Goal-directed Uses of the Replicability Concept

Eden T. Smith[^,1,2], Hannah Fraser[1,2], Steve Kambouris[1,2], Fallon Mody[2], Martin Bush[2], Fiona Fidler[3]

Authorship based on CRediT contributor roles:

1. These authors contributed to the writing
2. These authors contributed to editing and revision
3. These authors contributed through project supervision and resources

[^] Corresponding author: eden.smith@unimelb.edu.au

All authors affiliated with MetaMelb Research Group, University of Melbourne, Melbourne, Victoria, Australia.

Abstract

The replicability of a research claim is often positioned as an important step in establishing the credibility of scientific research. This expectation persists despite ongoing disagreements over how to characterise replication practices in various contexts. Rather than attempt to explain or resolve these disagreements, we propose that there is value in exploring the variable uses of the replicability concept. To this end, we treat the replicability concept as a goal-directed tool for studying scientific practices. This approach extends scholarship on the goal-directed uses of concepts within investigative practices to explore the value of reflexively interrogating those concepts used to study the sciences as practiced. In doing so, we highlight the importance of considering how and when a given concept is an appropriate tool for a given goal when studying scientific practices. For instance, in the case of the replicability concept, our examination suggests that there is value in clarifying how a given characterisation of replicability is appropriate to a given context-specific analytic and/or evaluative goal. In addition, we hope to draw attention to how Metaresearch and HPS contexts provide complementary insights for those seeking to understand when and how replication practices can help in assessing a given research claim.

Contents

Abstract **1**

Introduction **1**

Characterising replication practices **4**

Replicability as a concept for studying scientific practices **6**

Contextualising different uses of the replicability concept **10**

Metaresearch on replication practices 11

Historical, philosophical, and social studies of replication practices 13

Using the replicability concept as a goal-directed tool **16**

Acknowledgements **20**

References **20**

Introduction

Insights about the uses of scientific concepts as goal-directed tools within investigative practice offer an avenue for reflecting on how concepts are used to study the process of generating and disseminating research claims within the broader dynamics of scientific practices.[1] To illustrate this potential, we explore some of the variable uses of the replicability concept within studies that evaluate and/or analyse the dynamics of scientific practices.

The replicability of a research claim is a contested concept. Used broadly, replicability refers to any practice of assessing a research claim by repeating a prior study and comparing any findings about the phenomena of interest with those originally reported.[2] Given the difficulties of repeating any given study, this concept is often refined by specifying additional characteristics. Which characteristics are associated with a given use of the replicability concept often reflect the criteria expected to identify those practices that count as replications and the measures proposed for determining replication success/failure. For example, additional details sometimes characterise replication practices in terms of the ideal degree of similarity between the original and repeated study. Alternatively, replication practices are characterised by a set of criteria for determining when a repeated finding successfully reproduced those originally reported. Yet another way to characterise replication practices is by articulating guidelines for assessing the validity of the reported relationships between a repeatable finding and the scope of the research claim. To further complicate the situation, a range of ambiguous terminology has emerged in attempts to characterise the various practices associated with establishing the replicability of a research claim (Barba, 2018; Plesser, 2018). There are a range of productive approaches to examining variable characterisations of replication practices, and their associated terminological ambiguities. These include studying the contexts within which various terminology for replications came to be used (e.g., Peng & Hicks, 2021; Plesser, 2018), exploring variable understandings of what counts as replication practices (e.g., Alister et al., 2021; Fraser et al., 2020), outlining context-specific approaches to replication practices (e.g., Machery, 2020; Nosek et al., 2021), and examining replication practices as part of other historical, philosophical, and social studies considerations (e.g., Guttinger, 2020; Peterson & Panofsky, 2020; Romero, 2019).

In this paper we focus on understanding variable uses of the replicability concept for studying scientific practices, rather than attempt to resolve disagreements over how to characterise these practices, or define relevant terms within specific epistemic

[1] Unless otherwise specified, 'research claim', or 'claim', is used synonymously with 'first-order knowledge claim' – that is, a claim generated by scientific research practices and presented as a unit contribution to a field of scientific knowledge such that, if it were to be incorporated into the structure of that field, could contribute to accepted scientific facts (Leydesdorff, 1991, p. 75). This includes, yet is broader than, narrower definitions of a research claim as a statistically-based assertion about the strength and direction of an effect. Relatedly, throughout this paper 'investigative practices' indicates epistemic activities that generate first-order units of knowledge (research claims), including experimental interventions, as well as field observations, modelling, and more. These are an element of the broader range of 'scientific practices' through which first-order knowledge claims contribute to the generation and dissemination of established scientific knowledge (including peer-review and publication practices, funding constraints, and so forth).

[2] Here we use the term phenomena for 'stable regularities in the world', as distinct from an instance of a given phenomenon that is reported as the effect found in a specific investigation/experiment.

communities.[3] This focus on the uses of the replicability concept builds on scholarship detailing how specific concepts can contribute to the localised practices of scientific investigation.[4] This scholarship supports an understanding of scientific concepts as structured within communities of practice to function as goal-directed tools.[5] Recognising that scientific concepts can function as tools within investigative practices suggests that the concepts used to analyse and evaluate these practices also need interrogation. Taking this reflexive view, we explore how literature on the uses of concepts as goal-directed tools within investigative practices can be extended to interrogate the uses of concepts for evaluating and analysing the dynamics of scientific practices.

With this extension we treat the replicability concept as a tool for studying scientific practices, rather than detail how the replicability concept functions as a goal-directed tool more generally. By taking this approach, we also hope to draw attention to how different goals-directed uses of the replicability concept can complement each other. To this end, we make an analytic distinction between two of the broad approaches to using the replicability concept. However, this distinction should not be taken as describing disciplinary boundaries. While specialised disciplinary approaches to studying scientific practices are important, Joseph Rouse (2011) notes that the enclaves that form around specialist approaches to studying scientific practices can often obscure their overarching insights. Therefore, we do not intend to comment on any of the disputed boundaries between the multiple disciplines within which the replicability concept is used. Instead, we hope to highlight some overarching insights that are emerging from the uses of the replicability concept within two valuable approaches to studying scientific practices.

The first of these approaches is exemplified by approaches that emerge from the perspective of scientists studying the sciences as practiced. Within this 'Metaresearch context', studying replication practices often involves gathering empirical evidence that facilitates the evaluation of specific scientific practices and the development of improvement initiatives across multiple scientific disciplines.[6] The second of these approaches is often found within studies of replication practices that draw on perspectives informed by the historical, philosophical, and/or social studies of the sciences as practiced (HPS). Within this 'HPS context', the perspective cultivated focuses on developing accounts that analyse the dynamic conditions within which various elements of scientific knowledge are generated and disseminated over time.[7]

[3] Throughout this paper 'replicability' refers to the notion that a given research claim can be assessed in terms of its likely replicability, with the term 'replications' used for the wide variety of specific practices associated with this concept (including when those practices are described in other terms). This is consistent with Guttinger's (2020) distinction between the practice of attempting to reproduce a given prior finding (replications) and the quality of a given finding which can be assessed in terms of the feasibility of the relevant study being replicated (replicability).

[4] For an overview of these strands, see (Smith, 2019).

[5] This approach is outlined below. For more detail, see (Smith, 2018a, 2018b).

[6] For accounts of Metaresearch emphasising the importance of evaluating and improving scientific practices, see (Ioannidis, 2018; Ioannidis et al., 2015). For an account of the 'social movement of Metascience' as emerging at the intersection of earlier metaresearch practices within open science activism, the science of science, and statistical and methodological critiques, see (Peterson & Panofsky, 2020).

[7] For accounts of HPS studies emphasising the value of analysing the historical, philosophical, and social implications of the sciences as practiced, see (H. Chang, 2014; Soler et al., 2014).

It is worth reiterating that this is an analytic distinction. Neither the Metaresearch or HPS contexts take a uniform approach to studying the sciences as practiced, and there are notable overlaps between their respective contributions to evaluating and analysing scientific practices. As detailed below, distinguishing between Metaresearch contexts and HPS contexts is a device to highlight the complementary insights converging from differing goal-directed uses of the replicability concept when studying scientific practices. Making this analytic distinction allows us to focus on exploring examples of the replicability concept being used to pursue two goals: i) evaluating contemporary investigative practices within Metaresearch contexts that seek to inform initiatives for improving the generation and dissemination of scientific knowledge, and ii) analysing a range of historical and contemporary investigative practices to contribute to understandings within HPS contexts of the long-term dynamics of generating and disseminating scientific knowledge.[8]

While focusing on the differences between these two goals-directed approaches to using the replicability concept, it is important to note that they overlap in many ways. For instance, both goals reflect shared assumptions about the relative value of the multiple interdependent epistemic functions attributed to replication practices within both Metaresearch and HPS contexts. These interdependent functions include testing the reliability of the findings reported in support of a given research claim, as well as helping to establish the robustness and/or generalisability of a research claim that a set of findings is reported to support.[9] As detailed later, these shared assumptions can be seen in the asymmetrical value attributed to these overlapping epistemic functions in different contexts.

