttinger (2020) calls them trust establishing practices. We call them systems of justifications inspired by the concept of justified beliefs in the philosophy of knowledge, because they aim to ensure that the knowledge claims that are being produced during research are justified according to specific epistemic quality criteria. Each kind of research has their set of appropriate epistemic criteria. These are systems and not just one system due to the prevailing epistemic diversity within and across disciplines and research approaches. Hence, the plurality and diversity of systems of justification is a result of the interaction between factors like the contexts, epistemologies, ontologies, nature of the subject, research goal, methods, practices, and the different kinds of knowledge claims. Epistemology is especially important because it not just determines which kinds of knowledge claims are possible, but it also guides the application of the systems of justification. Hence, how these practices and criteria are used and interpreted is crucially influenced by the epistemologies. See Figure 2 for an illustration of how epistemology determines and guides the justification of knowledge claims.

A diagram of a method

Description automatically generated

Figure 2 – inspired by Carter & Little (2007) conceptualization of how epistemology and the system of justification determine as well as justify the resulting kinds of knowledge claim. This depiction also illustrates that the application of methods and the system of justification are determined by the specific epistemological stance.

An example of epistemic quality criteria are the four characteristics for naturalistic and qualitative inquiry presented by Guba and Lincoln (1985): credibility, dependability, confirmability, and transferability. Importantly, this is not the credibility people talk about in the context of the credibility crisis/ revolution (see. e.g., Lincoln & Guba, 1985; Pratt et al., 2020; Yadav, 2022). Credibility here is about how well (accurate) specific realities are represented. Credibility is usually achieved or demonstrated through persistent observation, member checking, prolonged engagement, triangulation, or peer debriefing. Another example are the various kinds of validity, for instance construct validity, internal validity, and external validity in combination with different kinds of reliability in addition to generalizability for quantitative (positivistic), experimental, or laboratory research. For ethnographic research an example might be Locke and Golden-Biddle’s (1997) authenticity, plausibility, and criticality. These are all quality criteria for different kinds of research resulting in specific kinds of knowledge claims. These criteria function like signposts during the research process, representing epistemic characteristics about the quality or trustworthiness of the production process that led to specific knowledge claims. Similarly, as described for credibility, each kind of system of justification has its own set of relevant practices that are crucial for ensuring and demonstrating the quality of research in light of these signposts indicating whether the knowledge claims that are being produced are justified. For naturalistic and/ or qualitative (interpretivist and constructivist) research approaches practices such as thick descriptions, reflexivity, positionality, audit trails, triangulation, peer debriefing, member checking, persistent observation and prolonged engagement can be used to work towards the four criteria from Guba & Lincoln (1985). Traditionally, for quantitative and positivist research practices and criteria such as various kinds of ‘reproducibility’, predictive power, control, and standardization, as well as triangulation can be applied to ensure research quality and thereby justify the knowledge claims. See Figure 3 for an example of what we mean with systems of justification (for an even more extensive and different list in a qualitative research context see Yadav (2022)).

|  |  |  |  |
| --- | --- | --- | --- |
| **Tradition/ Approach** | **Criteria** | **Example Questions** | **Example Practices** |
| ***Naturalistic/ Qualitative Research***  ***(see Lincoln & Guba, 1985; Stahl & King, 2020)*** | Credibility | How well are realities or sources of data represented? | - Prolonged engagement - persistent observation - triangulation - peer debriefing - negative case analysis - referential adequacy - member checking - reflexivity - positionality - … |
| Dependability | How dependent are findings on specific aspects of the research setup or the (sampled) subject of investigation?  How consistent are findings? | - see practices for credibility plus: - stepwise replication within the dataset - inquiry audit/ audit trail - triangulation - rich description - … |
| Confirmability  (relevant in positivistic qualitative research) | How accurate does the data capture the reality? | - inquiry audit - reflexivity - triangulation - peer debriefing - … |
| Transferability | How well do findings apply to other (similar) contexts? | - Thick description of context and situation - … |
| ***Positivistic/ Quantitative Research***  ***(see e.g., Yin, 2003)*** | Construct Validity | How well is the measurement capturing the intended construct? | - use diverse set of evidence - … |
| Reliability | How well do studies, measurements or investigators converge on one outcome. | - replication - standardization - … |
| Internal Validity  (systematic error) | To what degree is the study able to identify causal links?  How much confidence can be put into causal inferences? | - control - randomization - blinding - … |
| External Validity | How well do findings apply (generalize) to different contexts. | - Representative sampling - replication - … |

