2022

The Virtue of Appropriate Transparency: Research Psychology and the Replication Crisis

Barry Mannolini-Winwood
The University of Notre Dame Australia

THE VIRTUE OF APPROPRIATE TRANSPARENCY: RESEARCH PSYCHOLOGY AND THE REPLICATION CRISIS

Barry Mannolini-Winwood

Bachelor of Science (Cognitive Neuroscience and Health Psychology; Philosophy)

Submitted in fulfillment of the requirements for the Master of Philosophy

School of Philosophy and Theology

Fremantle Campus

May, 2022

Declaration

To the best of the candidate's knowledge, this thesis contains no material previously published by another person, except where due acknowledgement has been made.

This thesis is the candidate's own work and contains no material which has been accepted for the award of any other degree or diploma in any institution.

SIGNATURE: 

PRINT NAME: Barry Mannolini-Winwood

DATE: 3 May 2022

Abstract

I argue that one step toward addressing the replication crisis in psychology is through the application of appropriate transparency in research communication. Drawing upon virtue epistemology, I propose that appropriate transparency can be defined as a contextually informed point between understatement and overstatement. As such, I contend that judging appropriate transparency in research psychology requires a shared research design nomenclature (RDN). As researchers' definitions of terms related to research psychology are not necessarily identical between individuals, it is beneficial to use transparent definitions of pivotal concepts to facilitate effective communication between researchers and their audience. However, the boundaries of appropriate transparency in research communications differ depending upon the primary audience. Thus, I argue that appropriate transparency in communication must be judged in the context of the intended audience of the specific research communication. The lack of a shared RDN introduces unnecessary strain on this process; whereas, a shared RDN would aid researchers in identifying the appropriate level of transparency for their research communication audience. A RDN should be as explicit as possible without being unnecessarily proscriptive; it will be most useful when retaining sufficient flexibility for experts to apply practical judgement regarding appropriate transparency. Having outlined the benefits of a RDN, I then build upon Stefan Schmidt's and others' work to propose a RDN that could provide a guide to researchers on appropriate transparency in communicating research-design information. While adhering to a shared RDN requires extra effort, it also provides common ground for primary researchers and replication researchers when designing projects and communicating efficiently. Moreover, a shared RDN can be leveraged to demonstrate appropriate transparency for non-experts with minimal expenditure of effort by the researcher. By adhering to a RDN, experts and non-experts alike are provided with a common point of reference, thus encouraging restored confidence in research interpretation.

Acknowledgements

Primarily, I wish to acknowledge the support of my wife and best friend, Sarai Mannolini-Winwood. I would never have completed this thesis without Sarai's unwavering support and infinite patience. Thank you for always being there.

I wish to thank Professor Renée Köhler-Ryan for stepping in as my Primary Supervisor, and her newborn son David for patiently lending me his mother. Without the support, guidance, insight, and encouragement of Professor Köhler-Ryan, I'm not sure if my thesis would have been as recognisably my own. Thank you for your mentorship, your time, and your patience during your own significant life event.

I wish to thank Doctor Brentyn Ramm for stepping in as my Secondary Supervisor. Doctor Ramm was first a friend and a colleague, neither of which (I hope) stopped him from honestly critiquing my work. Both Professor Köhler-Ryan and Doctor Ramm provided nuanced insight into the philosophical concepts within my thesis; however, having Doctor Ramm available to also critique the psychology/science concepts was invaluable. Thank you for your guidance and your ongoing friendship.

I wish to acknowledge the input of my original supervisory team: Professor Philip Mathews and Doctor Richard Hamilton.

This thesis was edited by Elite Editing, and editorial intervention was restricted to Standards D and E of the Australian Standards for Editing Practice.

Support for this research was provided via an Australian Government Research Training Program (RTP) Scholarship.

Finally, I wish to thank and apologise to my friends and family who have had to listen to my ramblings over the course of my research. If you continue to listen, that's on you.

Any mistakes herein are my own.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

In this thesis, I argue that judging appropriate transparency in research psychology requires a shared research design nomenclature (RDN). I discuss that all researchers and their audiences do not necessarily use identical definitions of concepts related to psychological research; I then elucidate the reasons it is useful to unpack pivotal concepts in this field to facilitate effective communication between researchers and their audience.[1] I build upon these points to explain the utility inherent in interpreting the replication crisis in research psychology as a crisis of confidence in research conclusions, which supports the position that transparency in research communication is a necessary step[2] in restoring confidence in research conclusions.[3] I highlight how transparency, a pivotal term in psychological research, is nevertheless conceived of differently by individual researchers, as well as their audience; leading to my argument that *appropriate transparency* in any psychological research communication can only be judged in the context of the intended audience for that communication.[4] I explore how the lack of a shared RDN introduces unnecessary strain on replication research interpretation and extrapolation and why a shared RDN would aid researchers in identifying the appropriate level of transparency for their intended audience.[5] I justify the argument that an RDN should be as explicit as possible without being unnecessarily proscriptive; yet, I also assert that an RDN is most useful when it retains sufficient flexibility for experts to apply practical judgement regarding appropriate transparency.[6] I provide examples to demonstrate that although a shared RDN requires extra effort, it also provides common ground for primary researchers and replication researchers when designing projects and for communicating efficiently with each other. Thus, these examples support my argument that a shared RDN can be leveraged to demonstrate appropriate transparency for non-experts with minimal expenditure of effort by researchers.[7] Last, I reaffirm my claim that identifying appropriate transparency in relation to replication research will help restore some of the confidence in research conclusions and thus be

[1] As discussed further in Chapter 1.
[2] In a range of other necessary steps.
[3] As discussed further in Chapter 2.
[4] As discussed further in Chapter 3.
[5] As discussed further in Chapter 4.
[6] As discussed further in Chapter 5.
[7] As discussed further in Chapter 6.

one step (among the many others needed) towards addressing the replication crisis in psychological research.

1.1. An Example of the Crisis

The rate of failure to replicate psychological research results is concerning. In Chapter 2, I will explore the discourse about appropriate and inappropriate replication failure in more detail. In short, replication failure is only appropriate when it is additive to the observations and theoretical knowledge base of the research topic; it is inappropriate when it is the result of avoidable issues within research design and practice. To illustrate the problem of failure that psychology researchers face, a large-scale replication study from 2015 is useful.

The Open Science Collaboration (OSC) organised a global project to reperform (direct replication[8]) 100 published psychological research projects.[9] Of these 100 original studies:

- three reported statistically non-significant findings and
- 97 reported statistically significant findings.

Of the 97 replication projects that reperformed the same research as the original statistically significant studies:

- 62 reported statistically non-significant findings and
- 35 reported statistically significant findings.

That is, only 36% (35/97) of the replication studies obtained results that supported the original studies' statistical conclusions—64% (62/97) of the original research results could not be replicated by the OSC replication studies. In short, almost two-thirds of the replications failed.

In the published report summarising the entire project, the OSC acknowledged the unavailability of a single method to evaluate the success of an individual replication attempt. Thus, it is worth considering that a comparison of significant findings between the original and the replication studies does not necessarily provide a comprehensive picture. However, this example usefully illustrates a problem facing

[8] A direct replication changes as little of the original research project design as possible, as further defined in Section 5.2.1.
[9] Open Science Collaboration, 'Estimating the Reproducibility of Psychological Science', *Science* 349, no. 6251 (2015). https://doi.org/10.1126/science.aac4716.

psychological research as a whole; namely, on a variety of measures, in psychological research, results are not being replicated by follow-up projects. The statistical lack of replication has led to the claim that psychological research is in a state of replication crisis. Nevertheless, researchers such as Wolfgang Stroebe and Fritz Strack argued against the use of the term crisis in relation to replication in psychological research.[10]

1.2. Blurred Disciplinary Demarcations

As identified by the OSC, the causes of replication failure cannot be reduced to a simple singular factor.[11] Therefore, an interdisciplinary approach must be adopted to probe the replication crisis. This thesis builds upon a virtue epistemology framework (philosophy) to identify and promote specific research design variables (science). As Sebastion De Haro contended, it works from the orientation that the boundary between science and philosophy is not clear-cut, and hence, work in philosophy is relevant to science and vice versa.[12] De Haro contended that because science and philosophy often explore the same concepts, it may be useful to extrapolate from the discourse around both. Similarly, Gabriel Abend asserted that concepts traditionally placed within one (science/philosophy) can be useful for exploring the specifics of the other (philosophy/science).[13] Thus, in line with De Haro and Abend, it is useful to conceptualise and respond to issues identified within the replication crisis using ideas from both.

An open, synergistic relationship between philosophy and science is useful to both disciplines. De Haro argued that despite some opinions to the contrary, the discipline of science still needs philosophy. He addressed the synergistic relationship between both disciplines via three key arguments:

i. Denying the need for philosophy in science necessitates the use of philosophical arguments.
ii. Philosophical reasoning is necessary to account for the gaps in empirical observation.

[10] I rebut this argument in Chapter 2.
[11] Open Science Collaboration, 'Estimating Reproducibility'.
[12] Sebastian De Haro, 'Science and Philosophy: A Love-Hate Relationship', *Foundations of Science* 25, no. 2 (2020). https://doi.org/10.1007/s10699-019-09619-2.
[13] Gabriel Abend, 'What Are Neural Correlates Neural Correlates Of?' *BioSocieties* 12, no. 3 (2017). https://doi.org/10.1057/s41292-016-0019-y.

iii. Everything science explores, philosophy also explores using different aims and methods.

This third argument (in particular) supports an interdisciplinary approach to topics such as the replication crisis. Such an approach is more useful than a monodisciplinary approach because, as Abend claimed, if both disciplines are already exploring the same concepts, then it would be a wasted effort for both to replicate the same foundational work.

An interdisciplinary approach is particularly useful when exploring the concept of transparency as a proposed solution to the replication crisis. Transparency is a concept regularly promoted as a solution to the replication crisis in psychology.[14] Nevertheless, there is little evidence of transparency being functionally (and actionably) defined within the discourse. The discourse around transparency focuses almost exclusively on justifying (or rejecting) it and fails to provide practical advice to psychology researchers. A conceptual definition of transparency alone does not provide the level of detail necessary to implement the concept in research practice. However, as the next section discusses, traditionally philosophical concepts provide a practical framework upon which to identify, justify and construct actionable advice useful to practicing researchers.

1.3. Philosophical Foundations

The focus of this thesis is the relationship between transparency and the replication crisis in psychological research. I approach this relationship from a foundation wherein it is assumed that differences in the (implicit) meaning used by individuals in relation to terms within the discourse have a subsequent effect on both research practice and research interpretation. In this section, I summarise philosophical reasons that it is invalid to assume shared meanings of research terminology between researchers, hence justifying the need for an explicitly defined research language. Alfred N. Whitehead proposed that a common fallacy of human thinking is the assumption that our thoughts are necessarily indicative of a concrete reality that is universally shared by all.[15] For Whitehead, it is invalid to assume that any human

[14] Dalson Figueiredo Filho et al., 'Seven Reasons Why: A User's Guide to Transparency and Reproducibility', *Brazilian Political Science Review* 13, no. 2 (2019). https://doi.org/10.1590/1981-3821201900020001; Kevin C. Elliott, 'A Taxonomy of Transparency in Science', *Canadian Journal of Philosophy* (2020): 1–14. https://doi.org/10.1017/can.2020.21.
[15] Alfred North Whitehead, *Science and the Modern World* (New York: Macmillan, 1954): 52–75.

demarcation encapsulates all possible characteristics of reality outside our own thoughts.[16] In a similar vein to Whitehead's fallacy of misplaced concreteness, David J. Chalmers argued that the cause of a dispute between individuals is often the result of an unacknowledged divergence in meaning between crucial, but contentious, terms.[17] Gabriel Abend similarly argued that it is common for neuroscience researchers to conceptualise and operationalise the objects of their inquiry in ways that are incommensurable with the ways of other neuroscience researchers.[18] Synthesising the ideas of Whitehead, Chalmers and Abend, I find examples of theoretical arguments in the replication crisis discourse that can be attributed (at least in part) to divergent meaning attribution between contentious terms.