In examining the uses of the replicability concept as a tool for studying scientific practices within Metaresearch and HPS contexts, our aim is to highlight the value of recognising that concepts are used for pursuing context-specific goals. In doing so, we explore how the variable goal-directed uses of the replicability concept can contribute to converging accounts of scientific practices. As demonstrated below, this includes a convergence between Metaresearch and HPS scholarship highlighting that expectations about replication practices need to be contextualised in relation to the specific scientific practices being studied.

Comparing the uses of the replicability concept within Metaresearch and HPS contexts also demonstrates the potential of reflecting on the goals towards which evaluative and analytic concepts are being used. In this case, treating replicability as a concept used as a goal-directed tool demonstrates that, just as with concepts used within investigative practice, there is value in justifying when and how to use concepts as tools to study scientific practices. To illustrate this potential, we treat the replicability concept as a tool for studying scientific practices so that we can focus on comparing how the replicability concept can be used to pursue the context-specific Metaresearch and HPS goals outlined above. As with

[8] While outside the present scope, one important point of difference is that Metaresearch studies often draw on old prescriptivist philosophical accounts of (an ideal) Science that are no longer considered relevant to contemporary HPS accounts of the sciences as practiced. For example, see Maarten Derksen's (2019) description of the influence of Poppererian philosophy of science on the reform movement in social psychology.

[9] Note that concepts such as reliability and robustness have equally varied uses across different contexts. For this discussion we are using these to distinguish between practices that establish whether a reported result can be reproduced by repeated trials of the same conditions (reliability) and those that establish whether equivalent results are found under a wide range of conditions (robustness).

concepts used within investigative practices, specific uses of the replicability concept rely on sets of characteristics that individuate those specific forms of replication practices relevant to context-specific goals. Given this, it is difficult to compare uses of the replicability concept without considering the range of associated characteristics. Therefore, we begin by contextualising disagreements over how to characterise practices associated with the replication concept.

Having outlined the contested characterisations of the replicability concept, we then review the strand of scholarship demonstrating the value of investigating the uses of concepts as goal-directed tools within investigative practices. Building on this scholarship, we extend the view of concepts-as-tools to argue for the value of reflexively examining the concepts we use *to study* how investigative practices contribute to scientific knowledge more generally. Taking this approach, we explore uses of the replicability concept within the contexts of Metaresearch and HPS studies of scientific practices – focusing on a comparison between studies with an evaluative or analytic goal.

In the final section we aim to demonstrate the value of examining the replicability concept as used for context-specific goals. In doing so, we illustrate the importance of justifying how a given characterisation of replicability appropriately individuates the replication practices relevant to a given context-specific analytic and/or evaluative goal. In addition, we draw attention to how variable uses of the replicability concept within Metaresearch and HPS contexts each provide complementary insights about when and how replication practices offer a relevant strategy for assessing a given research claim.

Characterising replication practices

There have been multiple attempts to characterise the set of practices required to assess the replicability of an individual study. For instance, one review of the social science literature identified 18 replication typologies and 79 replication types (Gómez et al., 2010). These different typologies highlight the range of overlapping characteristics proposed for various types of practices associated with repeating a study in ways that assess the replicability of a research claim. Additionally, accounts of replication-like practice often use alternative terms – including repetitions and reproductions – each of which have also been inconsistently characterised. The difficulties of establishing which characteristics individuate the set of practices required to assess the replicability of an individual study suggests that disagreements over defining the *term* replication are only part of the story. Therefore, before comparing context-specific uses of the replicability concept, it is important to appreciate the range of characteristics that have become associated with this concept within both Metaresearch and HPS contexts. As detailed in the next section, these characteristics help to individuate an instance of the concept being used to study a specific type of investigative practice. This process of individuation involves identifying how the practice of interest is distinct from other related practices.

One influential way to distinguish between key types of replication practices is based on the degree of similarity between the original and repeated study (Fidler & Wilcox, 2018; Schmidt, 2017). A prominent example is the continuum between 'direct' and 'conceptual' replications (Schmidt, 2009). At one end of the scale, the term 'direct replicability' refers to the practice of repeating an initial study by following the same procedures as closely as possible. At the other end of the scale, 'conceptual replicability' refers to the practice of

repeating the initial study while deliberately altering one or more key components of the original procedure. This direct/conceptual continuum has been both influential and controversial (Guttinger, 2020; Nelson et al., 2020).

There are also attempts to differentiate between types of replication practices focusing on the expected function of each practice. For example, Nancy Cartwright (1991) argued for three levels of practice: (i) routine checking of the implementation of a given study; (ii) replication of the results, and (iii) reproduction of the phenomenon. Hans Radder (1992) proposed a different three types: (i) exact repeatability of the experiment from the same point of view of the theoretical interpretation of the question; (ii) replicability of the experimental result; and (iii) reproducibility of the material realization of the experiment regardless of theoretical expectations. More recently, discussing machine learning in the biosciences, McDermott et al., (2019) argue that there are three criteria for demonstrating the reproducibility/replicability of a study: reproduction of results under identical technical conditions; reproduction of results under statistically identical conditions; and replicating results under conceptually identical conditions. Meanwhile, drawing on the approach taken by the National Academies of Sciences, Brian Nosek and colleagues (2021, p. 4) offer another set of distinctions: "Replication refers to testing the reliability of a prior finding with different data. Robustness refers to testing the reliability of a prior finding using the same data and different analysis strategy. Reproducibility refers to testing the reliability of a prior finding using the same data and same analysis strategy".

Types of replication practices have also been distinguished from each other based on characteristic differences in the scope within which the research claim is expected to be stable. For example, while the repetition of experimental findings by original researchers feature as one of the practices associated with establishing the replicability of a given research claim historically, contemporary replications are most commonly characterised as a practice conducted by researchers not involved in the original study.[10] Additionally, replication practices have been characterised based on distinctions between their role in establishing either the internal or external validity of a research claim (Fabrigar et al., 2020; R. McDermott, 2011).[11] In this context, practices with internal validity are characterised as demonstrating internal consistency between a reported effect and the associated inferences underpinning a claim. In contrast, practices with external validity are characterised as helping to establish that a claim can generalise beyond the initial conditions (either by demonstrating that the effect persists beyond initial study conditions or that the claim is supported by converging findings).

The challenges of characterising such a wide range of practices as replications are well recognised. For example, in explaining the value of studying the role of re-doing experiments within the history of the sciences, Jutta Schickore (2011, p. 345) notes that "the very concept of replication has not received very much analytic attention.... As yet, no consensus about these concepts, their meaning and significance has emerged". Likewise, Samuel C. Fletcher (2021) draws attention to the lack of justifications provided for the

[10] For examples of studies focusing on researchers repeating their own studies, see (Bogen, 2001; Guttinger, 2019; Schickore, 2011); for examples and discussion of the expectation of independence of a repeated study, see (Clemens, 2017; Gómez et al., 2010; Guttinger, 2019; Makel et al., 2012; Patil et al., 2016; Radder, 1992).

[11] For more on the distinction between internal and external validity, see (Guala, 2003).

various ways in which the success/failure of replicability is measured. In addition, surveys of scientists demonstrate that there are multiple practices associated with replication concepts more generally. For example, when ecologists were presented with a choice of six variations on how a study might be repeated and asked to select which they considered to be replications, less than 40% chose only one option (Fraser et al., 2020). Similarly, a dual-language survey of psychologists found a clear preference for two of the options on the list of replication variations – with Italian psychologists seeming to rely on a more restricted meaning of replication than Australian psychologists (Agnoli et al., 2020).

In response to the challenges raised by the range of practices characterised as replications, there have been attempts to specify which one set of such practices has the most value. This includes proposals to narrow the definition of 'replication' – definitions which necessarily exclude related practices previously conceptualised as forms of replication practices. For example, Edouard Machery (2020) proposed recategorizing conceptual replications as extensions based on the argument that replications should refer only to their 'proper function' of establishing the reliability of a research claim through the practice of repeating the key components of the investigation as closely as possible. Meanwhile, describing direct replications as an idealised expectation in the ability to test the reliability of an effect, Brian Nosek and Timothy Errington (2020) propose a differently narrowed definition that emphasises the role of multiple replications in establishing the generalisability of a claim. This re-definition specifies that a replication can be any kind of study that would provide "diagnostic evidence about a claim from prior research" (Nosek & Errington, 2020, p. 2). Definitions aside, other aspects of these accounts of replicability are compatible. However, focusing on attempts to define the term 'replications' can obscure this complementarity (and potentially add yet more candidates to the array of typologies available for differentiating replications from related practices).

In addition to the range of terms and inconsistent characteristics, there are also difficulties in establishing the relative value of different types of replication practices. Exploring this possibility, an analysis of the discourse around replication practices suggests that disagreements about replications reflect foundational differences in perspective – from entrained disciplinary differences, to differences in baseline assumptions about the intrinsic heterogeneity, or lack thereof, within the natural world (Nelson et al., 2020). One such difference in perspective is the context-specific goal which the replicability concept is used for when studying scientific practices. With this in mind, in the next section we examine how the replicability concept is used for studying scientific practices from perspectives exemplified within Metaresearch and HPS contexts respectively.