Figure 3 – Inspired by Pratt et al. (2020), illustrating example systems of justification for naturalistic/ qualitative research and positivistic quantitative research.

What kind of criteria and therefore what kind of practices are appropriate can also depend on whether the research is confirmatory (deductive) or exploratory (inductive and abductive) (Haig, 2022). Similarly, it can be affected by whether the research is nomothetic or idiographic. Those distinctions can be rather heuristic because in reality research is often probably not this clear cut and messier. There are all sorts of classifications for kinds of systems of justification to ensure the quality of diverse kinds of research. Brian Haig (2022), for example, distinguishes between reliabilist and coherentist systems of justification to understand the place of replication in science. With the former aiming for reliability of methods and the later for theoretical consistency. The field of investigation influences which epistemic criteria and practices are appropriate for a specific kind of research, depending on among other things how typical (prototypical) a specific system of justification is for a specific field which is interdependent with the uncertainty (Whitley, 1984/2000). It is interdependent with the specific methodology, the epistemological stance taken during the research and the nature of the subject of investigation influencing which kind of practices and criteria are actually relevant and feasible. For instance, qualitative approaches to research can be positivist or interpretivist. Importantly, what systems of justifications and how they are appropriately applied will differ depending on whether the underlying epistemological stance is positivist or interpretivist. This is also one of the reasons why the qualitative-quantitative distinction might be a nice indication that there are different ways to produce knowledge with different resulting knowledge claims, but it is too vague an umbrella distinction to allow for an appropriate evaluation of relevance and feasibility of reproducibility and replication.

**What is the nature of the Subject?**

The nature of the subject of investigation influences how research is conducted, because it provides a set of conditions under which investigation has to take place and which aspects have to be taken into account during research. Numerous continuums, spectra, distinctions, and characteristics have been identified to distinguish and describe the nature of different subjects of investigation (see e.g., de Rijcke & Penders, 2018; Gläser et al., 2018; Guttinger, 2020; Nadin, 2018; Penders et al., 2019, 2020). Nadin (2018) for instance provides a distinction between living and non-living subjects of investigation and how that influences the practice of experimentation. De Rijcke and Penders (2018) introduce the differentiation between interactive kinds and indifferent kinds arguing that the subjects of investigation can be interacting with the researcher or indifferent. In cases of interaction, the researcher and the researched can, for instance, co-produce the knowledge. De Rijcke and Penders (2018) introduce this distinction to reject the relevance of replication as an epistemic criterion for such research about meaning and limit its relevance to truth seeking about indifferent kinds. Some argue that an important aspect to consider when distinguishing kinds of research is the subject complexity (Nadin, 2018). The assumption that the subject matter in the social sciences is more complex than in the natural sciences and living kinds are supposedly more complex than non-living kinds (Nadin, 2018) due to context dependency, causal density, interactions between researcher and researched, and the subject being in perpetual flux. This is somewhat reminiscent of what Guttinger (2020) means with historicity and plasticity of the subject matter. All else being equal, the more stable and indifferent a subject of investigation, the more feasible specific kinds of reproducibility can be.

**What is the research situation?**

**Research goal**

What the research goal is like is highly related to the motivation behind the research, the context of the research, and what the research question is. For instance, part of the research goal is whether it is about furthering knowledge or about getting a competitive advantage in something. For example, in science-industry collaborations, the goals diverge from classical depiction of research goals like knowledge accumulation since considerations like financial gain, costs, priority, confidentiality, intellectual property, and patents will influence the goal of the research (Guzzo et al., 2022). Furthermore, to use Kuhnian terms the research goal depends on the nature of the problem that is addressed or solved, on the conventions within a specific field or discipline and on what is possible with the known or available methods and practices (thought to be possible). Therefore, as industry-science illustrates ‘reproducibility’ might be relevant on an epistemological basis, but irrelevant for the ultimate goal behind the research. Hence, considerations about the goal and motivation behind a study provide important information about the situatedness of research that influence the relevance of ‘reproducibility’.