Alfred N. Whitehead's fallacy of misplaced concreteness is a foundational orientation for this thesis that can be seen underpinning my arguments against universally shared meaning. According to Whitehead, it is not always acknowledged that although human thought exists on a spectrum between the abstract and the concrete, it can never achieve pure abstraction nor pure concreteness.[19] Whitehead asserted that it is common for individuals to assume that their thoughts, concepts or ideas represent a comprehensive understanding of the concrete reality outside of themselves. Similarly to Alfred Korzybski's argument that '[a] map is not the territory it represents', Whitehead argued that regardless of how accurate and/or useful an individual's conceptualisation of a given idea is, a thought is never an exact representation of the concrete thing itself; instead, it is a useful abstraction.[20] Whitehead did not argue against the utility of abstraction, nor against a concrete reality outside of human thought. Rather, he justified the importance of acknowledging the fallacy inherent in assuming that any human thought can contain all the details of a given thing. For example, when an individual conceives of an orange, it is likely that their thoughts only contain the abstract information necessary for them to usefully interact with the object; generally, the conception of a particular orange does not contain specifics of the atomic structure of said orange. On noting this interpretation of Whitehead's fallacy of misplaced concreteness, one can then consider how the abstract nature of

[16] Whitehead.
[17] David J. Chalmers, 'Verbal Disputes', *The Philosophical Review* 120, no. 4 (2011). https://doi.org/10.1215/00318108-1334478.
[18] Abend, 'What Are Neural Correlates?'
[19] Whitehead, *Science and the Modern World*, 19-20.
[20] Alfred Korzybski, *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics* (New York; Institute of General Semantics, 1958); Whitehead, 19-20.

thought can be applied to language. Whitehead argued that although all human thought is abstract and no human thought can exhaustively represent the world, abstractions tend to vary between individuals. The idiomatic (often implicit) meaning that individuals construct around ill-defined terms can affect simple communication between them; the impact of an ill-defined nomenclature to accurately convey idiosyncratically constructed meaning can be further unpacked via David J. Chalmers' writing on contentious terms in verbal disputes.

David J. Chalmers argued that often the cause of dispute between two parties can be resolved by unpacking a single contentious term into its constituent parts.[21] Chalmers contended that an argument between two individuals over a particular conclusion is frequently the result of the use of one (or more) terms to which each party (unknowingly) attributes different meanings. Chalmers used the example of different meanings attached to the term 'go round' as changing the outcome of the question '[d]oes the man go round the squirrel or not?' Using a previous example presented by William James, Chalmers described a scenario in which a man and a squirrel are both circling a tree; the squirrel is on the tree trunk, and the man is slightly away from the tree.[22] They remain on opposite sides such that the tree is perpetually between them. If the term 'go round' is assumed to include the squirrel's movement, the answer will differ from the answer if the term is assumed to be independent of the squirrel's movement. If the difference in meaning attribution is not addressed, then a disagreement may result that is purely dependent upon individual interpretation of the contentious term. The opposite, but related, argument was presented by Gottlob Frege when discussing sense and reference.[23] Frege described a scenario wherein two individuals argue about the specifics of the evening star and the morning star without understanding that they are both referring to the same object by different names. Thus, if it is accepted that all human thought is abstract and no human thought can exhaustively represent the world (Whitehead), and that not all abstractions between individuals are guaranteed to be identical (Chalmers and Frege), then Abend's critique of incommensurability between researcher conceptualisation of theoretical and methodological terminology is plausible.

[21] Chalmers, 'Verbal Disputes', 1.
[22] Chalmers, 1.
[23] Gottlob Frege, 'Sense and Reference', *The Philosophical Review* 57, no. 3 (1948): 214–6.

Abend argued that there is evidence that some researchers who explore psychological phenomena do not conceptualise and operationalise the objects of their inquiry in commensurable ways.[24] Abend also argued that neuroscientists often seek the neural correlates (physical manifestations within the brain) of concepts such as morality, empathy and creativity without a shared nomenclature to define expressions of the concept, and nor do they work from a shared nomenclature regarding ways to identify such expressions. That is, to rephrase Abend:

- Researcher A is exploring Concept X.
- Researcher A defines Concept X in relation to Meaning 1.

The problem Abend addressed is that Researcher B may also be exploring Concept X via a (implicitly) different meaning attribution:

- Researcher B is exploring Concept X.
- Researcher B defines Concept X in relation to Meaning 2.

Since both researchers attribute different meanings to one concept (Meaning 1 v. Meaning 2), their research is no longer (perhaps unintentionally) immediately comparable.

However, Abend's arguments need not be limited to neuroscience and differences in meaning attribution related to neural correlates; similar (unacknowledged) differences in theoretical meaning attribution can be found in general psychological discourse. One example of incommensurability between meaning attribution in the psychological discourse can be found in two articles arguing for and against classifying addiction as a brain disease. The answer/response to whether addiction can or should be classified as a disease of the brain is not relevant here. Instead, I focus on the problem wherein an unacknowledged difference in definitions of disease results in the rebuttal no longer addressing the same concept as the original argument.

Alan I. Leshner argued that addiction is a brain disease, and, in response to Leshner, Neil Levy argued that it is not.[25] Despite being published more than a decade apart, both researchers presented arguments defending their position, seemingly to support the importance of research into the topic. However, both papers defined the key term

[24] Abend, 'What Are Neural Correlates?' 415–8.
[25] Alan I. Leshner, 'Addiction Is a Brain Disease, and It Matters', *Science* 278, no. 5335 (1997): 45–7. https://doi.org/10.1126/science.278.5335.45; Neil Levy, 'Addiction Is Not a Brain Disease (and It Matters)', *Frontiers in Psychiatry* 4 (2013): 24. https://doi.org/10.3389/fpsyt.2013.00024.

'disease' through different criteria—a point that Levy, who wrote the later of the two essays, did not explicitly address. Rightly or wrongly, Leshner's definition did not require a biological component for disease classification. Rightly or wrongly, Levy's definition did require a biological component. Therefore, when Levy argued to reject Leshner's classification of addiction as a disease by arguing against an inherently biological component within addiction, Levy's definition of addiction differed subtly from that of Leshner. Regardless of the merits of either argument, the fact that both parties attached a different meaning to the term disease, and that the later study did not explicitly address the fact that both used the term differently, demonstrates that Abend's observations of undisclosed disputes around conceptualisation and operationalisation are also present in the broader psychological discourse.

Thus, the philosophical foundation for my exploration of the replication crisis in psychological research is built upon three principles:

(i) All human thought is abstract, and no human thought can exhaustively represent the world (Whitehead).
(ii) Not all human thoughts, even when expressed in the same words, have the same meaning (Chalmers and Frege).
(iii) Unacknowledged differences in meaning between key terms are an observable issue within psychological discourse (Abend).

With these three points assumed, I aim to show the reason that a shared RDN is an important component in addressing the replication crisis in psychological research. Specifically, I argue that while increased transparency is an oft-repeated solution for the replication crisis (independent of the needs of replication), inappropriate transparency can be detrimental to research interpretation. Without a shared RDN against which to judge transparency, arguments for and against transparency in research communication are more distracting, than actionable, for researchers.

The lack of a shared RDN is observed in the different characteristics attributed by different researchers to key terms within the discourse. In turn, these implicit verbal disputes result in invalid inferences being made because of differences between the idiosyncratic meaning that individuals tacitly apply to different terms. As will be discussed in Section 5.2, not all research replications (by design) are conducted using the same replication methodologies. However, that there is more than one valid way

of conducting a replication project is rarely explicitly addressed within the discourse. When a replication failure between research projects is observed, the strengths and limitations of the specific replication methodology used are rarely addressed. Thus, conclusions drawn about the reliability of a specific replication methodology are invalidly generalised to all replication methodologies.

1.4. Thesis Structure

The main argument of the thesis proceeds as follows.

Chapter 2 unpacks how the replication crisis in psychological research is more clearly understood as a crisis of confidence in research results. Replication failure in and of itself is not an inherently undesirable outcome. Rather, appropriate replication failure is a useful tool to refine knowledge developed through research methodologies. Instead, the crisis is found in the quantifiable level of inappropriate replication failure that indicates a systemic problem with research practices. Inappropriate replication failure is the result of avoidable errors in individual research that, in turn, reduces confidence in all research results. That much of the cause of replication failure is entirely avoidable adds to a reduction of confidence in research results and has encouraged the rise of agnotological thinking.

Chapter 3 then examines how transparency in research, as a concept for combating the replication crisis, has higher practical utility when positioned as a virtue between two vices. Although transparency in research practices is commonly proposed as an answer to the replication crisis, the specific ways to achieve transparency are often left vague. This vagueness has left room for some to argue that transparency is more harmful than useful. This chapter argues that when transparency is considered an epistemic virtue, it can be operationalised more effectively, particularly in its relation to two epistemic vices: understatement and overstatement. However, for the proposal that transparency be treated as a virtue to be useful, the specifics of what should be made transparent needs to be defined.

Therefore, Chapter 4 develops an outline of a shared nomenclature for psychological research. This is one step towards establishing a shared point of reference for researchers to consider when selecting an appropriate level of transparency. If transparency can be useful for combating the replication crisis, then there also needs to be some consensus about what should be made transparent. A shared RDN is one

way of actualising transparency in research design and research communication. The argument for transparency being too difficult to achieve stems in part[26] from the lack of a shared nomenclature for researchers to leverage when communicating their processes and decisions. A shared nomenclature provides a starting point to build in appropriate transparency in research communication and thus identify what to share with a combined audience. Sharing everything with everyone may be as unhelpful as sharing nothing with anyone.

Chapter 5 builds upon the foundation developed in Chapter 4 through the example of an RDN. The chapter proposes a revision of Stefan Schmidt's RDN. Schmidt's RDN privileged experimental methodologies, whereas I propose a RDN that can apply to all psychological research methodologies. The list of research design variables proposed is neither exhaustive nor prescriptive. Rather, the categories of research design variables that this chapter extrapolates from Schmidt's study are intended to aid researchers in making more informed, transparent research design decisions. Ultimately, not every variable will be appropriate to every research project, and a list of all possible research design variables would be untenable. Instead, the thesis highlights the primary categories of variables that researchers need to be aware of when designing and communicating their research. The chapter builds on Schmidt's study and argues that these categories will make it easier for replication researchers to recreate their projects, as well as aid in restoring confidence in research conclusions.

1.5. Chapter Summary

The primary conclusions of this chapter are as follows:

- All researchers and their audiences do not necessarily use identical definitions of concepts related to psychological research.
- It is useful to unpack pivotal terms in psychological research to facilitate effective communication between researchers and their audience.

In this chapter, I have outlined the philosophical foundation upon which I have based my justification about the utility found in a shared RDN. I framed the issue via discussion of the OSC's large-scale replication project, which demonstrated that the rate of replication failure is at a level concerning to some. However, as I will unpack further in the next chapter, it is worth reiterating that the rates of replication failure

[26] Further discussion can be found in section 2.3.

alone are not sufficient justification for the use of the term crisis. The reasons that a replication attempt failed are as important (if not more so) to the broader analysis and the development of knowledge, as the failure itself. As a step towards meeting the goal of ensuring appropriate interpretation and shared meaning, I summarised three key concepts from traditionally philosophical discourse that can be usefully applied to psychological research design. In accordance with Whitehead, my thesis rests on the assumption that all human thought is an imperfect abstraction of a concrete reality that exists independent of human experience. I then used arguments from Chalmers and Frege to show that the level and characteristics of specific abstractions (terms) can be expected to differ between individuals. Next, I provided support for this position in a research context via Abend's observations concerning the unacknowledged differences in the way individual researchers attach meaning to specific terms in the discourse about psychological research.

As I develop my argument further, I will build upon this interpretation to create a framework for judging appropriate transparency in research design and communication via the description of relevant research design terminology. I contend that a shared RDN will provide guidance for researchers to demonstrate appropriate transparency in their research design and communication; which, in turn, will aid in identifying the specific utility of individual replication failures and in rebuilding general confidence in psychological research practices and conclusions. Currently, it is difficult to distinguish between useful and incidental replication failures, which has reduced the confidence of experts and non-experts alike in much of psychological research.[27]

[27] Following Stephen John, I refer to individuals with a formal research training background as experts; Stephen John, 'Epistemic Trust and the Ethics of Science Communication: Against Transparency, Openness, Sincerity and Honesty', *Social Epistemology* 32, no. 2 (2018): 75–6. https://doi.org/10.1080/02691728.2017.1410864.

Chapter 2. The Replication Crisis in Psychological Research

In this chapter, I assume the following primary conclusions from the previous chapter:

- All researchers and their audiences do not necessarily use identical definitions of concepts related to psychological research.
- It is useful to unpack pivotal concepts in psychological research to facilitate effective communication between researchers and their audience.