Replicability as a concept for studying scientific practices

As outlined above, replicability is a disputed concept. In addition to the ambiguous terminology, there is no single set of characteristics that individuate all those practices associated with assessing the replicability of a research claim. Fortunately, there is a strand of scholarship detailing a range of productive ways in which disputed scientific concepts can

function productively even without consensus on the relevant terms or defining characteristics.[12]

As mentioned earlier, this strand of scholarship offers a view of scientific concepts as structured within communities of practice to function as goal-directed tools.[13] While the insights associated with this view emerged through studies of the uses of scientific concepts within investigative practices, they have the potential to prompt similar reflections on how concepts can also function as tools for studying scientific practices more generally. To explore this potential, we draw on the view of concepts as goal-directed tools to reflect on how the replicability concept is used to evaluate and/or analyse how specific types of investigative practices contribute to scientific knowledge. In taking this approach, we are not seeking to establish that the replicability concept functions as a goal-directed tool in general (that would require a far more extensive study). Instead, we focus on the more foundational possibility there is value in extending existing insights about how scientific concepts can be used as goal-directed tools within investigative practices, to help us reflect on how the replicability concept is used to evaluate and/or analyse specific types of scientific practices within Metaresearch and HPS contexts respectively.

While a full exposition of the literature behind this view of concepts would go beyond the present scope, it is important to clarify some of the key analytic terms we have adopted. Firstly, we understand *scientific concepts* to be accrued bodies of shared knowledge used to individuate instances of a type of phenomena investigated in pursuit of specific goals within a given knowledge domain (Smith, 2019). The approach treats scientific concepts as functioning at the scale of small communities of practice (i.e. as more than mere elements of individual cognition, yet not so pervasive as to be elements of major social systems of thought). Likewise, our focus on the *uses* of concepts draws on studies of scientific concepts that contributed to empirical knowledge in ways that extend beyond their more traditionally recognised roles in mental and linguistic representation (MacLeod, 2012; Nersessian, 2012). This appreciation of concept-use also aligns with research demonstrating that concepts need not refer to a universally accepted referent to function productively within scientific practices (Bloch, 2012b; Bloch-Mullins, 2020; Brigandt, 2003; Pöyhönen, 2013; Waters, 2014). Relatedly, there is a growing appreciation for the view that concepts can be neither reduced to their referents nor dismissed as mere social constructs (Kindi, 2012).[14] Instead, concepts are increasingly understood to emerge and evolve within locally contingent and historically dynamic patchworks such that they can resist their intended use (Kindi, 2012).

We also draw on the notion that concepts can be used as *tools* that enable scientific practice (Arabatzis, 2019).[15] As Uljana Feest (2010, p. 173) argues, scientific concepts can function as tools that individuate a given type of phenomena (for further investigation

[12] For an overview of the multiple strands of literature contributing to this area of scholarship see (Smith, 2018a, 2018b).

[13] For more detail, see (Smith, 2020).

[14] For more on the value of pluralistic realist views of scientific concepts that do not assume that natural kinds can be read off the world, see Bloch-Mullins (2020, p. 24).

[15] For examples, see (Bloch, 2012a; Boon, 2012; Feest, 2012; MacLeod, 2012; Steinle, 2012).

within an available body of knowledge) by singling out instances of that phenomena from related phenomena. This process of singling out an instance of a given type of phenomena requires knowing which characteristics, or 'phenomenal properties', are unique to the phenomena of interest (i.e., not shared by other types of potentially related phenomena).[16] In addition to further development of the notion of concepts as tools, our emphasis on how concepts are characterised draws most directly on the work of Corinne Bloch-Mullins (2012a; 2020). For example, Bloch-Mullins (2020) argues that the process of individuating a concept involves articulating the distinguishing characteristics of the phenomena of interest – characteristics which may describe referential targets and reflect the inferential associations and contextual goals of the concept used. This appreciation for analysing multiple components of a concept also draws on Ingo Brigandt's (2002, p. 4) description of three of the distinct roles that concepts can play in scientific practices: the referential component (the kinds of entities, properties, or processes a concept refers to); the inferential component (connecting beliefs about how a concept can be used to support the inferences and explanations between the concept and other concepts within a given language community); and the epistemic goal (a component which provides the role of setting standards for what a concept can be used for within a given investigative practice). Examining each of these components helps to highlight how inferences are revealed or obscured by the entrenched associations embedded in the conceptual relationships that support the use of that concept for investigating and explaining the referential target (the phenomena of interest) for the purposes of a specific epistemic goal. Within this dynamic the characteristics of the referenced type of phenomena, and the set of inferences about an instance of that phenomenon, can all be changed if deemed epistemically warranted by the standards set by the epistemic goal the concept is used to pursue (Brigandt, 2012). Given this, when talking about goals, we seek to build on research exploring the importance of contextualising any given scientific concept in relation to the goals it is used to pursue (Brigandt, 2012, p. 78; MacLeod, 2012, p. 68; Steinle, 2012, p. 107). Understood within their context, concepts are tools that can only be judged as appropriate, or not, in relation to their use for a given goal (Steinle, 2010a, 2010b).

Drawn together, these fields of research provide a strand of scholarship demonstrating that scientific concepts can function as tools that are structured within communities of practice for pursuing context-specific goals (Smith, 2018a, 2020). For example, in a comparison of the goal-directed uses of two historically interdependent concepts as independent tools, Eden T. Smith (2018a) details how unexamined inferential assumptions about the phenomena of interest became entrenched in the routinised uses of these concepts in neuroimaging experiments. Used in routine pursuit of specific epistemic goals, these concepts were found to function analogously to mundane material instruments: treated as reliably appropriate for the purpose at hand – rarely justified and largely taken for granted.

Building on this strand of literature, insights from examining the context-specific uses of concepts as tools within investigative practice can be extended to reflect on how concepts function as evaluative and/or analytic tools for studying how investigative practices contribute to the generation and dissemination of scientific knowledge more generally. This

[16] To use Farkas' (2013, p. 399) example, "if two experiences both involve feeling cold they share a phenomenal property; if two experiences both involve something appearing blue, they share another phenomenal property".

extension draws on each of the key insights outlined above. To detail this extension, we will consider each point as it relates to examining uses of the replicability concept in studies that seek to evaluate and/or analyse the role investigative practices that involve repeating a previously reported study in generating and disseminating scientific knowledge.

The first point from the above snapshot of this literature is that concepts are frequently used within investigative practices even when their referential components remain disputed. Further, such disputed concepts are shown to function as productive tools in the generation of experimental knowledge. This suggests that there is value in considering the productivity of the contested concept of replicability for studying specific, yet ambiguously delineated, sets of scientific practices.

The second point is that concepts can be used *as tools* to individuate instances of the type of phenomena being investigated within a given set of investigative practices. As detailed above, when used within investigative practices concepts can function as tools to *individuate* a given type of phenomena of interest for further investigation by singling out instances of that type of phenomena from related phenomena based on a unique set of characteristics from within the available body of knowledge. Now, to extend this point from concepts used within investigative practices to those concepts used to study scientific practices, we will treat the 'practices being studied' within Metaresearch and HPS as analogous to the 'phenomena of interest' within a specific investigative practice. With this view, concepts such as replicability can be understood as a tool for singling out an instance of scientific practice for further evaluation and/or analysis. For example, when the replicability concept is used to study scientific practices it relies (often implicitly) on there being a set of differentiating characteristics delineating the type of replication practice of interest from potentially related types of practices.

In the context of studying scientific practices, the replicability concept may similarly function as a tool when used to individuate a given type of replication practice for further study within the available body of knowledge provided by Metaresearch and/or HPS contexts. As with individuating an instance of a given scientific phenomena for further investigation, the process of individuating a specific set of scientific practices for further study involves articulating the distinguishing characteristics that help to single out instances of that practice within a broader typology of related practices. Reflecting the complexity of such a process, the characteristics individuating a given conceptualisation of the practice of interest may describe specific referential targets while also reflecting the inferential associations entrenched in the contextual goals of the concept as used within the relevant context.

This brings us to the point that the characteristics used to individuate a given conceptualisation of a phenomena/practice of interest simultaneously describe specific referential targets and reflect the inferential associations entrenched in the contextual goals of the concept as used within the relevant context. Thinking about this is in terms of how the replicability concept is used highlights the interdependence of each component: an implied *referential target* (such as a given set of practices involved in repeating a prior study); as well as range of *inferential associations* (such as expectations about how this specific way of repeating a study establishes the reliability or robustness of a research

claim); and the *epistemic goal* (that sets the standards within which the practices of interest are studied – such as in terms of evaluating those practices to inform improvement initiatives, or analysing the conditions within which these practices contribute to broader fields of scientific knowledge over longer time frames). Viewed in this way, the variable uses of concepts such as replicability can be assessed in terms of their appropriateness for the relevant goal, rather than expecting replicability to function as a universally applicable concept.