**Research setup**

The research setup comprises all the practices, methods, that are applied for knowledge production during a study. These include among others: statistical techniques, forms of experimentation, kinds of observational approaches, forms of survey research, ethnographies, text analysis, histographies, interviews, simulations, thought experiments, theorization, … etc. On a more abstract level, one can also look at the list of epistemic activities that Chang presents: “… describing, predicting, explaining, hypothesizing, testing, observing, detecting, measuring, classifying, representing, modeling, simulating, synthesizing, analyzing, abstracting, idealizing.” (Chang, 2012, p. 16). Put differently, with research setup we attempt to capture what kinds of instruments, technologies, procedures, and methods are used in specific studies.

There are certain traditional methodological approaches, such as, experimental studies, observational studies, interview studies, ethnographic studies (combination of observational and interview study), theoretical studies, and simulation studies. These different study types contain specific kinds of analysis, like statistical, hermeneutical, descriptive, textual etc. (see Szostak, 2004 for an attempt at a classification of methodological and analysis approaches). All of these studies and analyses have unique dependencies. Consequently, part of the research setup is also a consideration of how and to what degree certain methods and practices of a study depend on the availability of resources, technologies, materials, and infrastructures since those influence the feasibility of redoing certain methods and practices. Leonelli (2022) lists conceptual, theoretical, material, methodological, infrastructural, socio-cultural, and institutional diversity across the research landscape through the lens of Chang’s (2012) systems of practices as part of epistemic diversity and argues that knowledge production and especially the feasibility of open science practices depends on these characteristics of research. These dependencies include among other things the amount of funding that is required to conduct a specific type of study or the dependency of certain open science practices on the availability of kinds of technology (see e.g., Leonelli, 2022). This consideration of resource dependency of a specific research setup in the context of ‘reproducibility’ has also been called a logic of investment. The size of the necessary investment to redo an original study is determined by how much the original experiment depends on technologies, instruments, techniques, and the difficulty of capturing or creating the research object (Peterson & Panofsky, 2021). For instance, expecting that someone else has to be able to completely redo an identical experiment in high energy particle physics is often infeasible, because most researchers lack the resources and infrastructure in form of funding and specific particle accelerators to conduct such an experiment in the first place. Another scarce resource, especially in research is time, reproducibility can often no be ensured or considered due to time constrains (Schickore & Hangel, 2019). Penders and colleagues (2020) add a crucial dimension to the ‘reproducibility’ discussions, which is the continuum of KPMs operating with writing as reporting or writing as research. In other words, the spectrum from writing as the process of making knowledge and writing as witnessing or making visible knowledge that was already made (Penders et al., 2020). As Penders et al (2020) indicate, when writing is the research itself then ‘reproducibility’ as forms of redoing seems quite misspaced and irrelevant. They point out in a footnote that duplication of writing as research could even be seen as plagiarism. While ‘reproducibility’ as enabling can be very important as enabling the audience to understand, trace and follow the argumentation or logic of the writing.