I discuss the reasons that the replication crisis in psychological research is a tangible concern that has led to reduced confidence in its conclusions. I summarise and align my assessment with those of Harold Pashler and Eric-Jan Wagenmakers, namely, that the replication crisis in psychological research is best framed as a crisis of confidence in psychological research conclusions, due to the misinterpretation of the role of replication failure in research.[28] I argue that replication failure is a critical tool for refining the boundaries of research phenomena. However, failure alone will not result in refining the knowledge base about the research topic. Replication failure can be caused by tangential factors independent of the research topic. In these cases, replication failure will serve to instead aid in refining research practices. However, the difference between replication failure due to research design and due to gaps in the research topic is rarely explicitly identified in the discourse regarding the replication crisis. This lack of attention results in an overly narrow interpretation of the replication crisis, which:

- obfuscates both the utility of replication failure for knowledge development,
- as well as the importance of preventing predictable replication failure to avoid wasting resources and negatively influencing confidence in research conclusions.

2.1. A Crisis of Confidence

The label ‘replication crisis in psychological research’ is an umbrella term for a multiplicity of issues that have contributed to a rise in agnotological thinking in

[28] Harold Pashler and Eric–Jan Wagenmakers, ‘Editors’ Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?’ *Perspectives on Psychological Science* 7, no. 6 (2012). https://doi.org/10.1177/1745691612465253.

relation to psychological science.[29] The existence of some form of crisis in regard to replication of psychological research is (currently) relatively uncontroversial within the discourse; however, descriptions of the crisis are rarely identical.[30] The replication crisis is often oversimplified, with arguments for and against frequently privileging one symptom over other equally valid concerns, while simultaneously misinterpreting valid research outcomes.[31] Thus, to find an overarching theme that accounts for the diversity of the characteristics of the problem, some have instead interpreted the replication crisis in psychological research as a crisis of confidence in its research results. By reframing the crisis as one in which the quantity of observed replication failures leads to a crisis of confidence in psychological research conclusions, the minority of arguments against the existence of a crisis can be more accurately considered.

In 2012, Harold Pashler and Eric-Jan Wagenmakers explicitly argued that the replication crisis should be reframed as a crisis of confidence.[32] They argued that the rate of replication failures was damaging the confidence that could (and should) be placed in knowledge derived from psychological research. Pashler and Wagenmakers made explicit what is often either assumed or overlooked by other commentators; despite achieving focal attention, appropriate replication failure is not a problem. Rather, several comorbid problems are contributing to inappropriate replication failure.

The lack of confidence in psychological research conclusions can be understood in the context of reduced confidence in all research conclusions. In a broader research context, Stephen John identified internal expert disputes as contributing to the lack of non-expert confidence in research conclusions.[33] Although it is only a small component of his overall argument, John described how replication failure contributes to an overall rise in agnotological thinking around research conclusions

[29] Harold Pashler and Christine R. Harris, 'Is the Replicability Crisis Overblown? Three Arguments Examined', *Perspectives on Psychological Science* 7, no. 6 (2012). https://doi.org/10.1177/1745691612463401.
[30] Stefan Schmidt, 'Shall We Really Do It Again? The Powerful Concept of Replication Is Neglected in the Social Sciences', *Review of General Psychology* 13, no. 2 (2009). https://doi.org/10.1037/a0015108; Pashler and Wagenmakers, 'Crisis of Confidence?'
[31] Wolfgang Stroebe and Fritz Strack, 'The Alleged Crisis and the Illusion of Exact Replication', *Perspectives on Psychological Science* 9, no. 1 (2014). https://doi.org/10.1177/1745691613514450; Pashler and Harris, 'Is the Replicability Crisis Overblown?'; Pashler and Wagenmakers.
[32] Pashler and Wagenmakers.
[33] John, 'Epistemic Trust', 78.

amongst non-experts.[34] In summary, due to an inaccurate folk philosophy of science, the more instances of replication failure non-experts observe, the less confidence they display in any or all research conclusions.[35] However, a lack of confidence in research conclusions and practices is not limited to non-experts.

A lack of confidence in the conclusions of entire subdisciplines can also be observed among experts in psychological research from tangential fields. In an open letter, psychologist and researcher Daniel Kahneman publicly singled out *priming* researchers as a group who (in his opinion) needed to make more effort to support the confidence of their research conclusions.[36] Although a researcher himself, Kahneman self-identified in his letter as not being a member of the *priming* researcher community. However, Kahneman's open letter demonstrated a personal lack of confidence in commonplace *priming* research practices. Kahneman claimed—whether fairly or not—that the broader issues contributing to the replication crisis in psychology were more apparent in priming research. While the context around Kahneman's letter explains why he chose to single out *priming* research as 'the poster child for doubts about the integrity of psychological research', as I will now discuss, the issues he raised are not exclusive to *priming* research.

Published statistical errors are one cause of reduced confidence in psychological research. Statistical results have often been misreported in published psychology research. In 2011, Marjan Bakker and Jelte M. Wicherts analysed 281 published psychological research papers, reporting that around 18% (35) of the articles that reported statistics (194) included 'gross' statistical errors.[37] In further analysis, 39 articles were found to contain at least one unsupported statistical conclusion; Bakker and Wicherts deduced that one out of seven published research psychology articles contains a statistical conclusion not supported by the analysis presented. The

[34] John's full argument is unpacked in more detail in the discussion in Chapter 4; however, it is useful to summarise a subset of his argument here.
[35] John, 'Epistemic Trust', 78-80.
[36] The use of specific words or phrases to elicit an unconscious response in an individual; see Andrew M. Colman, *A Dictionary of Psychology*, 4th ed. (2015), s.v. 'priming'; Daniel Kahneman, 'A Proposal to Deal with Questions About Priming Effects', September 26, 2012, Letter, from *Nature*, accessed June 22, 2021, https://www.nature.com/news/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf.
[37] Marjan Bakker and Jelte M. Wicherts, 'The (Mis)Reporting of Statistical Results in Psychology Journals', *Behavior Research Methods* 43, no. 3 (2011): 666. https://doi.org/10.3758/s13428-011-0089-5.

researchers concluded that (commonly) mistakes were the result of various human errors, including, but not limited to:

- incomplete statistical information provided,
- incorrect mathematical symbols used,
- inappropriate statistical tests applied,
- failure to follow reporting conventions,
- rounding errors, and
- copy-paste errors.

Notably, identifying human error as a problem that must be addressed is not to suggest that human error will ever be entirely avoidable.

Nevertheless, as Bakker and Wicherts identified, the magnitude of errors making it through the peer-review process inappropriately inflate the prevalence of replication failures.[38] Identifying human error is a problem that replication research aims to detect, since misreported statistics are unlikely to be supported by independent replication. [39] However, human error is not the sole cause of reduced confidence in psychological research outcomes. Reduced confidence is a shared symptom pointing towards a multiplicity of underlying problems, fraud being another such problem.

Uncovered fraud in psychological research is another factor that has negatively affected the general confidence of psychological research outcomes. Replication is lauded as a primary tool for combating research fraud, and thus, the identification of fraudulent research is sometimes discussed under the umbrella term of the replication crisis.[40] Yet, in the biggest cases of fraud in psychological research, replication is generally used to re-examine research conclusions only once fraudulent practices are suspected.

Among the most prominent examples of widespread fraud within psychological research is the work of Diederick Stapel.[41] Stapel is a former professor of social psychology who returned his Ph.D. after the depths of his fraudulent research practices were uncovered. In one case, Stapel's research data aligned with the

[38] Bakker and Wicherts, 'The (Mis)reporting of Statistical Results', 675.
[39] Further discussion can be found in Section 5.1.
[40] Schmidt, 'Shall We Really Do It Again?', 93.
[41] Wolfgang Stroebe, Tom Postmes and Russell Spears, 'Scientific Misconduct and the Myth of Self-Correction in Science', *Perspectives on Psychological Science* 7, no. 6 (2012): 671–2. https://doi.org/10.1177/1745691612460687.

expectations of his hypothesis so well that a commentator joked '[i]t is as if he made up these data himself'.[42] Unfortunately, the joke was found to be an accurate assessment. At the height of the controversy, commentators outside the psychology discipline used Stapel to cast doubt on the entire research discipline, identifying the ease with which spurious conclusions can be manipulated to appear within psychological research.[43] Yet, while replication of Stapel's research has been used to identify which of his projects were likely built using fraudulent data, replication played little part in initially identifying Stapel's propensity towards fraud.

The combination of misreported statistics and intentional fraud are only two of the reasons for the reduced confidence in psychological research. With some overlap between misreporting and fraud, another factor is questionable research practices (QRPs). QRP is an umbrella term for any method of manipulating statistical data analysis to promote a predetermined research agenda. The raw data are generally validly and honestly collected; however, the legitimacy of the data analysis method is sometimes (at best) questionable.[44] Yet, significantly, although QRPs are undeniably problematic, they are distinct from outright fraud. With the identification of the various causes discussed (as well as a range not discussed), the psychological research community is introducing reform to rebuild confidence in research conclusions.

As a collective, researchers are attempting to rebuild confidence in research results through various initiatives. They are drawing attention to the rate of misreporting of statistics in publications.[45] Probably, some errors will inevitably be published. However, this does not justify apathy towards meeting the goal of ensuring accuracy in the field. The research community is currently challenging publication standards by raising questions about how the publication of psychological research can be more open and transparent.[46] This challenge to publication standards can be observed in the

[42] Stroebe, Postmes and Strack, 'Scientific Misconduct', 671.
[43] Jelte M. Wicherts, 'Psychology Must Learn a Lesson from Fraud Case', *Nature (London)* 480, no. 7375 (2011): 7. https://doi.org/10.1038/480007a; Stroebe, Postmes and Strack, 676–7.
[44] Further discussion can be found in Section 2.3.
[45] Bakker and Wicherts, 'The (Mis)reporting of Statistical Results'.
[46] John P. A. Ioannidis, 'Why Science is Not Necessarily Self-Correcting', *Perspectives on Psychological Science* 7, no. 6 (2012). https://doi.org/10.1177/1745691612464056; John P. A. Ioannidis, 'Why Most Published Research Findings Are False', *PLoS Medicine* 2, no. 8 (2005). https://doi.org/10.1371/journal.pmed.0020124.

rise of various web-based platforms on which research can be freely shared and discussed.[47]

Despite a minority of crisis deniers still maintaining their position, the disciplinary consensus encourages researchers to actively work towards improving research standards.[48] Evidence of this improvement can be observed through the commonplace public rejection of QRPs by practicing psychology researchers.[49] The problem may not be completely resolved, but the consensus no longer approves of disingenuous manipulations of research results.

With the discourse beginning to identify and address the problematic issues that lead to certain replication failures, replication research is being reaffirmed as a useful protective practice and increased rates of replication research are being encouraged.[50] Although specific websites have been established to provide an avenue for replication research to bypass the usual publication streams, dedicated replication projects have also been introduced. [51] While publication bias is being addressed, and increased efforts are still required, the importance and relevance of replication in psychological research is being addressed more positively, implying that a shift towards replication research is no longer considered an unnecessary burden.[52]

2.2. Utility of Failure

Replication failures are not inherently positive or negative. Rather, when interpreted in the appropriate context, a replication failure is informative. If warranted, replication failures are desirable. If research results are not reproducible, most members within any given discipline would generally want to know this fact. Again, it is not the failure to replicate that warrants the label of crisis. Rather, one of the

[47] Further discussion of various factors contributing to the replication crisis are discussed in Section 2.3.
[48] Abend, 'What Are Neural Correlates?'; Bakker and Wicherts, 'The (Mis)reporting of Statistical Results'; Open Science Collaboration, 'Estimating Reproducibility'; Pashler and Harris, 'Is the Replicability Crisis Overblown?'; Schmidt, 'Shall We Really Do It Again?'; Stroebe, Postmes and Strack, 'Scientific Misconduct'; Wicherts, 'Psychology Must Learn a Lesson'.
[49] Leslie K. John, George Loewenstein and Drazen Prelec, 'Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling', *Psychological Science* 23, no. 5 (2012). https://doi.org/10.1177/0956797611430953; Stuart Vyse, 'P-Hacker Confessions: Daryl Bem and Me', *Skeptical Inquirer* 41, no. 5 (2017).
[50] Sophia Crüwell et al., 'Seven Easy Steps to Open Science: An Annotated Reading List', *Zeitschrift Für Psychologie* 227, no. 4 (2019). https://doi.org/10.1027/2151-2604/a000387; Schmidt, 'Shall We Really Do It Again?'
[51] Open Science Collaboration, 'Estimating Reproducibility'.
[52] Schmidt, 'Shall We Really Do It Again?'; Open Science Collaboration, 'Estimating Reproducibility'.

results of persistent replication failure is damage to the confidence that can be placed in the knowledge synthesised from the broader research body. When (primary and/or replication) research projects are being performed and reported at an avoidably low standard, legitimate questions about the practical utility of the research are raised. However, the identification of appropriate/inappropriate replication failure is often confused owing to a lack of shared meaning when discussing key methodological components of replication research.[53]

Replication failure is not immediately a problem in and of itself. Wolfgang Stroebe and Fritz Strack contended that the *crisis* label should be rejected because replication failure is not inherently negative. They argued that the state of replication failure is overstated because of a misunderstanding of the role of replication. Hence, they rejected the notion that there is a replication crisis.[54] This opinion is less apparent in the discourse than it was earlier, but some dissent persists in the wider discussion on appropriate replication levels. The replication crisis is a label levied against the state of replication in various fields that use statistical analysis to justify research conclusions. Therefore, some of the arguments for and against there being a replication crisis in psychological research are applicable across all research-based disciplines.