The final point highlights that the uses of concepts emerge in relation to the historically contingent goals of communities of practice. While the historical context of Metaresearch and HPS perspectives are beyond the present scope, it is important to appreciate the possibility that uses of replicability concept came to be structured by entrenched inferential associations that accrued within the bodies of knowledge generated by these different communities of practice over time.

While brief, each of these points suggest that insights gained from examining the variable uses of scientific concepts as goal-directed tools in investigative practices offer avenues for reflecting on how disputed concepts are used to study scientific practices. In illustrating this extension, we also outlined the first step in exploring the potential of interrogating the replicability concept *as if* it is used as a goal-directed tool for studying scientific practices.[17] In the next section, we treat the replicability concept as a goal-directed tool as a way of exploring the complementarity of evaluative and analytic approaches to studying scientific practices. In doing so, we aim to demonstrate that there is value in identifying the goal-directed uses of the replicability concept as a tool for studying scientific practices.

Contextualising different uses of the replicability concept

There are many approaches to studying how investigative practices contribute to scientific knowledge. We compare two of these approaches to highlight the value of reflecting on the goal-directed uses of evaluative and analytic concepts within these contexts. To this end, we draw on examples of each of these approaches as exemplified by studies published within Metaresearch contexts and HPS contexts respectively. As detailed earlier, this is merely an analytic distinction: these two approaches often overlap, and both contribute to multiple contexts.

The approach exemplified by Metaresearch studies involves gathering empirical evidence about specific investigative practices with the aim of informing evaluations and improvement initiatives within one or more of the sciences. In line with this approach, Metaresearch studies that use the replicability concept often prioritise the goal of evaluating how contemporary scientific practices can be improved. Meanwhile, the approach exemplified by HPS studies using the replicability concept tend to focus on the goal of analysing the diachronic dynamics of the diverse practices involved in generating and disseminating scientific knowledge.

[17] This example is not intended to establish that the replicability concept functions as a goal-directed tool – that would require a historical analysis of the dynamic conceptual developments through which it came to be used as a tool within various contemporary practices, the extent of such a study is outside the present scope.

While exploring the differences between these approaches, we also draw on a view of concepts as goal-directed tools (detailed earlier) to explore how these different uses of the replicability concept in Metaresearch and HPS contexts converge. To this end, in this section we provide some examples of the goal-directed uses of the replicability concept within Metaresearch and HPS contexts. Then, in the final section, we build on these examples to outline how treating the replicability concept as a goal-directed tool highlights an underappreciated complementarity between the perspectives exemplified within Metaresearch and HPS contexts respectively.

Metaresearch on replication practices

Metaresearch accounts of replication practices tend to prioritise the goal of evaluating contemporary practices for generating and disseminating scientific knowledge. In doing so, Metaresearch approaches can help to improve practices associated with the concept of replication within one or more fields. Many of these studies take the value of replication (variously characterised) as a given, focusing on explaining the low rate of scientific research claims being replicated and proposing strategies for increasing this rate. For example, within the psychological sciences, replication has become a well-recognised criterion for improving the credibility of research claims (Lindsay, 2015; Vazire, 2018). Meanwhile, replicability concerns also feature in other disciplinary-specific reform discussions – from social sciences (A. C. Chang & Li, 2018; Christensen & Miguel, 2018), to ecology (Fraser et al., 2020; Kelly, 2019), biology (Ioannidis et al., 2015; Ryan, 2011) and the biomedical sciences (Begley C. Glenn & Ioannidis John P.A., 2015; Danchev et al., 2019).

Interest in evaluating replication practices received a boost following some large-scale projects that repeated multiple published studies. In these projects, repeated studies frequently failed to generate the originally reported findings (Camerer et al., 2016, 2018; Klein et al., 2014; Open Science Collaboration, 2015). These studies prompted investigation into the proportion of studies that attempt to replicate previous findings, and why replication studies so often find different results to original studies. For example, several studies focused on demonstrating how rarely replications are published within specific disciplines, including ecology (Kelly, 2006, 2019), psychology (Makel et al., 2012), and medicine (Begley & Ellis, 2012).

These types of studies demonstrated that, while assumed to be routine in principle, the actual practices required to establish the replicability of research are both infrequently attempted and more complicated to conduct than expected. One response to this realisation within Metaresearch contexts has been to examine the barriers to attempting replication studies. These barriers include institutionalised structures that disincentivise replicating research claims by preferentially funding and publishing novel research (Agnoli et al., 2020; Fraser et al., 2020). In addition, this pressure to publish large volumes of research with novel, exciting findings can lead researchers to engage in 'questionable research practices' (John et al., 2012). For instance, publication pressure can prompt researchers to adopt questionable research practices that increase the chance of finding a statistically significant relationship regardless of whether instances of the phenomenon exist as an intersubjectively testable regularity, (e.g. conducting large numbers of analyses and only reporting the ones with statistically significant effects). Relatedly, questionable research practices can make research seem more exciting than is warranted to increase the chance of publication (e.g. describing post hoc hypotheses as if they had been derived *a priori*). This

suggests that, as well as disincentivising replication attempts, the pressure to publish novel findings contributes to the prevalence of questionable research practices that, in turn, decrease the likely replicability of research claims.

Even when research is conducted thoroughly and without these questionable research practices, Metaresearch studies have highlighted how conducting replications can be difficult because of the incomplete reporting of methods. For instance, not only does incomplete reporting make it logistically more difficult to repeat a study, it also leaves room for hidden moderators. These hidden moderators are unreported aspects of the original study that are influential on the results. Jay Van Bavel and colleagues (2016) found that studies with high susceptibility to hidden moderators in the reproducibility project: psychology (Open Science Collaboration, 2015), were less likely to replicate. However, evidence from the Many Labs studies suggests that for most studies, undescribed differences in research setting have minimal effect on the results (Ebersole et al., 2016; Klein et al., 2014, 2018). The concern about hidden moderators commonly arises in response to new replication studies, despite evidence that, at least in psychology, the influence may be minimal (Klein et al., 2018) and concern that claiming hidden moderators post hoc makes science unfalsifiable (Zwaan et al., 2018).

Given the small number of replications being attempted and the low replicability rate, Metaresearch studies have arisen questioning the role of replication in the scientific discourse, including in the survey of ecologists mentioned earlier. In this survey, a wide range of practices were found to count as replication studies (Fraser et al., 2020). In addition, while an overwhelming majority of the surveyed ecologists expressed support for replication studies, few checked for replication studies when assessing research claims themselves (Fraser et al., 2020). As Hannah Fraser and colleagues (2020) note, the practice reported by ecologists echoes comments made about replication in interviews of biochemists by sociologists Michael Mulkay and Nigel Gilbert (1986). For example, most biochemists reported disinterest in replicating the work of other researchers, yet all claimed that their own work had been independently replicated (Mulkay & Gilbert, 1986). This tendency to have simultaneous yet contradictory expectations about replication practices has been described as a form of conceptual slippage (Mulkay & Gilbert, 1986). On the one hand, researchers drew on a narrow conceptualization of replicating the work of others – primarily in terms of repeating an experiment directly, to test the reliability of experimental results. Whereas, when thinking about whether their own research had been replicated, a broader concept of replication emerged – where researchers reported their own research claim as replicated given the persistence of their claimed effect when elements of their original experiment have been intentionally varied.

Despite these shifting assumptions about the actual practice of conducting replications, the value of establishing the replicability of research claims still tends to provide a foundational expectation when evaluating scientific practices. For example, O'Dea and colleagues (2021) respond to lack of replication studies being conducted within ecology and evolutionary biology by suggesting that incentives could help to better align the behaviours of researchers with beliefs of researchers (in the value of replications). This perspective can also be seen in the range of improvement initiatives prompted by the increasing awareness of the barriers to conducting replications in practice. For instance, several initiatives are targeted at improving the efficiency and effectiveness of peer review. Examples of this

include increasing or decreasing the anonymity of peer review, introducing results-blind peer review or Registered Reports to ensure that reviewers decisions aren't influenced by how exciting the results are (Hamilton et al., 2020) , and outsourcing peer review to a community of researchers who emphasise transparent research practices (O'Grady, 2021). Several of these proposals have been picked up at scale. For example, within psychology there has been a shift in norms towards more rigorous reporting statistical power and pre-registration requirements (Strickland & De Cruz, 2021). More recently, the pressure to quickly share medical research during the COVID-19 pandemic has driven an initiative to speed up the Registered Report process to reduce the preliminary review time to 48 hours (Brock, 2020).

In addition to developing initiatives to reduce the barriers to conducting replications as a way of establishing the reliability of research findings, some approaches within the Metaresearch context are shifting the focus towards evaluating how to improve the robustness of generalisable research claims beyond the initial study conditions. This shift can be seen in Nosek and Errington's (2020) proposed redefinition of the term replication mentioned earlier. Similarly, Tal Yarkoni (2020) argues that, while efforts to establish that an effect is repeatedly precise are important, this only matters if the research claim is meaningful at a scale that is generalisable beyond the constraints of one initial study. Relatedly, Peder Isager and colleagues (2020) suggests that decisions about what to replicate should consider the degree of uncertainty about the reliability of the research findings; the epistemic value attributed to how a research claim is replicated; and the utility and risks an un-replicated finding presents when the associated claims are disseminated to contexts with high-impact consequences (such as in engineering design or policy decisions).