**Uncertainty**

Our consideration of uncertainty is inspired by the concept of task uncertainty from Whitley (1984/2000) as well as Peterson and Panofsky’s (2021) use of task uncertainty as a conceptual lens to understand interviewees’ responses about what replication means in their research routine. Task uncertainty has two components: Strategic task uncertainty and technical task uncertainty. Whitley defines strategic task uncertainty as the degree of shared theoretical understanding about the subject of investigation which allows for standardization and control over the phenomena. Strategic task uncertainty is therefore about the ability to keep the elements of an investigation pure, homogeneous, and stable while restricting their properties and hold their features uniform. This part of strategic task uncertainty also illustrates that standardization and control are conceptualized as a consequence of understanding and not alone an indicator for understanding. Therefore, when the understanding of the subject changes the standardization has to be adapted. Hence, standardization devoid of understanding cannot be an indicator for low uncertainty. In addition to the reasons presented in our critique of Leonelli’s (2018) use of standardization in her framework, we do not incorporate Whitley’s (1984/2000) conception of standardization and control of research units and subjects of investigation into our understanding of uncertainty. This could systematically privilege certain fields and kinds of research and reinforce the hegemony of positivist approaches to research since the relevance of control 4and standardization depends on the KPM. Furthermore, it is hard to distinguish whether standardization is based on current understanding or just part of a tradition. Nevertheless, we capture those aspects with different elements of our framework.

Standardization and control are part of some systems of justification. Moreover, as Guttinger (2020) and Voelkl and Würbel (2016, 2021) indicated, to what degree standardization is possible and conducive to certain kinds of ‘reproducibility’ depends on the nature of the subject of investigation (e.g., complexity, time and context sensitivity). Therefore, we focus on the aspect of theoretical understanding of the subject of investigation. For Whitley (1984/2000) strategic task uncertainty also means the agreement about intellectual priorities and the prioritization of topics and intellectual goals, but also about what deserves and brings reputational gains.

Whitley (1984/2000) defines technical task uncertainty as the degree to which techniques, methods and procedures are understood and their ability to provide reliable results within a research community. It describes the amount of certainty about how to address specific problems, how and when to use specific procedures and methods, the degree of replicability, predictability, and visibility of results, as well as the number of conflicting interpretations and how much ambiguity is present in such interpretations. For Whitley, technical task uncertainty is about whether practices and interpretations are tacit, personal, and fluid or formal and uniform. The aspects of replicability, predictability and reliability of results and methods that are part of Whitley’s definition of technical task uncertainty are also captured differently in our KPM-framework. On the one hand, predictability and reliability are, similar to standardization and control, in our framework part of certain systems of justification and influenced by the research goal, methodology, and epistemological stance. On the other hand, ‘reproducibility’ is the epistemic criterion for which our whole framework is created to assess its relevance and feasibility through the lens of different KPMs.

Uncertainty is the characteristic to situate the feasibility of ‘reproducibility’ within the specific community of researchers who potentially are enabling or actually redoing something. Uncertainty is a community characteristic, because the degree to which theoretical understanding of the subject of investigation and the understanding of methods, which are an integral component of the research setup, are shared among researchers, depends on the characteristics of that community. Usually, that community is a research specialty. Consequently, we define methodological uncertainty as how well methods are understood as well as the degree to which understanding of methods, and how to use and justify them, is shared within a specific community. Furthermore, uncertainty in this sense is not about some average uncertainty regarding all the methods and practices within a specific field, but the other way around. It is about the uncertainty within the specific community that the authors are addressing regarding the methods and practices that are used in a specific study. In turn, we define theoretical uncertainty capturing how well the subject is understood as well as whether and to what degree this theoretical understanding is shared in a community and allows to guide investigation and interpretations. The aspect of theoretical understanding can provide guidance and make an investigator, as well as the application of a research setup more adaptable to different environmental and temporal contexts. What is relevant and what has to be accounted for in the form of, for instance, mediating and moderating factors can be informed by theoretical understanding of the nature of the subject of investigation regarding underlying mechanisms, interdependencies, context dependencies and temporal (in)stability. Hence, theoretical uncertainty, as to what degree background knowledge is shared in a community not just influences the practice of ‘reproducibility’ but also its appraisal (see e.g., Norton, 2015; Steinle, 2016).