Jeffrey Mogil (Canada Research Chair in Genetics of Pain; E. P. Taylor Chair in Pain Studies) insisted that a 50% replication failure rate in research is entirely acceptable.[55] Mogil raised the point that a 100% reproducibility rate (especially in exploratory research) is both unrealistic and undesirable: '[t]here is no way you could expect 100% reproducibility, and if you did, then the studies could not have been very good in the first place'. This view implies a universality of failure rates that is unrealistically oversimplified.

Acceptable replication failure rates depend upon the specific design decisions of individual research projects. There are three categories of replication research methodologies to which arguments about failure rates must be applied separately:

- Direct replications are as identical to the original research as possible.

[53] Schmidt, 'Shall We Really Do It Again?', 90.
[54] Stroebe and Strack, 'The Alleged Crisis'.
[55] Jeffrey Mogil quoted in Philip Hunter, 'The Reproducibility "Crisis"', EMBO Reports 18, no. 9 (2017): 1493. https://doi.org/10.15252/embr.201744876.

- Systematic replications make intentional iterative changes to the original research.
- Conceptual replications explore the same theoretical constructs as the original research using different research designs.[56]

Therefore, conflating the entire replication crisis with instances of direct replication failure illustrates a limitation in the arguments for the declaration of a crisis.

One piece of evidence is insufficient to automatically discredit two (or more) pieces of contradictory evidence, yet it is enough to justify further investigation. Stroebe and Strack argued against the perception that a single contradictory piece of evidence is grounds for discrediting all prior work.[57] However, they failed to acknowledge that it is equally invalid to automatically dismiss a contradictory result. In both cases, further investigation is warranted into the appropriateness of the (research and replication) methodologies and conclusions being reported by the research teams of all projects involved. Perhaps Stroebe and Strack assumed this position was the norm, but they only explicitly argued for further analysis of direct replication failures. They neglected to explicitly apply the same measure to primary research and other replication methodologies.

Individual cases of appropriate or inappropriate replication standards are not universally applicable. Although replication failure should not be immediately dismissed, neither should a single case of replication failure automatically discredit the original research. Stroebe and Strack argued against there being a crisis by focusing on the limitations of inferences that can be drawn from a failed direct replication.[58] They asserted that it is invalid to assume a single replication failure is sufficient evidence to discredit the results of prior research, particularly if a body of replication attempts supports the original claim. Yet, Stroebe and Strack, and—separately—Mogil, demonstrated invalid reasoning when they generalised from failures of direct replication to all replication research. That replication failure can be

[56] I argue in opposition to thinkers such as Brian A. Nosek and Timothy M. Errington, and Edouard Machery who reject the utility of the categorical classification of *conceptual replication*; Brian A. Nosek and Timothy M. Errington, 'What Is Replication?', *PLOS Biology* 18, no. 3 (27 March 2020): e3000691, https://doi.org/10.1371/journal.pbio.3000691; Edouard Machery, 'What Is a Replication?', *Philosophy of Science* 87, no. 4 (October 2020): 545–67, https://doi.org/10.1086/709701.
[57] Stroebe and Strack, 'The Alleged Crisis', 60.
[58] Stroebe and Strack, 60–1.

useful does not discredit the concerns that replication failure is too often the result of poorly conducted and/or communicated research.

I argue that replication failure between projects can often be identified as a result of the misunderstanding and misattribution of replication methodologies due to the lack of a shared RDN. Replication methodologies have specific characteristics with specific inferences that can be drawn from each. Thus, using one type of replication methodology to infer an unsupported conclusion is inappropriate. [59] Rather, appropriate replication failure occurs only when the failure to replicate is the result of factors inherent to the psychological phenomena being explored. Research replication based on insufficient effort (whether voluntarily or due to a lack of sufficient research design transparency by the original researchers) to appropriately conceptualise the necessary methodological framework for an accurate replication project is inappropriate. I argue that inappropriate replication failure can be described as the result of wilful participation in avoidable practices; specific approaches to replication research are useful for identifying QRPs.[60]

2.3. Avoidable Factors Contributing to the Crisis

Several factors have been identified as contributing to the avoidable rates of replication failure observed within published psychological research. Psychological research is not conducted in a bubble of unlimited funding outside the influence of human fallibility. It is currently undeniable that some areas of research would be impossible without large sums of money. Scott Lilienfeld, Sophia Crüwell et al. and others have identified that a bias towards predetermined outcomes has been observed within the literature because, unfortunately, careers are often tied to funding and/or publication.[61] This bias is particularly problematic since at least one solution proposed to correct the replication crisis (meta-analysis) requires that a representative spread of results be produced. However, if results being published are artificially

[59] Further discussion can be found in Section 5.2.
[60] Replications that change minimal design variables between projects are best suited for detecting QRPs; further discussion can be found in Section 5.2.
[61] Scott O. Lilienfeld, 'Psychology's Replication Crisis and the Grant Culture: Righting the Ship'. *Perspectives on Psychological Science* 12, no. 4 (2017): 660. https://doi.org/10.1177/1745691616687745; Crüwell et al., 'Seven Easy Steps', 237-238.

skewed, then the meta-analysis will also be skewed. Lilienfeld suggested that because of these biases, a culture of QRPs in psychology is unsurprising.[62]

Biases have encouraged the circumstances of the replication crisis. Lilienfeld indicated that given that the careers of psychology researchers have been tied to publication goals, it is almost inevitable that issues have arisen within the discipline. Scott Lilienfeld noted how the grant culture in psychological research is biased towards generating positive results. In short, Lilienfeld identified that it is not unexpected that when a researcher's livelihood (in his arguments, grant money) is tied to the generation of positive results, some researchers will do everything in their power to secure positive results. This includes QRPs as well as outright fraud. However, grant money is not the only way in which researchers are biased towards producing confirmatory results.

Similarly, there has been a publication bias towards predetermined results. Regarding original research, negative results were/are rarely published, and in replication research, positive (confirmatory) results were/are rarely published. The former has been justified by the argument that something not occurring is less interesting than something that is occurring;[63] therefore, publications should focus on positive results. Likewise, the latter has been justified by the argument that something already observed is less interesting than something contradictory; therefore, publications should focus on replication failures. While not necessarily easily countered, these arguments fail to address the import of all research results in synthesising a representative picture of a phenomenon (in psychological research as well as other research). Systematic data analysis is not necessarily interesting, but it is nevertheless an important component of scientific knowledge development.[64] This neglect of comprehensive data collection is further exacerbated by tying researchers' job security and success to the number of papers they publish.[65]

In addition, Lilienfeld asserted that when researchers' livelihoods are influenced towards the production of predetermined results, QRPs will inevitably become a significant issue. QRPs are being addressed, and more outlets for the publication of

[62] Lilienfeld, 661–3.
[63] Hunter, 'The Reproducibility "Crisis"', 1495.
[64] Sherri L. Jackson, *Research Methods and Statistics: A Critical Thinking Approach*, 5th ed. (Boston, Massachusetts: Cengage Learning, 2016): 11–2; Schmidt, 'Shall We Really Do It Again?'
[65] Lilienfeld, 'Psychology's Replication Crisis'661–3; Vyse, 'P-Hacker Confessions'.

non-significant results are being provided, but a clearer path forward does not automatically revive confidence in the reliability of existing research. Lilienfeld argued that when results are being externally driven towards predetermined conclusions, it is unsurprising that psychology researchers have adopted questionable practices as the disciplinary norm.[66]

Psychological research has a history of systemic QRPs. It is now commonly acknowledged that in the recent past, psychology researchers have routinely engaged in practices that do not meet current standards. Stuart Vyse is an experimental psychologist who specialises in critical thinking and the nature of pseudoscience. Vyse also confessed to regularly manipulating the statistical analysis of research data to make results appear more meaningful than they were.[67] Vyse reported that this was common practice for a time across all disciplines that use statistical analysis.

QRPs were/are commonly used by psychology researchers; an anonymous survey of over 2,000 psychology researchers reported a high prevalence of QRPs.[68] Of the researchers surveyed, over 90% admitted to using at least one QRP. Although it is not as simple as inferring that, thus, over 90% of psychological research conclusions are the result of QRPs, the scope of the problem casts serious doubt on the entire discipline. It is not unreasonable for confidence to be withheld until the research standards of any individual project have been independently audited. From the survey, the self-admission rate for selected QRPs were:[69]

- neglecting to report all dependent measures (63%);
- resuming data collection after analysing the significance of existing data (56%);
- excluding data after analysing the impact of doing so on the results (38%);
- neglecting to report all of a research project's conditions (28%); and
- stopping data collection after analysing the significance of preliminary data (16%).

If this is an indication of the prevalence of such practices in the broader research, then the failure of most replication attempts is almost inevitable. Unfortunately, the study

[66] Lilienfeld.
[67] Vyse, 'P-Hacker Confessions'.
[68] John, Loewenstein and Prelec, 'The Prevalence of Questionable Research Practices', 524.
[69] John, Loewenstein and Prelec, 525.

analysed that most of the QRPs identified likely occur at rates higher than the survey demonstrated. This view is supported by how these practices relate to p-hacking as well as by the prevalence of p-hacking within the discipline.

In statistical analysis, a common probability threshold (.05) alone has been upheld as the benchmark for determining the value of a research project. Results below .05 were more likely to be published and receive public attention. Results above the threshold were more likely to be ignored and remain unpublished. Thus, a culture arose around manipulating data in such a way as to guarantee a result of .05 or below (p-hacking).

Vyse admitted to p-hacking his own research via various QRPs, as well as aiding colleagues to do the same with their own.[70] Vyse has now recognised that p-hacking is harmful to the credibility of psychological research and has regretted the part he played in the pervasiveness of this practice.[71] However, he has expressed optimism that the shift towards more transparent accountability within research design is the correct move for the discipline.[72] As Vyse identified, p-hacking has contributed to the widespread failure of replication research attempts.[73] With the prevalence of p-hacking now coming to light, confidence in many previously accepted phenomena is being (justifiably) questioned, but accurately describing the problem is the first step to correcting it.

2.4. Individual Replication Failure as a Single Data Point

It is often overlooked (or assumed) to the detriment of the broader discourse that individual instances of replication failure are only a single data point in a larger conversation. A single negative research result is insufficient to disprove a large quantity of positive research results, and vice versa. As mentioned previously, priming research has been a focal target (possibly unfairly) in critiquing research standards and replication in psychological research. Wolfgang Stroebe and Fritz Strack used a particular case study from priming research to explore and contextualise

[70] Vyse, 'P-Hacker Confessions'.
[71] Vyse.
[72] Vyse.
[73] Vyse.

the appropriateness of individual replication failure, as being a single data point in a larger conversation.[74]

The concept of priming suggests that activating a social concept in an individual's memory can lead to unconscious changes in their behaviour.[75] John A. Bargh, Mark Chen and Lara Burrows performed one notable study on the topic, reporting that participants who were unconsciously primed with words related to old age consistently walked slower when later leaving a room.[76] The study concluded that the observed behaviour was the result of the words used having an unconscious psychological impact on the participant's behaviour. However, a replication of the original research project failed to observe identical outcomes.

Doyen et al.'s (systematic) replication study made some key adjustments to the original study's methodology but sought to replicate the original experimental conditions as closely as possible.[77] The replication used a less subjective measure to record the walking speed of participants, as well as initially kept the nature of the study being conducted from the experimenters who interacted with the participants. The results of the first part of this replication attempt failed to support the conclusions of the prior research; no difference was measured between the experimental participants and the control group.