Complementing this shift, others are exploring how, in addition to methodological reforms, Metaresearch approaches can help to challenge the current institutional and cultural conditions that disincentivise and restrict improvements to scientific practices, including but not limited to replications. For example, Rink Hoekstra and Simine Vazire (2020) argue that a key aspect of credibility is exercising intellectual humility – including being transparent about research limitations and non-defensive in response to the constructive criticism replication practices can generate. Building on this point, Richard Ramsey (2021) argues that there are also limitations specific to a given field of research and greater modesty about these would facilitate more collaborative and cumulative practices (including replications).

As these examples of the evaluative perspective found within Metaresearch demonstrate, there are many challenges to conducting replications in practice. These challenges draw attention to the importance of specifying when and how to conduct specific forms of replications in relation to the resource constraints researchers operate within, and the broader conditions within which research claims are generated and disseminated. In doing so, these Metaresearch approaches contribute a range of strategies for identifying and improving how replication practices are evaluated.

Historical, philosophical, and social studies of replication practices

While sometimes overlapping with Metaresearch contexts, studies of replication practices within the fields of history, philosophy, and/or the social studies of the sciences (HPS) often start from a different perspective. Within this context, examinations of specific investigative practices tend to contribute to broader analytic accounts of the diachronic conditions within

which bodies of scientific knowledge emerge.[18] As such, HPS accounts of replication practices often prioritise the goal of analysing the dynamic processes within which investigative practices contribute to the generation and dissemination of bodies of scientific knowledge over long periods of time. Within this context, fewer initiatives are proposed for improving the conditions for attempting replications or increasing the rate of successful replications. Instead, HPS approaches tend to focus on investigating the disputed role of replication practices in generating and disseminating research claims.

This analytic approach has drawn attention to the multiple epistemic functions attributed to various types of replications. For example, one epistemic function attributed to replication practices that repeat studies (both as a close copy of the original and by altering key components) is to *increase confidence in the reliability of a research finding* by establishing that the claimed effect is genuine instance of the target phenomena (i.e., an intersubjectively testable regularity) rather than an isolated coincidence or an artefact of the study design.[19] Another common epistemic function attributed to replication practices is to *increase confidence in the generalisability of a research claim* by establishing that the relevant effects are robust enough to persist in all relevant conditions.

These two epistemic functions have been explicated in philosophical discussions of replication (Cartwright, 1991; Feest, 2019; Romero, 2019), and are also apparent within historical and sociological studies of scientific practice (Mulkay & Gilbert, 1986; Steinle, 2016). Of the two functions, less epistemic value tends to be ascribed to establishing reliability than establishing robustness and generalisability. This difference in value can be attributed to an entrenched assumption that procedurally routine scientific practices (such as calibration of instruments and experimental equipment) are epistemically unimportant (Feest, 2019; Romero, 2020; Schickore, 2011). These assumptions can be seen in philosophical accounts of replicability where the routine practices of establishing the reliability of research findings are taken for granted (Radder, 1992). The relative epistemic unimportance of establishing reliability also reflects a disinterest in the (presumed) routine practice of checking whether a given set of numerical/statistical calculations is reliably repeatable (Cartwright, 1991; Collins, 1991).

While routine reliability checks may be of minimal interest, HPS accounts of scientific practices do include a range of approaches to examining the role of replications in establishing that the persistence of a reported effect justifies the robustness and generalisability of the associated research claim. Within this context, replication practices (of various types) are positioned as part of a collection of similarly important practices. For example, historical studies have demonstrated that repetition of research findings can play a minimal role, if any, in establishing the robustness of research claims (Leonelli, 2018; Steinle, 2016). Relatedly, historical studies of neuroscientific practices detail the dissemination of un-replicable research claims, demonstrating that establishing replicability is not always necessary for resolving doubts about research claims (Bogen, 2001).

In this context, while establishing the reliability of a finding is important, it is only one step (and not always even an essential step) in a collection of practices involved in establishing

[18] As within Metaresearch contexts, consensus about the relevant analytic concepts for studying the replication practices within HPS has been elusive (Schickore, 2011). In addition, despite there being historical and sociological accounts of the limited role of repeating experiments within the sciences, many philosophers of science took it for granted that experiments were repeatable, research findings replicable, and research claims reproducible (Radder, 1992).

[19] For more on phenomena in terms of their intersubjectively testable regularity, see (Bogen, 2001, p. 493).

the persistence of a reported effect, and the generalisability of the associated research claim. For example, Mieke Boon (2012) argues that multiple elements of scientific practice work together to establish the robustness of a scientific claim. While these practices do include those that help to establish the reliability of a finding when repeating the initial experiments (more or less closely), other practices are required to help establish whether the stability of the phenomena of interest can persist under different research conditions (Boon, 2012).

These accounts highlight the limitations of specific forms of replications and reflect the analytic approach within HPS to describing the range of practices that contribute to generating and disseminating robust scientific knowledge.[20] This includes HPS accounts that highlight that no single characterisation of replicability can account for all the actual replication practices. For instance, Stephan Guttinger's (2020) examination of the role of 'micro-replications' within the life sciences draws attention to the limits of expecting a one-size-fits-all conceptualisation of replicability as a universal epistemic norm.

The variability of replication practices documented across different scientific fields also offers some context for the responses from HPS scholars to recent proposals to re-define the term 'replication', and to prescribe criteria for establishing the validity of research claims. These responses often acknowledge the value of replications of research findings in fields where establishing the procedural reliability of reported findings are both possible and demonstrably valuable. For example, the procedural repeatability of reported findings is considered a reasonable expectation in those fields with a high degree of control over research variables (e.g., laboratory experiments), and a reliance on statistical inferences (such as in null-hypothesis significance testing) (Leonelli, 2018; Schickore, 2011).

However, these approaches also highlight that there is a gap between the expected role of replications (as a crucial and routine part of all investigative practices) and what actually happens in the messy social contexts these practices are situated within. For instance, Sabina Leonelli (2018) notes that there are many different possible criteria for evaluating whether or not a given replication attempt was successful. In addition, even in scenarios where there are systematic attempts to replicate research findings, a range of constraints limit the impact of these on establishing (or revising) a given research claim (Romero, 2016).

Analyses of replication practices in a range of high-control contexts have led some HPS scholars to resist the expectation that the in-principle replicability of research findings ought to be able to extend to fields and contexts where these practices are not already found. For example, Feest (2019) notes the justified anxiety around low replication rates within fields dominated by hypothesis-testing studies, yet resists the extension of these replication expectations to all of the sciences (especially exploratory research). Leonelli (2018) also argues that there is limited epistemic value in expecting the replication of research findings in irreproducible studies, such as those involving perishable samples carried out in idiosyncratic environmental conditions. Leonelli (2018) goes on to suggest that the replicability of research findings is not an appropriate way to establish the credibility of research claims in research fields with limited control of variables (such as in ecological observational practices) or those involving hermeneutical practices (such as qualitative

[20] For more on how HPS approaches cultivate interest in the 'mosaic' formed by the wide range of localised scientific practices, see (H. Chang, 2011).

analyses of interview data). Relatedly, Bart Penders et al., (2019, p. 2) proposes that "replicatory expectations should be contextualised in relation to specific epistemic communities rather than across the board". Meanwhile, Felipe Romero (2020) offers an argument that supports efforts to engage with the social structures of scientific practices in reforms that address failures to replicate research claims.

These calls for caution regarding efforts to prescribe 'one right way' to conduct replications can be understood within the context of the broader trends within HPS scholarship of analysing localised investigative practices. For instance, the examples from HPS accounts offered each contribute to broader analyses of historical, philosophical, and sociological accounts of scientific practices. Within this context, replications are positioned as just one of many practices that can contribute to establishing reliability of research findings and/or robustness of research claims across various scientific fields. Further, this perspective within HPS scholarship emphasises the value of being explicit about the context-specific expectations for a given replication. In doing so, this analytic approach contributes to our understanding of the disputed roles of (various types of) replications in the generation and dissemination of robust scientific knowledge.

Using the replicability concept as a goal-directed tool

Drawing on the view of the replicability concept as a goal-directed tool for studying investigative practices, we draw attention to some converging insights emerging from the evaluative and analytic approaches exemplified within Metaresearch and HPS contexts. To this end, we draw on the insights from scholarship on concept-use outlined in the earlier section as well as examples of context-specific uses of the replicability concept described above. In doing so, we hope to show that there is value in examining the replicability concept as used to pursue context-specific goals within Metaresearch and HPS respectively.