These methodological and theoretical understandings can and to a certain degree do remain tacit in practice, although to what degree such tacit understanding is shared between researchers is community specific (Collins, 1985; Collins et al., 2023). Additionally, we would like to emphasize that tacit knowledge is not the same as personal knowledge and the presence of tacit knowledge is not an indicator for a lack of formalization in a research community or knowledge production mode. Tacit knowledge is implicit knowledge about practices and conventions within a community that are acquired through the process of socialization (Collins, 1991, 2016). Hence, it can be personal or shared. Whether it is personal or shared tells you something about the cohesiveness of a community regarding the elements that are shared or not. In other words, to what degree tacit knowledge is personal or shared in a specific research community influences the uncertainty associated with a study. A certain amount of theoretical and methodological understanding can however explicitly be shared in form of detailed descriptions about what was done (protocols), as well as underlying theories and maybe even more importantly why it was done demonstrating theoretical and methodological understanding. This often falls under the umbrella of transparency (production & analytical transparency), but reflexivity and thick descriptions can be equally insightful for these purposes (see Field & Derksen, 2021; Jamieson et al., 2023). Since uncertainty is to a certain degree a community and tacit characteristic someone from within the specific knowledge community might have to be part of the assessment. See Figure 4 for an example illustration of where, for instance, in a theory testing research process theoretical and methodological uncertainty are located.

A diagram of a diagram

Description automatically generated with medium confidence

Figure 4 – Inspired by Trafimow (2023). Depiction of the location of theoretical and methodological uncertainty. Theoretical uncertainty envelopes the whole investigative process while methodological uncertainty spans from data collection to the inference/ conclusion.

**Application**

**Relevance based on the function of a specific kind of replication or reproducibility:**

Based on the answers to the questions for each of the research aspects the question has to be answered whether the stated functions or purposes of replication or reproducibility are compatible with or appropriate for them.

Epistemology

● What is the epistemological position underlying the study?

○ Is the aim to identify one underlying truth?

○ Is it about meanings, lived experiences, opinions or interpretations?

○ Is it about plurality and diversity, or convergence towards consensus/ unification?

○ Is it about protecting objectivity or valuing subjectivity?

Systems of Justification

● What are the established epistemic criteria for the type of research in a study that indicate trustworthiness or quality?

○ e.g., are (internal, external) validity, generalizability and reliability important?

○ e.g., are confirmability, credibility, transferability, and dependability important?

○ e.g., are plausibility, criticality, and authenticity important?

● What are the established practices used to ensure the epistemic criteria of the specific type of research?

○ Are there other practices than replication or reproducibility that fulfill similar enough or more appropriate purposes for the specific kind of research e.g., does triangulation or crystallization already fulfill a similar enough or more relevant function than kinds of replication or reproducibility?

Research Goals

● What is the motivation or goal behind the study?

○ What is the research question?

e.g., are we investigating question of style or meaning?

○ Is it about profit or getting some kind of head start in a market competition?

**Feasibility based on practice of actual redoing or enabling redoing**

Nature of the subject of investigation

● How complex is the subject, ….

● How time sensitive is the subject of investigation, is it changing over time or stable?

● How context dependent is the subject of investigation, is it interacting with the surroundings or even the investigator, or is it indifferent?

Research setup and its resource dependence

● To what degree do redoing or enabling depend on specific resources like technologies, subjects (e.g., participants or materials), instruments or machines?

○ How available and expensive are those resources?

Theoretical Uncertainty

● To what degree is the theoretical understanding of the subject of investigation shared between the researcher(s) of the original study and the researcher(s) considering to redo it? How well understood is the subject?

● To what degree is the theoretical understanding underlying an original study explicitly shared? In other words, to what degree is it clear how theoretical understanding guides the proceedings of the original study?

Methodological Uncertainty

● To what degree is the understanding of how to use and interpret the methods shared between the researchers of the original study and the researchers considering to redo it? How well is the method understood?

● To what degree is what was done in an original study described and explained. In other words, how detailed are the descriptions and explanations of methods, analysis and interpretations.

These questions are not a substitute for independent thinking, but a decision aid for an informed decision based on criteria that illuminate the relevance and feasibility of different kinds of reproducibility or replication for individual studies.