In offering a critique of the failed replication, rather than addressing the change in methodological precision that may have contributed to the different results, Stroebe and Strack focused on the differences between participants.[78] In their criticism of this replication attempt, Stroebe and Strack identified the sociodemographic differences between the source and the replication study as potentially being responsible for the difference in results. The basis for this critique appears reasonable since they defended it with reference to supporting observations within the discourse. However, Stroebe and Strack neglected to address why this effect would not also apply to the second part of the experiment where the researchers did observe the priming effect. If the sociodemographic differences between the original and the replication studies

[74] Stroebe and Strack, 'The Alleged Crisis', 66.
[75] Colman, *A Dictionary of Psychology*, s.v. 'priming'.
[76] John A. Bargh, Mark Chen and Lara Burrows, 'Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action', *Journal of Personality and Social Psychology* 71, no. 2 (1996). https://doi.org/10.1037/0022-3514.71.2.230.
[77] Stéphane Doyen et al., 'Behavioral Priming: It's All in the Mind, but Whose Mind?' *PloS One* 7, no. 1 (2012). https://doi.org/10.1371/journal.pone.0029081.
[78] Stroebe and Strack, 'The Alleged Crisis', 67–8.

were the only cause of failure, then it would be reasonable to assume the second attempt would fail for similar reasons.

In the second part of the Doyen et al. (systematic) replication attempt, the original conclusions were found to be valid if the experimenters' expectations were accounted for as an additional confounding variable.[79] When the experimenter expected a participant to walk slower because of being exposed to words related to old age, the participant walked slower because of being exposed to these words. They concluded that the results of their second replication suggested the effect of priming on participants was not only a result of the words used, but also the unconscious effect of experimenters' bias influencing their behaviour with the participant. However, Stroebe and Strack countered this conclusion stating that as the as the experimenters in the Bargh, Chen and Burrows research were unaware of the experimental conditions, expectations could not have played a part in the original experiment.[80]

Yet, the controversy around priming does not end there. Stroebe and Strack offered critiques of additional replication attempts. They considered alternative interpretations of the reasons for replication failures as well as offered alternative methodologies they judged relevant to exploring the concept further. However, while they focused on salvaging the theoretical work around priming, they somewhat obfuscated their original point. Namely, individual replication projects need to be regarded as a single data point in the broader conversation.

Regarding the replication crisis in psychological research, the conclusions of the studies by Bargh, Chen and Burrows, and by Doyen et al., are less salient than the ways those same conclusions are then interpreted within the broader priming discourse.[81] Stroebe and Strack demonstrated the necessity of further analysis and discussion rather than reflexive defence or denial of the initial conclusions.[82] As illustrated by the case from priming research, on the surface, an individual replication failure can be interpreted as denying the conclusions of the original research, but a more complex and nuanced picture may evolve through further exploration.

[79] Doyen et al., 'Behavioral Priming', 3–6.
[80] Stroebe and Strack, 'The Alleged Crisis', 67–8
[81] Bargh, Chen and Burrows, 'Automaticity of Social Behavior'; Doyen et al.
[82] Stroebe and Strack, 'The Alleged Crisis'.

In conclusion, instances of replication failure are only validly understood when regarded as single data points that need to be explored in the context of the research discourse. It is equally invalid to interpret individual cases of replication failure as:

- sufficient evidence to reject all prior research on the same topic; and
- irrelevant against a body of successful replication.

Single cases of replication failure should be analysed from as unbiased a position as possible to consider the strengths and weaknesses of the project in and of itself. Then, the research should be evaluated against the existing discourse. Again, individual cases of replication failure are single data points in a broader conversation.

2.5. Arguments Against the *Crisis* Label

Currently, such arguments are apparently less popular, but some still imply that the use of the term *crisis* is inaccurate. Unlike how Pashler and Wagenmakers chose to argue a reframing of the crisis (not a relabelling), Pashler and Harris directly addressed common arguments from within the discourse that reject any form of crisis.[83] Despite the term *replication crisis* being used to refer to the multiplicity of problems that result in the failure to replicate psychological research, some object to the use of the specific term *crisis*.[84] One common argument calls for patience, suggesting that the term crisis implies an unnecessary call to action. Another rejects the problem altogether, suggesting instead that the observed percentage of failure is acceptable.[85]

2.5.1. Argument for Patience

There is an argument for patience in the self-correcting nature of science as being the solution to any perceived crisis in psychological research. Proponents of this argument insist that the 'alleged crisis' is an unnecessarily dramatic call to action.[86] Pashler and Harris interpreted this type of argument as implying that unrefined research conclusions will inevitably be retested and rejected over time, or that the discourse will follow the conclusions with the most evidence, and thus, additional

[83] Pashler and Wagenmakers, 'Crisis of Confidence?'; Pashler and Harris, 'Is the Replicability Crisis Overblown?'
[84] Pashler and Harris; Stroebe and Strack, 'The Alleged Crisis'.
[85] Pashler and Wagenmakers, 'Crisis of Confidence?'
[86] Stroebe and Strack, 'The Alleged Crisis'.

replication is unnecessary. However, Pashler and Harris (among others[87]) counter-argued that there is nothing inevitable about replication research, and that there is no guarantee that the shifting focus of academic research will follow the more robust theory. In addition, another point against patience (not addressed by Pashler and Harris) is the potential harm of a delay in the rigorous validation of theories to the most vulnerable members of society. For example, the practice of using lobotomies to treat certain psychological disorders was widely accepted for decades, acceptance permeating beyond academic discourse.[88] Yet, the research behind it has been harshly criticised and ultimately discredited.[89] It may be accurate to say that scientific knowledge did inevitably self-correct, but there was a cost to those outside of academia that needs to be factored into the discussion.

A common argument against there being a replication crisis urges for more patience about the inevitability of research: 'Science is self-correcting but slow—although some erroneous results may be published, eventually these will be discarded. Current discussions of a replicability crisis reflect an unreasonable impatience'.[90] The argument is that claiming there is a crisis distracts from the inevitable refinement of ideas through accumulated observations. The idea of the replication crisis being a part of the natural progression of science is not exclusive to the realms of psychology. Erkki Ruoslahti (a cancer researcher) argued that '[f]indings that are not correct will disappear because others can't reproduce them or publish divergent results, after an adequate try and hopefully also explaining why the results are different'.[91] An important distinction to note is that commentators such as Ruoslahti did not argue that replication failures are not occurring. More so, they suggested that individual failures to replicate are being overblown, and that a broader, more patient view needs to be taken as regards the ongoing development of scientific knowledge. As discussed, a single replication failure alone is insufficient to automatically discredit a theory. However, there are several key problems with this position in relation to psychological research.

---

[87] Simine Vazire and Alex O Holcombe, 'Where Are the Self-Correcting Mechanisms in Science?', *Review of General Psychology*, 2020.
[88] Rachel Cooper, 'On Deciding to Have a Lobotomy: Either Lobotomies Were Justified or Decisions under Risk Should Not Always Seek to Maximise Expected Utility', *Medicine, Health Care and Philosophy* 17, no. 1 (2014). https://doi.org/10.1007/s11019-013-9519-8.
[89] Cooper.
[90] Pashler and Harris, 'Is the Replicability Crisis Overblown?', 534.
[91] Erkki Ruoslahti quoted in Hunter, 'The Reproducibility "Crisis"', 1494.

It is unreasonable to assume inaccurate findings will necessarily be retested and/or contradicted. One interpretation of the argument for patience suggests that inaccurate findings will necessarily be retested over time and, thus, contradicted. Pashler and Harris argued that (at least in psychological research) this interpretation was unrealistically optimistic. In 2012, Pashler and Harris performed a crude assessment of the state of direct replications in psychological research. They concluded that because most of the replication research they identified was generally performed within 4 years of the original, there was no evidence to support the view about the correction of invalid results in the long term. However, assuming the results of Pashler and Harris's assessment were accurate at the time, they are no longer entirely valid. Thanks to the attention the replication crisis has directed onto psychological research, initiatives such as the OSC have implemented large-scale replication projects with the explicit intent of exploring or validating prior research. Yet, it cannot be known whether the observed increase in replication rates would have necessarily occurred without the topic of the replication crisis permeating throughout the discourse. Similarly, that rates of replication are increasing is tangential to the causes of replication failure because the rate of failure appears consistent.

Another interpretation of the argument for patience suggests that even if replication of specific theories or conclusions is not attempted, inaccurate theories will inevitably fall out of vogue. Proponents of this position assert that incorrect conclusions will fall out of favour and be superseded by more robustly supported claims.[92] Thus, the argument infers that replication research is not entirely necessary – science will inevitably correct over time as more observations are accumulated. Pashler and Harris suggested that this is not necessarily true due to the 'faddish' nature of academic research.[93] They contended that entire avenues of research may simply lie dormant due to the zeitgeist being oriented in more popular directions. Thus, Pashler and Harris observed no support for the inevitability of this confidence in scientific progress.

Further, both these arguments for patience neglect to account for the impact of research on sociopolitical outcomes. As Pashler and Harris argued, a lack of academic focus on a theory may also simply be the result of its acceptance, and not that it is

[92] Pashler and Harris, 'Is the Replicability Crisis Overblown?', 534–5.
[93] Pashler and Harris, 534–5.

considered outdated. However, the assertion that a theory is either accepted or ignored neglects the impact of inaccurate theories outside of academic discourse.

2.5.2. Argument for Acceptable Failure Rates

Another argument against the use of the term *crisis* derives from examining the probability that some false positives will always be published. Even if a research project is conducted without any methodological issues, there is always a chance that the statistical analysis will result in a false positive. In statistics, the likelihood of a false positive is what the probability value (*p*-value) indicates. The *p*-value is an abstract representation of the probability that the results of a research project are representative of a repeatable phenomenon, rather than the result of chance. A *p*-value of .05 is equivalent to saying that there is a 5% chance that the statistical analysis was the result of chance alone.

Crisis deniers, such as Wolfgang Stroebe and Fritz Strack, and Philip Hunter, have argued that false positives are observed in published research at a rate that is tolerable; therefore, there is no crisis.[94] Given that the probability threshold of a false positive has often been set to 5%, it has been argued that 5% of false positives being published is acceptable.[95] However, it has been identified that even without replication failure, the rate of published inaccurate statistics exceeds 5% (again, the OSC reported a replication failure rate of 64%).[96] Although false positives are statistically (not necessarily practically) inevitable, the issues around intentional fraud, misreporting, QRPs, and biases all combine to raise serious questions about the rate of publication of false positives. Yet, building upon this argument, some commentators have rejected the assertion that only 5% of false positives being published is acceptable. Instead, some have contended that in certain areas of research, a 50% rate of false positives is an acceptable norm.[97] However, even if it may be appropriate/inappropriate for certain methodological approaches, it does not mean that that it will be appropriate/inappropriate for all approaches.

The acceptable replication failure rates differ depending on the method of replication research used. Sometimes, the complexities of different methodological approaches to

[94] Stroebe and Strack, 'The Alleged Crisis'; Hunter, 'The Reproducibility "Crisis"'.
[95] Pashler and Harris, 'Is the Replicability Crisis Overblown?', 531–3.
[96] Open Science Collaboration, 'Estimating Reproducibility'.
[97] Hunter, 'The Reproducibility "Crisis"', 1493.

research are neglected in the discourse about replication research in psychology. Stroebe and Strack mentioned the different degrees of appropriateness of replication failures for direct replications in both applied (intervention-based) and basic (theory-based) research.[98] They also argued that direct replications are more appropriate in applied research, paying attention to the development of specific treatments and/or interventions. Applied research mistakes can have substantial (or fatal) consequences and, thus, researchers must take all possible care.[99]

2.6. Chapter Summary

The primary conclusions of this chapter are as follows:

- It is useful to interpret the replication crisis in psychological research as a crisis of confidence (in research conclusions).
- Transparency in research communication is a necessary step in restoring confidence in replication research conclusions.

In this chapter, I have unpacked the topic of the replication crisis in psychological research to demonstrate the reasons it is important to conceptualise the crisis as one of confidence in both psychological research practice and the conclusions that are then drawn. The negative impact of a lack of a shared RDN is observed in the disregard for the appropriateness of a given inference being made in relation to specific replication failure. Ill-defined terminology sometimes results in a conclusion being deduced from replication failure that is not appropriate for the type of replication that was performed. Replication failure alone cannot be appropriately interpreted without additional information regarding the circumstances that lead to this failure. However, because of the ill-defined RDN, this interpretation of replication failure is often overlooked; specific instances of replication failure are misinterpreted to justify agnotological thinking in relation to all psychological knowledge. In later chapters, I begin to synthesise an RDN designed to aid in the appropriate interpretation of replication failure. However, such appropriate interpretation is a symptom of an oft-identified issue in research communication, namely, that without transparency in research design and communication it is

[98] Stroebe and Strack, 'The Alleged Crisis', 60.
[99] However, Stroebe and Strack argued that basic research is limited to exploratory research. Therefore, they argued that it is both invalid and detrimental to assume successful direct replications provide meaningful insight into the validity of theories. While not entirely uncontentious, this argument is still supported by a faction within the discourse; Stroebe and Strack.

inevitable that the large percentage of avoidable replication failures currently observed will continue to contribute to the replication crisis.