As detailed earlier, even when the referential component of a scientific concept is disputed it can be used productively in pursuit of context-specific goals. Similarly, we explored how the replicability concept is used productively for pursuing context-specific goals despite the unresolved disagreements about what counts as (any given type of) replication practice. Building on the sketch of how the replicability concept is used in Metaresearch and HPS contexts above, we suggest that variable uses of the replicability concept contribute to (at least) two complementary goals. Focusing on the Metaresearch context, we provided examples of how the replicability concept can be used for evaluating contemporary scientific practice in ways that contribute to improvement initiatives. Turning to the HPS context, we offered examples of how the replicability concept can be used for analysing the dynamic conditions through which replication practices contribute to a range of diverse practices involved in generating and disseminating knowledge over time. Reflecting on these variable goal-directed uses of the replicability concept offers an avenue for exploring some of the shared challenges and converging insights emerging from these two approaches.

The most obvious challenge shared by both evaluative and analytic approaches to studying replication practices is that there is a range of inconsistent terminologies and ambiguous characteristics associated with the replicability concept. Drawing on the insight that scientific concepts can function productively even when consensus about the referent remains unresolved, we take the view that this variability need not undermine any of the uses of the replicability concept for studying scientific practices. Taking this view, and

treating the replicability concept as a tool, suggests that this challenge can be mitigated by clarifying how the specific (implied or explicit) characterisation of replication practices being studied is relevant for the goal of evaluating and/or analysing scientific practices.

Another challenge is that, as with scientific concepts used within investigative practices, uses of the replicability concept for studying scientific practices can carry-along inferential associations that go unexamined. These expectations about the relative epistemic value of various types of replication practices are interconnected yet contribute to different assumptions depending on the goal-directed uses of the replicability concepts. For instance, Metaresearch approaches often start from the premise that scientific findings should always be replicable (at least in principle). Taking the epistemic value of replications as a given, Metaresearch approaches then focus on evaluating, explaining, and changing those scientific practices that disincentive replication practices and/or contribute to low rates of successful replications. In this context, the disagreements about what counts as replications and the disputes over how to measure if a replication is successful are both hurdles to overcome in improving how investigative practices contribute to scientific knowledge. In contrast, within HPS contexts there tends to be analytic focus on examining the epistemic functions of replications in and of themselves – both as actually practiced within specific sciences and as a more general expectation about how various sciences should be practiced.[21] However, within both contexts, the variable characterisations of replicability suggest that replications are expected to contribute to at least two broadly understood epistemic functions: (i) checking the reliability of a claim by testing that the reported effect is stable when repeating the original procedures as closely as possible in all relevant respects;[22] (ii) confirming the robustness of a claim – either by testing whether the reported effect persists when deliberately altering key aspects of the original procedures,[23] or by demonstrating that the claim generalises to similar effects found under different conditions.[24]

Of these functions, the difficulty of establishing the reliability of research findings is often obscured by the entrenched assumption that published scientific findings can be treated as reliable in principle. For instance, HPS approaches tend to position the practices involved in establishing the reliability of research findings as just one of many practices required to establish the robustness of research claims. By taking this approach, HPS studies of scientific practices help to highlight how different types of replications contribute to the many different practices for establishing the reliability and/or robustness of research findings. Relatedly, a similar assumption underlies the crisis-response to the high-profile Metaresearch failures to reproduce findings when repeating previously published studies as

[21] See Guttinger (2020) for more on the tension between descriptive accounts of replication practices and normative accounts of replicability expectations.

[22] Also see: routine checking practices (Cartwright, 1991; Collins, 1991); exact repeatability (Radder, 1992); procedural replication (Bogen, 2001); verification tests (Clemens, 2017); direct replications (Altmejd et al., 2019); micro-replications (Guttinger, 2019); computational reproducibility (Leonelli, 2018; Peng, 2011); internal validity testing (Fraser et al., 2020; Schmidt, 2009); and reproducibility (Nosek et al., 2021).

[23] Also see: replications (Radder, 1992); reproductions (Schickore, 2011); direct replications (Feest, 2019; Fidler & Wilcox, 2018; Schmidt, 2009); re-sampling replications (Machery, 2020); robustness (Nosek et al., 2021).

[24] Also see: hypothesis validity tests (Cartwright, 1991); conceptual replications (Fidler & Wilcox, 2018; Schmidt, 2009); construct validity tests (Fraser et al., 2020); extensions (Machery, 2020; Zwaan et al., 2018); replications (Nosek et al., 2021; Nosek & Errington, 2020).

closely as possible. This crisis led to Metaresearch approaches attempting to identify and resolve the conditions within which such failures occur (maintaining the expectation that research findings should be reliable in principle even if difficult in practice). In evaluating a range of constraints that limit replication attempts and/or hinder the success of such attempts in practice, Metaresearch approaches help to demonstrate that the practices required to establish the reliability of research findings are not as routine as expected.

More recently, some Metaresearch uses of the replicability concept have shifted interest towards the potentially more tractable problem of establishing the robustness of generalisable research claims (e.g., Nosek & Errington, 2020; Yarkoni, 2020). Meanwhile, HPS uses of the replicability concept have started to consider the epistemic relevance to replication practices of procedural considerations such as measurement errors and software maintenance (e.g., Hocquet & Wieber, 2021; Machery, 2020). While these trajectories overlap, the evaluative Metaresearch and analytic HPS approaches each generate distinct insights about replication practices. On the one hand, Metaresearch studies highlight the range of resource constraints and disciplinary conditions that constrain replication practices. For instance, in evaluating contemporary scientific practices Metaresearch uses of the replicability concept draw attention to disincentives that constrain attempts to establish both the reliability of research findings and the robustness of research claims. In response, Metaresearch approaches have supported a range of initiatives that seek to overcome these disincentives. As the earlier example of the rapid-version of Registered Reports implemented for COVID-19 research illustrates, these initiatives are also impacted by conditions well beyond specific investigative practices. Meanwhile, HPS studies offer a view of replications as one of the many scientific practices through which research claims can become established and disseminated. This view is supported by studies that analyse the diachronic dynamics of scientific practices, to suggest that, even if better incentivised, replication practices may not necessarily be an appropriate way to assess the findings generated by all the different investigative practices. Reflecting this, the earlier examples help to illustrate why proposals for improving the rates of (successful) replications common within the Metaresearch context are not as common within HPS studies of replication practices.

While pursuing different goals, the evaluative and analytic uses of the replicability concept can be drawn together to provide converging insights about when and why establishing the reliability of a research finding is relevant to demonstrating the robustness of the associated research claim. For example, the insights emerging from both Metaresearch and HPS contexts mentioned above converge to support a view that attempts to prescribe a universal off-the-shelf set of criteria for assessing the reliability of research findings and/or robustness of research claims inevitably obscure the diverse situated contingencies of actual scientific practices. This convergence is often overlooked.

This includes Metaresearch accounts of replications that draw attention to the range of barriers inhibiting attempts to replicate research findings. As detailed above, these barriers include the logistical difficulties of repeating studies, poorly reported methods and questionable practices in published research, disputes over how to evaluate the success of a replication attempt, and broader institutionalised disincentives for both conducting replicable research practices and attempting to replicate prior research. Complementing this consideration, HPS approaches highlight that replications function as one of many

scientific practices that help establish research claims in ways that contribute to the generation and dissemination of robust scientific knowledge. This perspective suggests that it is important to situate expectations of replication practices within relevant epistemic communities – as demonstrated by the earlier examples of investigative practices, such as exploratory research and observational field studies, that should not be expected to provide replications to establish the credibility of research claims (Feest, 2019; Leonelli, 2018; Penders et al., 2019).

While far from a systematic review, this exploration demonstrates that there is value in examining the context-specific uses of concepts as goal-directed tools for studying scientific practice. In the case of the replicability concept, considering two different goals – exemplified by the evaluative goals within Metaresearch and the analytic goals within HPS contexts respectively – has helped to highlight distinct yet converging insights. Given the entrenched assumptions these different uses share, this exploration also hints at additional lessons we could learn from the scholarship on the uses of concepts as tools within investigative practices. For instance, the variable uses of the replicability concept suggest that rather than unify the multiple characterisations of replication practices there is space for appreciating each in terms of their appropriateness for the goal of evaluating and/or analysing scientific practices. Similarly, viewing concepts as tools (used for individuating types of practices for further study) draws attention to the importance of justifying and documenting how and why a given concept is appropriate for a specific goal within a given context. For instance, as our exploration of the range of characteristics associated with the replicability concept suggests, there are often unexamined inferential associations carried along by ambiguously individuated concepts. One way to raise awareness of these inferential associations would be to cultivate practices of documenting and justifying how the concepts used to study replication practices draw on sets of characteristics for individuating those practices that are appropriate for a context-specific goal. While the details of cultivating such practices are beyond the present scope, we hope to have demonstrated that there is value in considering how a given use of the replicability concept contributes to context-specific analytic and/or evaluative goals.