**Acknowledgements**

This research is supported by the Horizon Europe project TIER2 under grant agreement No 101094817***.***

**References**

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Barba, L. A. (2018). *Terminologies for Reproducible Research*. https://doi.org/10.48550/ARXIV.1802.03311

Baumgaertner, B., Devezer, B., Buzbas, E. O., & Nardin, L. G. (2018). *Openness and Reproducibility: Insights from a Model-Centric Approach*. https://doi.org/10.48550/ARXIV.1811.04525

Bennett, E. A. (2021). Open Science From a Qualitative, Feminist Perspective: Epistemological Dogmas and a Call for Critical Examination. *Psychology of Women Quarterly*, *45*(4), 448–456. https://doi.org/10.1177/03616843211036460

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

Buzbas, E. O., Devezer, B., & Baumgaertner, B. (2023). The logical structure of experiments lays the foundation for a theory of reproducibility. *Royal Society Open Science*, *10*(3), 221042. https://doi.org/10.1098/rsos.221042

Carter, S. M., & Little, M. (2007). Justifying Knowledge, Justifying Method, Taking Action: Epistemologies, Methodologies, and Methods in Qualitative Research. *Qualitative Health Research*, *17*(10), 1316–1328. https://doi.org/10.1177/1049732307306927

Chang, H. (2012). *Is Water H2O?: Evidence, Realism and Pluralism* (Vol. 293). Springer Netherlands. https://doi.org/10.1007/978-94-007-3932-1

*Classifying Science* (Vol. 7). (2004). Springer Netherlands. https://doi.org/10.1007/978-1-4020-3095-6

Collins, H. (2016). Reproducibility of Experiments: Experimenters’ Regress, Statistical Uncertainty Principle, and the Replication Imperative. In H. Atmanspacher & S. Maasen (Eds.), *Reproducibility: Principles, Problems, Practices, and Prospects* (pp. 65–82). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118865064.ch4

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

Collins, H., Shrager, J., Bartlett, A., Conley, S., Hale, R., & Evans, R. (2023). Hypernormal Science and its Significance. *Perspectives on Science*, *31*(2), 262–292. https://doi.org/10.1162/posc\_a\_00572

de Rijcke, S., & Penders, B. (2018). Resist calls for replicability in the humanities. *Nature*, *560*(7716), 29–29. https://doi.org/10.1038/d41586-018-05845-z

Derksen, M. (2019). Putting Popper to work. *Theory & Psychology*, *29*(4), 449–465. https://doi.org/10.1177/0959354319838343

Devezer, B., Nardin, L. G., Baumgaertner, B., & Buzbas, E. O. (2019). Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity. *PLOS ONE*, *14*(5), e0216125. https://doi.org/10.1371/journal.pone.0216125

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

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00621

Fanelli, D. (2018). Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, *115*(11), 2628–2631. https://doi.org/10.1073/pnas.1708272114

Fidler, F., & Wilcox, J. (2021). Reproducibility of Scientific Results. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/

Field, S. M., & Derksen, M. (2021). Experimenter as automaton; experimenter as human: Exploring the position of the researcher in scientific research. *European Journal for Philosophy of Science*, *11*(1), 11. https://doi.org/10.1007/s13194-020-00324-7

Field, S. M., Van Ravenzwaaij, D., Pittelkow, M.-M., Hoek, J. M., & Derksen, M. (2021). *Qualitative Open Science – Pain Points and Perspectives* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/e3cq4

Fricker, M. (2007). *Epistemic Injustice*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Gibbons, M. (Ed.). (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. SAGE Publications.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (2010). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. SAGE Publications Ltd. https://doi.org/10.4135/9781446221853

Gläser, J., Laudel, G., Grieser, C., & Meyer, U. (n.d.). *Scientific fields as epistemic regimes: New opportunities for comparative science studies*.