Chapter 3. Transparency in Psychological Research as an Epistemic Virtue

In this chapter, I assume the following primary conclusions from the previous chapters:

- All researchers and their audiences do not necessarily use identical definitions of concepts related to psychological research (Chapter 1).
- It is useful to unpack pivotal concepts in psychological research to facilitate effective communication between researchers and their audience (Chapter 1).
- It is useful to interpret the replication crisis in psychological research as a crisis of confidence in research conclusions (Chapter 2).
- Transparency in research communication is a necessary step in restoring confidence in replication research conclusions (Chapter 2).

I contend that too much transparency in research design and communication can be as undesirable as too little transparency. I apply an interpretation of virtue epistemology to the concept of transparency in scientific communication, highlighting that the transparency level that is contextually useful for combating the replication crisis is distinct from the level that is contextually useful for communicating research to non-experts. However, given that I have aligned with those arguing that the replication crisis in psychological research is better understood as a crisis of confidence in research design and conclusions, I consider that communicating research appropriately to non-experts is as important as communicating it to experts. Thus, by positioning *appropriate transparency* (contextually desirable state) between *overstated transparency* (contextually undesirable state) and *understated transparency* (contextually undesirable state), I argue that researchers currently lack the tools (a shared RDN) necessary to judge the level of transparency appropriate for their target audience. I contend that appropriately transparent communication between researchers and their different audiences is an important step towards restoring confidence in the results of psychological research.

In this chapter, I build upon arguments that transparency in research communication is an ill-defined concept that requires further development before the concept can usefully be leveraged by researchers. The common-sense definition of transparency is that it comprises concepts such as visibility, communication and accountability, and facilitation of research evaluation; however, the quantification of these concepts is

rarely addressed. Moreover, I am unaware of any research that uses epistemic virtue theory as a framework to consider transparency in relation to research communication. I argue the utility of leveraging an epistemic virtue framework as a tool to endorse arguments for and against transparency in research communication. In my research, I was unable to identify any argument for or against transparency that provided researchers with a guide towards addressing the variable needs of the research communication audience; rather, what I did find implied a single level of transparency as the goal for all research communications.[100]Instead, in this chapter, I define appropriate transparency as an epistemic virtue that lies on a spectrum between *understated transparency* and *overstated transparency*. Virtue epistemology (epistemic virtue) being a theory (or theories) of knowledge that leverages the concept of *virtue*—as developed through *virtue ethics*—to consider the characteristics useful for refining knowledge.[101] Despite the name, virtue theory[102] has as strong a focus identifying the right thing to do in each context (virtue) as it does the wrong things to do (vices). It is useful to understand that the virtue exists between two vices: one of deficiency (of the virtue) and one of excess (of the virtue). What I refer to as the *vice/virtue/vice trichotomy*.[103] Thus, I defend the classification of *understated transparency* and *overstated transparency* as an epistemic vice of deficiency and of excess, respectively.[104]

Epistemic virtue theory provides a useful framework for judging the concept of transparency as a desirable state found between opposing undesirable states. Thus, appropriate transparency in research communication is a desirable state between two (related) undesirable states; transparency in research is an oft-discussed topic in the discourse around the replication crisis. However, the various ways in which

[100] For example: Elliott, 'A Taxonomy of Transparency'; Figueiredo Filho et al., 'Seven Reasons Why'; John, 'Epistemic Trust'.
[101] Craig A Boyd and Kevin Timpe, *The Virtues: A Very Short Introduction*, Very Short Introductions (Oxford: Oxford University Press, 2021); John Turri, Mark Alfano and John Greco, 'Virtue Epistemology', in *The Stanford Encyclopedia of Philosophy*, Fall 2019 ed., ed. Edward N. Zalta (Stanford: Metaphysics Research Lab, Stanford University, 2019), https://plato.stanford.edu/archives/fall2019/entries/epistemology-virtue/; Richard Kraut, 'Aristotle's Ethics', in *The Stanford Encyclopedia of Philosophy*, Summer 2018 ed., ed. Edward N. Zalta (Stanford: Metaphysics Research Lab, Stanford University, 2018), https://plato.stanford.edu/archives/sum2018/entries/aristotle-ethics/; Simon Blackburn, *A Dictionary of Philosophy*, 3rd ed. (2016), s.v. 'virtue epistemology'.
[102] Inclusive of, but not limited to, virtue epistemology and virtue ethics; Boyd and Timpe, *The Virtues*.
[103] That I use this term to highlight the importance of both vices in relation to the associated virtue does not feel unique and I suspect I am not the first to use this language. However, I was unable to locate a reference using the term *vice/virtue/vice trichotomy*.
[104] Further discussion can be found in Sections 3.2 and 3.3.

transparency is discussed in the literature often imply different meaning attributions to the term. Thus, transparency can be understood as an example of a contentious term fallaciously discussed as if possessing an immutably concrete definition.[105] Some commentators have argued that increased transparency will resolve many of the problems identified by the replication crisis, but they do not provide specific guidance on ways through which transparency should be actualised. Alternatively, some commentators have rejected transparency as a solution, arguing that the more the research information available, the more likely that it will be misinterpreted, but they neglect to account for the necessity of shared information in the replication process.In evaluating various arguments for and against transparency in research, it is useful to conceptualise transparency as a desirable state that can be judged (contextually) as existing along a spectrum between two undesirable states. The validity of arguments both for and against being transparent makes it apparent that there is less utility in arguing the term from the position of a false dichotomy; researchers lack a universally appropriate way to apply the concept of transparency to their research design and/or communication. Pro-transparency arguments provide justification for transparency in research specifics appropriate to the context of the situation; conversely, in specific circumstances, transparency can be judged as being in a state of excess. Anti-transparency arguments provide justification for a lack of transparency in research specifics appropriate to the context of the situation; conversely, in specific circumstances, transparency can be judged as being deficient. On the surface, it appears that each side of this argument hinges upon the definition of transparency; whereas by exploring the concept of appropriate transparency as being contextually bound, the dispute dissolves as each argument is validated when considered within specific contexts.

However, although I contend that transparency in research communication is one step towards correcting the replication crisis, I maintain that transparency should be informed by the research objectives. Trying to be transparent in one's research without identifying what needs to be transparent would be aimless. In later chapters, I propose that a shared RDN is necessary to identify appropriate transparency. To be appropriately transparent, one must know what details to judge as necessary and/or

[105] As discussed in Section 1.3.

unnecessary. Without a shared nomenclature, appropriate transparency cannot be easily identified.

3.1. Ill-Defined Transparency

Transparency in research is commonly promoted as one way to correct for the replication crisis in research;[106] yet, transparency is ill-defined as a research design concept.[107] Little consensus is found about details related to transparency in research: Stephen John argued against transparency in research communication by equating the concept with uncritical openness, whereas Kevin C. Elliot promoted transparency as a taxonomic term with multiple dimensions and variations.[108] In these examples (as well as in the broader discourse), transparency is discussed in relation to both research design, and (sometimes implied, and at other times explicitly discussed), by extension, research communication. Thus, although not always outlined as such (from the context of its use), transparency in research design is suggested as a component of communication of said research. Therefore, in the discourse, arguments for and against transparency in research can be interpreted as synonymous with arguments for transparent research communications, and vice versa.

However, this ambiguity surrounding transparency recalls the already discussed issue surrounding a lack of shared nomenclature in research psychology.[109] Elliot summarised this ambiguity by both acknowledging it as a concept important to all researchers, while simultaneously addressing the complexity of defining the concept. Elliot identified that whereas researchers and philosophers are both quick to promote transparency as useful and important, its definitions often widely vary.

I argue that both sides of this dilemma are correct, but only under specific contexts. It is possible to assume that the reasons given for and against transparent research communication imply the contextual appropriateness of each position. Rather than assuming a singular answer for a universally correct level of transparency in research, I argue two related points. First, the appropriate level of transparency should be evaluated to suit the intended audience of a given research communication. Second, an exhaustive level of relevant detail concerning the research process should be made

[106] Figueiredo Filho et al., 'Seven Reasons Why'; Crüwell et al., 'Seven Easy Steps'; Open Science Collaboration, 'Estimating Reproducibility'.
[107] Elliott, 'A Taxonomy of Transparency'.
[108] John, 'Epistemic Trust'; Elliott, 'A Taxonomy of Transparency'.
[109] Schmidt, 'Shall We Really Do It Again?'

readily available for those who desire it. Moreover, similarly to Elliot, I argue that uncritical assumption of shared meaning when discussing transparency would be unhelpfully ironic.

Dalson Figueiredo Filho, Rodrigo Lins, Amanda Domingas, Nicole Janz and Lucas Silva acknowledged a lack of consensus in research discourse regarding the expected characteristics of transparency across and within multiple disciplines.[110] In an effort to correct for this lack of specificity, Figueiredo Filho et al. argued that transparency can be defined as 'the full disclosure of the research design, which includes the methods used to collect and analyze data, the public availability of both raw and manipulated data, in addition to the computational scripts employed along the way'.[111] However, Figueiredo Filho et al.'s arguments for transparent research practices did not address the potential broader social implications of 'full disclosure' on the accessibility of research for non-experts.[112]

Stephen John, who opposed transparency in research communication, argued that transparency in research does more harm than good.[113] John built his argument against transparency in research by focusing on how climate sceptics use communications between climate researchers to promote an anti-research rhetoric. He argued that (as non-experts are untrained in parsing the complexities of the research process) transparency concerning raw research data and research processes introduces unnecessary confusion into the dissemination of research conclusions to non-experts. However, even though John focused on the dangers of exhaustive transparency for non-experts, he makes little reference to the necessity of transparency between experts.

In this regard, transparency in scientific communication is undeniably beneficial between experts, but too much transparency with non-experts may obfuscate the appropriateness of research conclusions. Figueiredo Filho et al. argued for a complete, unfiltered level of transparency.[114] They presented a reasoned argument that insufficient transparency regarding the communication of the research process is less helpful to the ongoing research in a field. Conversely, John presented a reasoned

[110] Figueiredo Filho et al., 'Seven Reasons Why'.
[111] Figueiredo Filho et al., 3–4.
[112] Figueiredo Filho et al., 3–4.
[113] John, 'Epistemic Trust'.
[114] Figueiredo Filho et al., 'Seven Reasons Why'.

argument that transparent communication of the research process is less helpful for the dissemination of research conclusions to non-experts.

It is worthwhile to explain the importance of these points still further. Uncritical transparency in research communication can create confusion for non-expert readers. Raw, unfiltered research data has been shown to occasionally encourage some non-experts to mistrust experts. As mentioned, John argued against transparency in research by discussing the discourse around leaked climate research in the 2009 controversy colloquially referred to as Climategate. [115] John (informed by a specific model of learning that describes how non-experts learn from experts) contended that allowing non-experts unfiltered access to raw research data is demonstrably problematic. He described how individuals without formal training can misinterpret raw research data; he highlighted that the literature has shown non-experts to be unaware of acceptable trends in data analysis, and therefore, they misunderstand the appropriateness of research conclusions. [116] John focused the discussion on climate change and passionately defended the position that experts should be allowed free rein to perform their research without non-expert oversight. John did acknowledge that it may be inappropriate for research to be shrouded in secrecy, but he nevertheless encouraged obfuscation of both raw research data and design justifications from non-experts.

The levels of misinterpretation by experts give credibility to John's argument that it is unreasonable to expect non-experts to be able to accurately interpret raw research data and design decisions. A problematically high rate of mistakes has been identified in research communication written or edited or peer-reviewed by experts.[117] Therefore, it is reasonable to conclude that non-experts will be at least as susceptible to misinterpretation as experts.

Unfortunately, John conflated all research communications as a singular homogeneous concept. He did not argue that raw research data and design decisions should not be shared between experts. Instead, John discussed research communication as almost exclusively something provided by experts to non-experts.