In conclusion, we drew on literature detailing the uses of concepts within investigative practice to treat the replicability concept as a tool for studying scientific practices. By taking this view, we were able to explore the productive uses of two goal-directed uses of the replicability concept: i) evaluating contemporary investigative practices to inform initiatives for improving the generation and dissemination of scientific knowledge, and ii) analysing a range of historical and contemporary investigative practices to contribute to understanding the long-term dynamics of generating and disseminating scientific knowledge. To compare how uses of the replicability concept function as a tool for pursuing these two goals, we drew on examples of primarily evaluative-uses of the replicability concept from Metaresearch contexts, and primarily analytic-uses of the replicability concept from HPS contexts. As mentioned earlier, this distinction between Metaresearch and HPS contexts is a device to highlight the complementary insights emerging within studies of scientific practices from two context-specific uses of the replicability concept for pursuing different goals. In offering this exploration of the variable uses of the replicability concept we have sought to illustrate that there is value in examining the uses of concepts as tools for studying scientific practices. In the case of the replicability concept, this value includes highlighting how Metaresearch and HPS contexts provide complementary insights about

when and how replication practices offer a relevant strategy for assessing a given research claim.

Acknowledgements

In 2021, this manuscript was been submitted as a chapter in an edited volume (forthcoming):
Bloch-Mullins, C. L., & Arabatzis, T. (Eds.) *Concepts, Induction, and the Growth of Scientific Knowledge.*

This research was conducted on the unceded lands of the Wurundjeri peoples of the Kulin nation. It was conducted as part of a broader project that was funded in part by DARPA's Systematizing Confidence in Open Research and Evidence (SCORE) program under cooperative agreement N660011924016.

References

Agnoli, F., Fraser, H., Thorn, F. S., & Fidler, F. (2020). *Australian and Italian Psychologists' View of Replication* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/ks48e

Alister, M., Vickers-Jones, R., Sewell, D. K., & Ballard, T. (2021). How Do We Choose Our Giants? Perceptions of Replicability in Psychological Science. *Advances in Methods and Practices in Psychological Science*, *4*(2), 25152459211018200. https://doi.org/10.1177/25152459211018199

Altmejd, A., Dreber, A., Forsell, E., Huber, J., Imai, T., Johannesson, M., Kirchler, M., Nave, G., & Camerer, C. (2019). Predicting the replicability of social science lab experiments. *PLOS ONE*, *14*(12), e0225826. https://doi.org/10.1371/journal.pone.0225826

Arabatzis, T. (2019). What are scientific concepts? In K. McCain & K. Kampourakis (Eds.), *What is Scientific Knowledge?: An Introduction to Contemporary Epistemology of Science* (pp. 85–99). Routledge.

Barba, L. A. (2018). *Terminologies for Reproducible Research* [Preprint]. http://arxiv.org/abs/1802.03311

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. https://doi.org/10.1038/483531a

Begley C. Glenn & Ioannidis John P.A. (2015). Reproducibility in Science. *Circulation Research*, *116*(1), 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819

Bloch, C. L. (2012a). Early concepts in investigative practice—The case of the virus. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 191–218). De Gruyter.

Bloch, C. L. (2012b). Scientific kinds without essences. In A. Bird, B. D. Ellis, & H. Sankey (Eds.), *Properties, powers, and structures: Issues in the metaphysics of realism* (pp. 233–256). Routledge.

Bloch-Mullins, C. L. (2020). Scientific Concepts as Forward-Looking: How Taxonomic Structure Facilitates Conceptual Development. *Journal of the Philosophy of History*, *1*(aop), 1–27. https://doi.org/10.1163/18722636-12341438

Bogen, J. (2001). `Two as good as a hundred': Poorly replicated evidence in some nineteenth-century neuroscientific research. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *32*(3), 491–533. https://doi.org/10.1016/S1369-8486(01)00013-9

Boon, M. (2012). Understanding Scientific Practices: The Role of Robustness Notions. In L. Soler, E. Trizio, T. Nickles, & W. Wimsatt (Eds.), *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science* (pp. 289–315). Springer Netherlands. https://doi.org/10.1007/978-94-007-2759-5_12

Brigandt, I. (2002). *A Theory of Conceptual Advance: Explaining Conceptual Change in Evolutionary, Molecular, and Evolutionary Developmental Biology—D-Scholarship@Pitt* [Doctoral Dissertation, University of Pittsburgh]. http://d-scholarship.pitt.edu/8849/

Brigandt, I. (2003). Species Pluralism Does Not Imply Species Eliminativism. *Philosophy of Science*, *70*(5), 1305–1316. https://doi.org/10.1086/377409

Brigandt, I. (2012). The dynamics of scientific concepts. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 75–103). De Gruyter.

Brock, J. (2020, April 14). Rapid Registered Reports initiative aims to stop coronavirus researchers following false leads. *Nature Index*. https://www.natureindex.com/news-blog/rapid-registered-report-coronavirus-aims-to-stop-researchers-following-false-research-leads

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436. https://doi.org/10.1126/science.aaf0918

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

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

Chang, A. C., & Li, P. (2018). Is Economics Research Replicable? Sixty Published Papers From Thirteen Journals Say "Often Not." *Critical Finance Review*, *7*. https://doi.org/10.1561/104.00000053

Chang, H. (2011). The Philosophical Grammar of Scientific Practice. *International Studies in the Philosophy of Science*, *25*(3), 205–221. https://doi.org/10.1080/02698595.2011.605244

Chang, H. (2014). Epistemic activities and systems of practice: Units of analysis in philosophy of science after the practice turn. In L. Soler, S. Zwart, M. Lynch, & V. Israel-Jost (Eds.), *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science* (pp. 67–79). Taylor and Francis.

Christensen, G., & Miguel, E. (2018). Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature*, *56*(3), 920–980. https://doi.org/10.1257/jel.20171350

Clemens, M. A. (2017). The Meaning of Failed Replications: A Review and Proposal. *Journal of Economic Surveys*, *31*(1), 326–342. https://doi.org/10.1111/joes.12139

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

Danchev, V., Rzhetsky, A., & Evans, J. A. (2019). Centralized scientific communities are less likely to generate replicable results. *ELife*, *8*, e43094. https://doi.org/10.7554/eLife.43094

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

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

Fabrigar, L. R., Wegener, D. T., & Petty, R. E. (2020). A Validity-Based Framework for Understanding Replication in Psychology. *Personality and Social Psychology Review*, *24*(4), 316–344. https://doi.org/10.1177/1088868320931366

Farkas, K. (2013). A sense of reality. In F. Macpherson & D. Platchias (Eds.), *Hallucination: Philosophy and psychology* (pp. 399–415). MIT Press.

Feest, U. (2010). Concepts as tools in the experimental generation of knowledge in cognitive neuropsychology. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, *4*(1), 173–190. https://doi.org/10.4245/sponge.v4i1.11938

Feest, U. (2012). Exploratory experiments, concept formation, and theory construction in psychology. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 167–190). De Gruyter.

Feest, U. (2019). Why Replication is Overrated. *Philosophy of Science*, 705451. https://doi.org/10.1086/705451

Fidler, F., & Wilcox, J. (2018). Reproducibility of Scientific Results. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018). Metaphysics Research Lab, Stanford University.

Fletcher, S. C. (2021). How (not) to measure replication. *European Journal for Philosophy of Science*, *11*(2), 57. https://doi.org/10.1007/s13194-021-00377-2

Fraser, H., Barnett, A., Parker, T. H., & Fidler, F. (2020). The role of replication studies in ecology. *Ecology and Evolution*, *10*(12), 5197–5207. https://doi.org/10.1002/ece3.6330

Gómez, O. S., Juristo, N., & Vegas, S. (2010). Replications Types in Experimental Disciplines. *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, 3:1-3:10. https://doi.org/10.1145/1852786.1852790

Guala, F. (2003). Experimental Localism and External Validity. *Philosophy of Science*, *70*(5), 1195–1205. https://doi.org/10.1086/377400

Guttinger, S. (2019). A New Account of Replication in the Experimental Life Sciences. *Philosophy of Science*, *86*(3), 453–471. https://doi.org/10.1086/703555

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

Hamilton, D. G., Fraser, H., Hoekstra, R., & Fidler, F. (2020). Meta-Research: Journal policies and editors' opinions on peer review. *Elife*, *9*, e62529.

Hocquet, A., & Wieber, F. (2021). Epistemic issues in computational reproducibility: Software as the elephant in the room. *European Journal for Philosophy of Science*, *11*(2), 38. https://doi.org/10.1007/s13194-021-00362-9

Hoekstra, R., & Vazire, S. (2020). *Intellectual humility is central to science* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/edh2s

Ioannidis, J. P. A. (2018). Meta-research: Why research on research matters. *PLOS Biology*, *16*(3), e2005468. https://doi.org/10.1371/journal.pbio.2005468

Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: Evaluation and Improvement of Research Methods and Practices. *PLOS Biology*, *13*(10), e1002264. https://doi.org/10.1371/journal.pbio.1002264

Isager, P. M., van Aert, R. C. M., Bahník, Š., Brandt, M. J., DeSoto, K. A., Giner-Sorolla, R., Krueger, J., Perugini, M., Ropovik, I., van 't Veer, A. E., Vranka, M. A., & Lakens, D.