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

Guba, E.G. and Lincoln, Y.S. (1994) Competing paradigms in qualitative research. In:   
 Denzin, N.K. and Lincoln, Y.S., Eds., Handbook of Qualitative Research,   
 Sage Publications, Inc., Thousand Oaks,105-117

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

Guzzo, R. A., Schneider, B., & Nalbantian, H. R. (2022). Open science, closed doors: The perils and potential of open science for research in practice. *Industrial and Organizational Psychology*, *15*(4), 495–515. https://doi.org/10.1017/iop.2022.61

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/10.1038/s41567-018-0342-2

Holbrook, J. B., Penders, B., & de Rijcke, S. (2019). *The humanities do not need a replication drive*. CWTS. Retrieved September 22, 2023, from https://www.cwts.nl:443/blog?article=n-r2v2a4&title=the-humanities-do-not-need-a-replication-drive

Huijnen, Pim, & Huistra, Pieter. (2022). *On the Use of Replications in History*. https://doi.org/10.5281/ZENODO.7037400

Kaplan, A. (1964). The Conduct of Inquiry. San Francisco, CA: Chandler.

Kitzes, J., Turek, D., & Deniz, F. (Eds.). (2018). *The practice of reproducible research: Case studies and lessons from the data-intensive sciences*. University of California Press.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* University of Chicago Press:   
 Chicago.

Kuhn T. S. (1970). *The structure of scientific revolutions* (Second edition enlarged).   
 University of Chicago Press.

Legg, C., & Hookway, C. (2021). Pragmatism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/pragmatism/

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

Leonelli, S. (2022). Open Science and Epistemic Diversity: Friends or Foes? *Philosophy of Science*, *89*(5), 991–1001. https://doi.org/10.1017/psa.2022.45

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications Beverly Hills, Calif.; WorldCat.

Locke, K., & Golden-Biddle, K. (1997). CONSTRUCTING OPPORTUNITIES FOR CONTRIBUTION: STRUCTURING INTERTEXTUAL COHERENCE AND “PROBLEMATIZING” IN ORGANIZATIONAL STUDIES. *Academy of Management Journal*, *40*(5), 1023–1062. https://doi.org/10.2307/256926

Malich, L., & Rehmann-Sutter, C. (2022). Metascience Is Not Enough – A Plea for Psychological Humanities in the Wake of the Replication Crisis. *Review of General Psychology*, *26*(2), 261–273. https://doi.org/10.1177/10892680221083876

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

Moon, K., & Blackman, D. (2014). A Guide to Understanding Social Science Research for Natural Scientists: Social Science for Natural Scientists. *Conservation Biology*, *28*(5), 1167–1177. https://doi.org/10.1111/cobi.12326

Nadin, M. (2018). Rethinking the experiment: Necessary (R)evolution. *AI & SOCIETY*, *33*(4), 467–485. https://doi.org/10.1007/s00146-017-0705-8

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. (2019). *Strategy for Culture Change*. Retrieved August 22, 2023, from https://www.cos.io/blog/strategy-for-culture-change

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Peels, R. (2019). Replicability and replication in the humanities. *Research Integrity and Peer Review*, *4*(1), 2. https://doi.org/10.1186/s41073-018-0060-4

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

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

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

Plesser, 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

Pownall, M. (2022). *Is replication possible for qualitative research?* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/dwxeg

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

Robins, R. W., Gosling, S. D., & Craik, K. H. (1999). An empirical analysis of trends in psychology. *American Psychologist*, *54*(2), 117–128. https://doi.org/10.1037/0003-066X.54.2.117

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, Re-Implementation, Re-Analysis, and Re-Use in Computational Literary Studies* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-2657846/v1

Shan, Y. (2022). Philosophical foundations of mixed methods research. *Philosophy Compass*, *17*(1), e12804. https://doi.org/10.1111/phc3.12804

Sikorski, M., & Andreoletti, M. (2023). Epistemic Functions of Replicability in Experimental Sciences: Defending the Orthodox View. *Foundations of Science*. https://doi.org/10.1007/s10699-023-09901-4

Stahl, N. A., & King, J. R. (2020). Expanding Approaches for Research: Understanding and Using Trustworthiness in Qualitative Research. *Journal of Developmental Education*, *44*(1), 26–28. JSTOR.