[115] John, 'Epistemic Trust', 81–2.
[116] Notably, some of John's arguments about acceptable trends strayed towards a defence of various QRPs, but his arguments did not hinge upon these defences; therefore, I do not engage with them further (John.).
[117] Ioannidis, 'Why Most Findings Are False'; Bakker and Wicherts, 'The (Mis) reporting of Statistical Results'.

John made no mention of situations wherein replication researchers are required to design their projects with only published research reports for guidance.[118] Therefore, any inference that research communication is exclusively the practice of experts summarising conclusions for non-experts is obviously inaccurate.

To account for arguments both for and against transparency, I propose that it is useful to understand it as an epistemic virtue that orients researchers to consider how best to apply the concept to their current situation. I agree with arguments that propose transparency is necessary for encouraging more robust replication attempts, but I disagree that it is useful to communicate exhaustive detail to non-experts. I agree with arguments that propose ill-considered transparency can damage the expert – non-expert relationship, but I disagree with the views that transparency should be rejected in favour of closely guarded secrets. Overall, I propose that full transparency in research design is necessary, but also that practical wisdom should be applied when considering the appropriate level of transparency in research communication.

Replication researchers and non-expert readers are two distinct categories of groups, each requiring different levels of transparency. The exhaustive level of detail ideal for a direct (and/or systematic) replication attempt is sub-optimal for inclusion in most research communications. In line with John, it is reasonable to argue that complete transparency in all research communication would be detrimental to some audiences. Yet, although John argued that non-experts are susceptible to misinterpreting raw research data, it is unclear why topical experts without research expertise would not also be susceptible to misinterpreting raw data. An individual may be an expert on a given topic without possessing the skills necessary to accurately interpret statistical data. Therefore, it is reasonable to conclude that ill-considered transparency in some research communications may engender more confusion than necessary among experts and non-experts. However, again this does little to account for the level of transparency optimal for replication researchers, and to this we will now turn.

3.2. Epistemic Virtue in Research Communication

An epistemic virtue is a desirable state that exists between two undesirable states (epistemic vices). At one end of the spectrum, there is a vice of deficiency; that is, depending upon the specifics of a given scenario, there is a state wherein too little of

[118] Open Science Collaboration, 'Estimating Reproducibility'.

the quality is judged as undesirable. At the other end of the spectrum, there is a vice of excess; that is, depending upon the specifics of a given scenario, there is a state wherein too much of the quality is judged as undesirable. The desirable state, the virtuous state, is judged upon the specifics of a given scenario as existing somewhere between the two extremes. However, it is important to note that whereas the virtuous state is neither universal nor static, it is also not individually relative.[119] These terms and points will now be explored.

An epistemic virtue is an orienting principle that is identified due to the context within which it is sought. To reiterate, I describe the virtue as an orienting principle in that it is not a static rule that identifies a fixed position equally between both vices. Rather, the virtuous state is contextually evaluated, positioned on a spectrum between the two vices, as informed by any relevant variables that effect the situation. Depending upon the specifics of the circumstances, the correct positioning of the virtue may be found in perfect balance between the vices, or the virtue may be judged as positioned more towards one vice than the other. However, this is not the same as claiming that epistemic virtues are relative; rather, all individuals with the same skills and abilities, in the same circumstances, will identify the epistemic virtue at the same point between the opposed epistemic vices.[120]

In the related field of virtue ethics, (moral) courage is often given as a virtue that exists between the vices of cowardice (deficiency) and of rashness (excess).[121] Consider then how the courageous action in a circumstance may contain acts that appear more in line with one of the vices. For example, if Individual A were present while Individual B was drowning, the courageous action would depend upon the broader context of the situation. If Individual A were a professional lifeguard witnessing Individual B drowning in a public pool, the virtuous action would be for Individual A to attempt to rescue Individual B. The risk to Individual A is not zero, but it is not so high as to warrant inaction. Individual A has the skills to act with low risk to themselves; therefore, the courageous thing would be for Individual A to act. Yet, if Individual A were a professional lifeguard witnessing Individual B drowning in a public pool in which a live electrical wire had been submerged, it may be judged as rash for Individual A to attempt a rescue before the electrical wire was dealt with.

[119] Further discussion can be found in Section 3.3.
[120] Boyd and Timpe, *The Virtues*; Kraut, 'Aristotle's Ethics'.
[121] Boyd and Timpe, *The Virtues*; Kraut, 'Aristotle's Ethics'.

Likewise, if instead of a live wire in the pool, Individual A had two broken arms, regardless of their training, attempting a rescue would also be judged as rash. As with moral courage, intellectual courage is found in the details of the situation, not via a predetermined set of criteria.

Similar consideration can be applied to intellectual courage as an epistemic virtue. According to John Turri, Mark Alfano and John Greco, intellectual courage is defined as an epistemic virtue akin to the ethical virtue of moral courage.[122] An individual publicly acknowledging what they believe to be the truth irrespective of external pressures to remain silent is given as one example of intellectual courage. However, intellectual courage may also orient an individual to remain silent if the negative outcome of publicly acknowledging the truth were judged as being greater than the positive.

Further, it needs to be noted that an epistemic virtue is not a single, simple, universal rule that can be applied to all situations. The epistemically virtuous state must be reflected upon, and calculated in relation to the specific context, lest the outcome be incorrectly aligned too closely to either of the epistemic vices. With this definition in mind, I now position the two extremes of transparency as undesirable states, justifying the positioning of transparency as a desirable middle ground.

3.3. Vices of Transparency

The level of transparency that will be a useful tool for combating the replication crisis in psychology is likely to be judged as skewed more towards the side of excess than towards deficiency. However, before the appropriate level of transparency can be judged in any context, a working definition of the spectrum of transparency in scientific communication must first be established.

Thus, it is relevant to consider that appropriate transparency in psychological research communication is not a binary concept. Some arguments that promote transparency in research communication position the concept in opposition to (what I define as) understated transparency (understatement/understated).[123] That is, transparency is argued as one solution for a perceived lack of communication regarding relevant

[122] Turri, Alfano, and Greco, 'Virtue Epistemology'.
[123] Further discussion can be found in Section 3.4.1.

research details that negatively affects the replication process.[124] However, some arguments against transparency in research communication propose that the transparency of details is detrimental to the expert – non-expert relationship, leading to (what I define as) overstated transparency (overstatement/overstated).[125] That is, too much detail can be as harmful to scientific communication as too little detail.[126] Considering understatement and overstatement, I instead propose to situate the epistemic virtue of transparency in psychological research communication in the context of the communication.

Although I argue that transparency should be positioned between the extremes of understatement and overstatement, I do not intend to suggest that it is a static concept with universally applicable boundaries. Instead, I define transparency as a useful orientation found between understatement (too little transparency) and overstatement (too much transparency). However, I also argue that transparency is not a static, universally applicable concept. The correct level of transparency should be tailored towards the specifics of the circumstances. In different circumstances, appropriate transparency would differ. In identical circumstances, appropriate transparency would appear identical.

Given that research communication is not exclusively targeted at expert readers, the level of research specifics communicated can be overstated. In his argument against transparency, John framed the issue as being one of how best to encourage non-experts to learn from experts.[127] Some may argue that the target audience for a research report is not necessarily a non-expert, but there are two key problems with this viewpoint:

- Regardless of the intended audience, research reports are often publicly accessible.
- Even experts on a topic are not necessarily experts on the specifics of the research methodologies.

If psychological research communications are only intended for other experts within both the topic of the research and the specifics of the given research paradigm, then

[124] Ben Almassi, 'Climate Change, Epistemic Trust, and Expert Trustworthiness', *Ethics and the Environment* 17, no. 2 (2012). https://doi.org/10.2979/ethicsenviro.17.2.29.
[125] John, 'Epistemic Trust', 80–2.
[126] Further discussion can be found in Section 3.4.1.
[127] John, 80–2.

the audience becomes untenably narrow. Yet, regardless of who is assumed to be the target audience, it is relevant to consider that the research communication will be available to a wide spectrum of readers with varying degrees of technical understanding. Therefore, John's warning against overstating unnecessary detail in general research communication is warranted.

Since research communication is not exclusively targeted at non-expert readers, the level of research specifics communicated can be understated. As discussed, one argument against transparency in research communication proposed that too much detail in research communication can lead to inaccurate interpretations by non-experts. John used the controversy about the misinterpretation of raw climate science data by non-expert readers to argue against the appropriateness of complete transparency in research communication.[128] Although he used the example of climate science, John proposed his argument to be applicable to all forms of research communication.

For example, consider two independent psychology researchers with identical training preparing unrelated reports for the same board of directors. The level of transparency in research communication would be contextual to the experience of the researcher, as well as their understanding of the experience level of the board members. It is conceivable that the level of transparency appropriate for a board of psychology researchers would differ from that appropriate for a board of ethicists. If both psychology researchers are successful in finding the correct level of transparency for the context, then it is assumed that the level of raw data and design justifications between the two reports would be identical. If too much detail were included in one of the reports, then the level of transparency would be in excess (overstated). If insufficient detail were included in one of the reports, then the level of transparency would be in deficiency (understated). However, although potentially less of a problem, overstatement could still be an issue for research communication intended for replication researchers.

Various categories of research design variables will be discussed in Chapter 5. However, (as will be discussed in Section 5.3) it is important that researchers demonstrate practical judgement when determining the salient design variables to communicate. Depending upon the research context, a detail such as the colour of the

[128] John, 81–2.

walls in an experimental laboratory may or may not be relevant. If judged (through explicit consideration as informed by the topical discourse) as irrelevant to the context of the given research project, the inclusion of this information may distract from the relevant details and would therefore also be an example of overstated transparency.

3.4. Contextual Transparency

Transparency is a necessary aid for improved replication research practices. Bias-driven results and QRPs are two of the categorical factors contributing to the replication crisis in psychological research; Stefan Schmidt and Rachel Rosen have independently argued that increasing the levels of transparency in research practices is a solution to both concerns.[129] Both argued that transparency can serve as a protective measure against conscious and unconscious biases in research evaluation, as well as a protective measure against QRPs in research analysis. However, as Kevin C. Elliot identified, consensus about ways to define as well as apply transparency are lacking.[130]

Many of the arguments for increased levels of transparency in research design decisions promote the need to include increased levels of detail in research communication. Yet, how research is communicated plays a large role in the prevalence of replication failure in psychological research. Replication attempts designed exclusively from details provided in a research report require design assumptions that may or may not be representative of the original research.[131] However, two of the three methodological approaches to replication research (direct and systematic) are best served when comprehensive details are known about the original research.[132] When replication projects are limited to incomplete primary research descriptions harvested from published research reports, it is unclear whether the replication research design would meet the necessary criteria to be categorised as direct or systematic replication research. If it is unknown how similar the design of a replication research project was to the original research project, then it may be more appropriately labelled as a conceptual replication (irrespective of the replication

[129] Schmidt, 'Shall We Really Do It Again?'; Rachel Rosen, 'Research Transparency and Replication at MDRC', *MDRC*, January, 2018, accessed June 22, 2021, https://www.mdrc.org/publication/research-transparency-and-replication-mdrc.
[130] Elliott, 'A Taxonomy of Transparency'.
[131] Open Science Collaboration, 'Estimating Reproducibility'.
[132] Further discussion can be found in Section 5.2.

researchers' intent).[133] That it is more common for a replication attempt to fail when one or more of the original researchers are not involved in the replication attempt suggests that the ways in which research is commonly communicated may be insufficiently tailored for effective replication practices.[134] However, how research design transparency should be actualised is rarely agreed upon.[135]

3.4.1. Against Understated Transparency

Too little transparency in research communication has been proposed as a contributing factor to the replication crisis in psychological research.[136] A common argument for one cause of the replication crisis in various research fields is that necessary details are excluded from research communication.[137] Two replication research methodologies, direct and systematic, require detailed description of the original research's contextual background, research complex, conceptual operationalisation and characteristics of participants/subjects, if any. Yet, this information is often excluded from research publications.[138] Therefore, researchers interested in performing a replication of the original research must contact the original researcher(s) to obtain the details necessary to perform either a direct or a systematic replication. That the current situation encourages communication between research teams is not in and of itself problematic. Rather, increased communication between original researchers and replication researchers may be desirable. However, historically there has been little guarantee that replication researchers will be able to contact the original research team.[139] In addition, if contactable, there is no guarantee that the original research team will either:

- be willing to disclose the information necessary to perform a direct/systematic replication; or
- have comprehensively documented the information necessary to perform a direct/systematic replication.