(2020). *Deciding what to replicate: A formal definition of "replication value" and a decision model for replication study selection.* [Preprint]. https://doi.org/10.31222/osf.io/2gurz

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kelly, C. D. (2006). Replicating Empirical Research In Behavioral Ecology: How And Why It Should Be Done But Rarely Ever Is. *The Quarterly Review of Biology*, *81*(3), 221–236. https://doi.org/10.1086/506236

Kelly, C. D. (2019). Rate and success of study replication in ecology and evolution. *PeerJ*, *7*, e7654. https://doi.org/10.7717/peerj.7654

Kindi, V. (2012). Concept as vessel and concept as use. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 23–46). De Gruyter.

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

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225

Leonelli, S. (2018). Rethinking Reproducibility as a Criterion for Research Quality. In *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise* (Vol. 36B, pp. 129–146). Emerald Publishing Limited. https://doi.org/10.1108/S0743-41542018000036B009

Leydesdorff, L. (1991). In Search of Epistemic Networks. *Social Studies of Science*, *21*(1), 75–110. http://www.jstor.org.ezp.lib.unimelb.edu.au/stable/285323

Lindsay, D. S. (2015). Replication in Psychological Science. *Psychological Science*, *26*(12), 1827–1832. https://doi.org/10.1177/0956797615616374

Machery, E. (2020). What Is a Replication? *Philosophy of Science*, *87*(4), 545–567. https://doi.org/10.1086/709701

MacLeod, M. (2012). Rethinking scientific concepts for research contexts: The case of the classical gene. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 47–74). De Gruyter.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research: How Often Do They Really Occur? *Perspectives on Psychological Science*, *7*(6), 537–542. https://doi.org/10.1177/1745691612460688

McDermott, M. B. A., Wang, S., Marinsek, N., Ranganath, R., Ghassemi, M., & Foschini, L. (2019). *Reproducibility in Machine Learning for Health* [Preprint]. arXiv:1907.01463 [cs, stat]. http://arxiv.org/abs/1907.01463

McDermott, R. (2011). Internal and External Validity. In J. N. Druckman, D. P. Greene, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 27–40). Cambridge University Press.

Mulkay, M., & Gilbert, G. N. (1986). Replication and Mere Replication. *Philosophy of the Social Sciences*, *16*(1), 21–37. https://doi.org/10.1177/004839318601600102

Nelson, N., Ichikawa, K., Chung, J., & Malik, M. (2020). *Mapping the discursive dimensions of the reproducibility crisis: A mixed methods analysis* [Preprint]. MetaArXiv. https://doi.org/10.31222/osf.io/sbv3q

Nersessian, N. J. (2012). Modeling practices in conceptual innovation: An ethnographic study of a neural engineering research laboratory. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 245–270). De Gruyter.

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., Almenberg, A. D., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L., Schönbrodt, F., & Vazire, S. (2021). *Replicability, Robustness, and Reproducibility in Psychological Science* [In Press]. PsyArXiv. https://doi.org/10.31234/osf.io/ksfvq

O'Dea, R. E., Parker, T. H., Chee, Y. E., Culina, A., Drobniak, S. M., Duncan, D. H., Fidler, F., Gould, E., Ihle, M., Kelly, C. D., Lagisz, M., Roche, D. G., Sánchez-Tójar, A., Wilkinson, D. P., Wintle, B. C., & Nakagawa, S. (2021). Towards open, reliable, and transparent ecology and evolutionary biology. *BMC Biology*, *19*(1), 68. PDF. https://doi.org/10.1186/s12915-021-01006-3

O'Grady, C. (2021, April 19). Fifteen journals to outsource peer-review decisions. *Science | AAAS*. https://www.sciencemag.org/news/2021/04/fifteen-journals-outsource-peer-review-decisions

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

Patil, P., Peng, R. D., & Leek, J. T. (2016). What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. *Perspectives on Psychological Science*, *11*(4), 539–544. https://doi.org/10.1177/1745691616646366

Penders, B., Holbrook, J. B., & de Rijcke, S. (2019). *Rinse and repeat: The situated value of replication across epistemic cultures* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/nfjb6

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, *334*(6060), 1226–1227. https://doi.org/10.1126/science.1213847

Peng, R. D., & Hicks, S. C. (2021). Reproducible Research: A Retrospective. *Annual Review of Public Health*, *42*(1), 79–93. https://doi.org/10.1146/annurev-publhealth-012420-105110

Peterson, D., & Panofsky, A. (2020). *Self-Correction in Science: The Diagnostic and Integrative Motives for Replication*. SocArXiv. https://doi.org/10.31235/osf.io/96qxv

Plesser, H. E. (2018). Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Frontiers in Neuroinformatics*, *11*. https://doi.org/10.3389/fninf.2017.00076

Pöyhönen, S. (2013). Natural Kinds and Concept Eliminativism. In *EPSA11 Perspectives and Foundational Problems in Philosophy of Science* (pp. 167–179). Springer, Cham. https://doi.org/10.1007/978-3-319-01306-0_14

Radder, H. (1992). Experimental Reproducibility and the Experimenters' Regress. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1992*, 63–73. JSTOR. http://www.jstor.org/stable/192744

Ramsey, R. (2021). A Call for Greater Modesty in Psychology and Cognitive Neuroscience. *Collabra: Psychology*, *7*(1). https://doi.org/10.1525/collabra.24091

Romero, F. (2016). Can the behavioral sciences self-correct? A social epistemic study. *Studies in History and Philosophy of Science Part A*, *60*, 55–69. https://doi.org/10.1016/j.shpsa.2016.10.002

Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, *14*(11), e12633. https://doi.org/10.1111/phc3.12633

Romero, F. (2020). The Division of Replication Labor. *Philosophy of Science*, *87*(5), 1014–1025. https://doi.org/10.1086/710625

Rouse, J. (2011). Philosophy of Science and Science Studies in the West: An Unrecognized Convergence. *East Asian Science, Technology and Society*, *5*(1), 11–26. https://doi.org/10.1215/s12280-010-9145-y

Ryan, M. J. (2011). Replication in Field Biology: The Case of the Frog-Eating Bat. *Science*, *334*(6060), 1229–1230. https://doi.org/10.1126/science.1214532

Schickore, J. (2011). The Significance of Re-Doing Experiments: A Contribution to Historically Informed Methodology. *Erkenntnis (1975-)*, *75*(3), 325–347. JSTOR. http://www.jstor.org/stable/41476727

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

Schmidt, S. (2017). Replication. In M. C. Makel & J. A. ; Plucker (Eds.), *Toward a more perfect psychology: Improving trust, accuracy, and transparency in research* (pp. 233–253).

Smith, E. T. (2018a). *The Structured Uses of Concepts as Tools: Comparing fMRI Experiments that Investigate either Mental Imagery or Hallucinations* (MINERVA) [Doctor of Philosophy, School of Historical and Philosophical Studies, University of Melbourne]. http://hdl.handle.net/11343/219955

Smith, E. T. (2018b). Interdependent Concepts and their Independent Uses: Mental Imagery and Hallucinations. *Perspectives on Science*, *26*(3), 360–399. https://doi.org/10.1162/posc_a_00278

Smith, E. T. (2019). Examining the Structured Uses of Concepts as Tools: Converging Insights. *Filozofia Nauki*, *28*(4), 7–22. https://doi.org/10.14394/filnau.2019.0024

Smith, E. T. (2020). Examining tensions in the past and present uses of concepts. *Studies in History and Philosophy of Science Part A*, *84*, 84–94. https://doi.org/10.1016/j.shpsa.2020.08.004

Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (2014). Introduction. In L. Soler, S. Zwart, M. Lynch, & V. Israel-Jost (Eds.), *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science*. Taylor and Francis.

Steinle, F. (2010a). Concepts, facts, and sedimentation in experimental science. In D. J. Hyder & H.-J. Rheinberger (Eds.), *Science and the life-world: Essays on Husserl's Crisis of European sciences* (pp. 199–214). Stanford University Press.

Steinle, F. (2010b). Scientific facts and empirical concepts: The case of electricity. In M. Epple & C. Zittel (Eds.), *Science as cultural practice* (pp. 32–43). Akademie Verlag.

Steinle, F. (2012). Goals and fates of concepts: The case of magnetic poles. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 105–126). De Gruyter.

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). Wiley.

Strickland, B., & De Cruz, H. (2021). Editorial: Replicability in Cognitive Science. *Review of Philosophy and Psychology*, *12*(1), 1–7. https://doi.org/10.1007/s13164-021-00531-y

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences*, *113*(23), 6454–6459. https://doi.org/10.1073/pnas.1521897113

Vazire, S. (2018). Implications of the Credibility Revolution for Productivity, Creativity, and Progress: *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691617751884

Waters, C. K. (2014). Shifting Attention from Theory to Practice in Philosophy of Biology. In M. C. Galavotti, D. Dieks, W. J. Gonzalez, S. Hartmann, T. Uebel, & M. Weber (Eds.), *New Directions in the Philosophy of Science* (pp. 121–139). Springer International Publishing. https://doi.org/10.1007/978-3-319-04382-1_9

Yarkoni, T. (2020). The generalizability crisis. *The Behavioral and Brain Sciences*, 1–37. https://doi.org/10.1017/S0140525X20001685

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. https://doi.org/10.1017/S0140525X17001972