Steinle, F. (2016). Stability and Replication of Experimental Results: A Historical Perspective. In H. Atmanspacher & S. Maasen (Eds.), *Reproducibility: Principles, Problems, Practices, and Prospects* (pp. 39–63). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118865064.ch3

Steltenpohl, C. N., Lustick, H., Meyer, M. S., Lee, L. E., Stegenga, S. M., Standiford Reyes, L., & Renbarger, R. L. (2023). Rethinking Transparency and Rigor from a Qualitative Open Science Perspective. *Journal of Trial and Error*. https://doi.org/10.36850/mr7

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

Trafimow, D. (2023). A New Way to Think About Internal and External Validity.   
 *Perspectives on Psychological Science*, *18*(5), 1028–1046.   
 https://doi.org/10.1177/17456916221136117

Tuval-Mashiach, R. (2021). Is replication relevant for qualitative research? *Qualitative Psychology*, *8*(3), 365–377. https://doi.org/10.1037/qup0000217

Voelkl, B., & Würbel, H. (2016). Reproducibility Crisis: Are We Ignoring Reaction Norms? *Trends in Pharmacological Sciences*, *37*(7), 509–510. https://doi.org/10.1016/j.tips.2016.05.003

Voelkl, B., & Würbel, H. (2021). A reaction norm perspective on reproducibility. *Theory in Biosciences*, *140*(2), 169–176. https://doi.org/10.1007/s12064-021-00340-y

Whitley, R. (2006). *The intellectual and social organization of the sciences* (2. ed., reprinted). Oxford Univ. Press.

Würbel, H. (2000). Behaviour and the standardization fallacy. *Nature Genetics*, *26*(3), 263–263. https://doi.org/10.1038/81541

Yadav, D. (2022). Criteria for Good Qualitative Research: A Comprehensive Review. *The Asia-Pacific Education Researcher*, *31*(6), 679–689. https://doi.org/10.1007/s40299-021-00619-0

Yanow, D., & Schwartz-Shea, P. (2010). Interpretive Research: Characteristics and Criteria: *Revue Internationale de Psychosociologie*, *Vol. XV*(35), 29–38. https://doi.org/10.3917/rips.035.0029

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed). Sage Publications.

1. We use ’reproducibility’ for all the “re-terms” like replication, replicability, reproducibility, reproduction, repeatability, repetition, reanalysis, reinterpretation, conceptual replication, direct replication, (in)exact replication, constructive replication, literal replication, scientific replication, statistical replication, method reproducibility, result reproducibility, inferential reproducibility, computational reproducibility, hypothetical replicability and many more. [↑](#footnote-ref-1)
2. Inspired by Peng’s (2006) reproducibility spectrum. [↑](#footnote-ref-2)
3. For an attempt at translating this culture change pyramid into practice see the TOP guidelines (Nosek et al., 2015) [↑](#footnote-ref-3)
4. See also <https://tier2-project.eu/news/reproducibility-all> and <https://reproducibility.umn.edu/> [↑](#footnote-ref-4)
5. We do not mean Gibbons and colleagues’ (Gibbons, 1994, Gibbons et al., 2010) knowledge production modes with mode 1, mode 2 and mode 3. However, if the name knowledge production mode is problematic, we can also refer to our framework as the knowledge production matrix framework. [↑](#footnote-ref-5)
6. It is important to emphasize that the focus regarding relevance of ‘reproducibility’ is here mostly on what we might call epistemic relevance, which is about whether the epistemic practices and criteria fit into the knowledge production mode mostly with regards to the epistemological stance and the system(s) of justification. Social relevance would then be the degree to which a practice or criteria is relevant due to the mutual dependence (interdependencies) of researchers, e.g., through the peer review system, within and across certain research communities (Gläser et al., 2018; Whitley, 1984/2000). Whether social and epistemic relevance are really two distinct aspects, or whether this is just a question of perspective and prioritization is an issue for another time and place. [↑](#footnote-ref-6)
7. We do recognize the importance of ontology for the relevance of ‘reproducibility’. However, we somewhat merged epistemology and ontology into one category under epistemology to make the framework more manageable. [↑](#footnote-ref-7)