[133] Further discussion can be found in Section 5.2.
[134] Open Science Collaboration, 'Estimating Reproducibility'.
[135] Further discussion can be found in Chapter 4.
[136] Figueiredo Filho et al., 'Seven Reasons Why'; Rosen, *Research Transparency and Replication at MDRC*.
[137] Schmidt, 'Shall We Really Do It Again?'; Crüwell et al., 'Seven Easy Steps'; Elliott, 'A Taxonomy of Transparency'.
[138] Crüwell et al.
[139] Crüwell et al., 240–1.

Comprehensive research design detail is necessary for replication research. Direct replications of original research should align to the conditions of the original research as closely as possible, with no intentional, incremental changes. Conversely, systematic replications of original research should align to the conditions of the original research as closely as possible with specific, intentional, incremental changes. While it is a nomological impossibility for replication research to exactly mimic every detail of an original research project, aligning as closely as possible when performing a direct or systematic replication is an important step in knowledge development on any research topic.[140]

Understating details concerning the research design and practice are harmful to replication attempts. Whether omitted through intention or apathy, failure to document any (original and replication) research project comprehensively may distract from the research conclusions. Replication of a thinly documented research project is not necessarily impossible. Failure to record and/or share the specifics of the research design (e.g. contextual background, conceptual operationalisation and characteristics of participants/subjects) may be irrelevant to the replicability of the research conclusions. Certain research conclusions are robust enough that they can be replicated under vastly different research conditions.[141] However, this style of replication is more accurately positioned within the bounds of the third type of replication: conceptual replication.

One common solution proposed for research design transparency is that research design variables, as well as raw, unanalysed data, should be registered with a third party.[142] As Schmidt suggested, this step will prevent QRPs and benefit replication research design. However, although pre-registration will help to diagnose research that has been identified as questionable, as well as to guide replication research design, it is unclear how it will protect against QRPs in the analysis stage of research. Pre-registration alone does not guarantee QRPs will be absent from research analysis; rather, with pre-registration, QRPs may be easier to identify after research conclusions are called into question.

[140] Further discussion can be found in Section 5.2.
[141] Schmidt, 'Shall We Really Do It Again?', 96.
[142] Crüwell et al., 'Seven Easy Steps', 240–1.

Moreover, unexpected confounding factors may be introduced during the research process that have a quantifiable effect on the research outcomes, without changing any of the pre-registered conditions. For example, a project exploring the levels of wellness and anxiety in a general population may be affected if a global pandemic occurs partway through the project. It is conceivable that the original research project was not designed to include the effect of the pandemic. However, it would be a QRP if the pandemic were not accounted for in the data analysis.

Another solution proposes that as part of research communication, researchers need to explain/justify the reasoning behind their research design decisions.[143] However, again, this may be helpful for post-hoc research diagnosis and replication research design, but it is unclear how this would protect against miscommunication of research conclusions. While transparency through pre-registration and design justification will undoubtedly be beneficial for post-hoc diagnosis of research and replication research design, it is unclear how as standalone concepts either is necessarily protective against QRPs and bias-driven results.

Yet, both pre-registration and design justification are useful for replication research design. Although it is unclear how transparency might serve as the sole solution to the replication crisis,[144] I nevertheless conclude that it is a useful and necessary component. As Schmidt asserted, both direct and systematic replication research design are best served via comprehensive levels of communication on research design variables.[145] Anything less reduces the utility of the replication research in those areas that direct and systematic replication research are tailored to illuminate. Consequently, I argue that although complete transparency in research design and practice may not protect against research communication failures, it is useful for independent post-hoc analysis of research, as well as for direct and systematic replication research design.

The utility of salient research details for replication researchers supports the idea that the lack of transparency can be detrimental to the ongoing development of a given field. However, I propose that this idea demonstrates a contextually undesirable level

[143] Elliot, 'A Taxonomy of Transparency'; Schmidt, 'Shall We Really Do It Again?'
[144] I find insufficient evidence to support transparency alone as being sufficient to correct for the entire replication crisis.
[145] Schmidt, 'Shall We Really Do It Again?'

of transparency, rather than serving to demonstrate the necessity of complete, unjudged transparency.

If transparency is an epistemic virtue, then the utility of transparency for replication research supports the positioning of understated transparency as an epistemic vice. However, reasonable concerns are found in arguments that suggest unfiltered transparency can also be undesirable.

3.4.2. Against Overstated Transparency

Some have argued that too much transparency in research communication is both onerous and a contributing factor to the rise in agnotological doubt about research conclusions. Kevin C. Elliot identified two primary arguments against transparency in research communication:

- Transparency is too difficult.
- Transparency is too dangerous.

The first argument against transparency suggests that the level of detail proposed as being necessary for research to be truly transparent is unreasonably difficult. As summarised by Elliot, arguments that transparency is too difficult propose that it is unreasonable to expect researchers to exhaustively document all aspects of the research process.[146] The level of detail proposed by Stefan Schmidt as being useful in the reporting of research design decisions goes beyond current norms.[147] Elliot did not address Schmidt's propositions in relation to the argument for difficult transparency, but Elliot's description of the argument can easily be applied as a counter to Schmidt. If even the levels of research design documentation commonly proposed in the discourse are argued to be unreasonably laborious, then, for similar reasons objections would be raised to providing the additional level of detail that Schmidt proposed.

The second argument against transparency suggests that the level of detail proposed as being necessary for research to be truly transparent is confusing and/or distracting for non-expert readers. As Elliot summarised, there is concern that unnecessarily detailed research reporting may be off-putting for non-expert readers, and therefore

[146] Elliot, 'A Taxonomy of Transparency'.
[147] Further discussion can be found in Section 5.3; Schmidt, 'Shall We Really Do It Again?'

detrimental to the role of research communication in general society.[148] Elliot summarised a point exhaustively made by John, that certain levels of transparency in the research process have been misused by agnotologists to promote an anti-research agenda.[149] The argument for dangerous transparency suggests that because non-experts are untrained to correctly interpret research details, they should be shielded from it. However, overstatement can still be a problem for replication researchers.

Some differences in research conditions between original research and replication research have little bearing upon the research outcomes. Thus, inclusion of this level of specificity may be unreasonably distracting. For example, if a research project explores the effect of psychedelic compounds on participant depressive symptoms in a controlled environment, then (unless there is evidence to suggest otherwise) the level of lumens emitted by the laboratory lightbulbs is likely unnecessary detail for inclusion in any research communication. By itself, a single superfluous detail may not be problematic; however, if details of every possible conceivable variable are communicated, the relevant details may be obscured.

The argument for difficult transparency and the argument for dangerous transparency both support the idea that an excess of transparency can be detrimental to research communication. Rather than reject either of these arguments, I instead agree that they raise valid concerns. However, I propose that they better describe a contextually undesirable level of transparency, rather than serving to reject transparency entirely.

If transparency is an epistemic virtue, then the argument for difficult transparency and the argument for dangerous transparency support the positioning of overstated transparency as an epistemic vice.

3.4.3. Virtue of Transparency in Research

Transparency in research communication is an orienting principle that can be used to find the appropriate level between understatement and overstatement. The appropriate level of transparency necessary to correct the replication crisis in psychological research is distinct from the appropriate level of transparency for general research communication. I agree that exhaustive recording of relevant research design decisions is both necessary and useful towards combating the replication crisis, but

[148] Elliot, 'A Taxonomy of Transparency'.
[149] Elliot, 'A Taxonomy of Transparency'; John, 'Epistemic Trust'

providing the same level of transparency in all research communication is likely unnecessary (and distracting) for the non-expert reader.[150] As the OSC identified, finding sufficient detail when performing a replication attempt of prior research can be unreasonably difficult.[151] Proposals for registration of prior research data as well as a comprehensive detailing of salient research design variables may be useful for researchers to appropriately engage in replication research. However, as John argued, dogmatically including this level of detail in all research communication is possibly detrimental for some readers.[152]

Therefore, contextual appropriateness should be reflected upon to identify the correct level of transparency for a given audience. For the sake of future analysis and the correct design of replication research, raw data and comprehensive details of research design variables, as well as justification for design decisions, should be recorded with independent bodies. Arguments in favour of this level of detail reasonably align with the needs of robust replication research. However, when choosing how to communicate their results, researchers should consider who the target audience of their communication is before they decide upon the appropriate level of detail to include. The level of detail appropriate for a research audit may differ from that appropriate for a research journal article, which may again differ from the level appropriate for a popular media article.

3.5. Chapter Summary

The primary conclusions of this chapter are as follows:

- Transparency is a pivotal term in psychological research that is conceived of differently by individual researchers as well as the researchers' audience.
- Appropriate transparency in any psychological research communication can only be judged in the context of the intended audience for that communication.

In this chapter, I have justified the characterisation of transparency in research communication as a contextually relative state. Arguments that promote a universal

[150] As has been discussed in Section 2.5.2 and will be again in 5.1.1; some practical errors in research practices are likely inevitable; however, this should not be taken as justification for not trying; Schmidt, 'Shall We Really Do It Again?'
[151] Open Science Collaboration, 'Estimating Reproducibility'.
[152] John, 'Epistemic Trust'.

level of transparency in research communication neglect the observable variety of states, wherein the appropriate level of transparency is highly dependent upon the intended audience of the communication. That there is an implicit assumption of a universal level of transparency distracts from the specific needs of a given research communication to record the necessary details. The level of transparency in research communication required for a personal research log differs from that required for an academic research report, which differs again from that needed for an executive summary of research conclusions. As John identified, too much transparency can lead to non-experts misinterpreting research practices and thus negatively affects the confidence of those same non-experts in research conclusions.

Therefore, as a framework for judging the level of transparency necessary for a given piece of research communication, I proposed *appropriate transparency* as an epistemic virtue. I positioned *appropriate transparency* as a desirable state that can only be judged in the specific context of the research communication to which it is being applied. Depending upon the context, too little transparency would indicate an epistemic vice of deficiency, which I refer to as *understated transparency*. Depending upon the context, too much transparency would indicate an epistemic vice of excess, which I refer to as *overstated transparency*. *Appropriate transparency* is a desirable state wherein the research communication conveys only that which is useful for the intended audience to accept the conclusions of the research project. I contend that providing audiences with that which is necessary for them to interpret the research project correctly is one tool towards correcting for the replication crisis in psychological research. However, the capacity to judge *appropriate transparency* requires the development of a shared RDN.

Chapter 4. Judging Appropriate Transparency Requires a Shared Nomenclature

In this chapter, I assume the following primary conclusions from the previous chapters:

- All researchers and their audiences do not necessarily use identical definitions of concepts related to psychological research (Chapter 1).
- It is useful to unpack pivotal concepts in psychological research to facilitate effective communication between researchers and their audience (Chapter 1).
- It is useful to interpret the replication crisis in psychological research as a crisis of confidence in research conclusions (Chapter 2).
- Transparency in research communication is a necessary step in restoring confidence in replication research conclusions (Chapter 2).
- Transparency is a pivotal term in psychological research that is conceived of differently by individual researchers, as well as their audience (Chapter 3).
- Appropriate transparency in any psychological research communication can only be judged in the context of the intended audience for that communication (Chapter 3).

I contend that a shared RDN provides a foundation upon which researchers can judge the appropriate level of transparency for the intended audience of their research communication. When this intended audience is a replication expert, a shared RDN provides a categorical framework for research designers to ascertain the information necessary for appropriate replication of their research projects. I argue that a shared RDN will encourage better research record keeping, which will aid in improved research communication, reduce verbal disputes and divergent research definitions between researchers and, ultimately, help to rebuild confidence in psychological research conclusions. Since it will reduce the effort required for researchers to identify the relevant research design variables applicable to their research project, they should be more likely to record the details necessary and sufficient for effective replication research. The provision of a shared point of reference for the definitions of various research design variables should enable researchers to identify the detail necessary to communicate to their intended audience more effectively.

A minimal shared RDN in psychological research can already be observed in the discourse. For example, one existing cultural expectation is that a research project's dependent variable/s and independent variable/s will be explicitly communicated. The expectation that dependent and independent variables will be communicated in (most) research communications, without the need for dependent variable and/or independent variable to be redefined in every research report, suggests that the utility (of at least some level) of a shared RDN is already broadly accepted.

However, more targeted effort needs to be made to develop the nomenclature further towards three related goals:

- First a robust nomenclature will aid researchers in identifying and recording the research design variables relevant to the specific research project on which they are working.
- Second, a shared point of reference for the utility of specific research design variables will make it simpler for researchers to identify the appropriate details to share with their intended audience.
- Third, a shared RDN will promote accurate, relevant communication of information and thus correct some easily avoidable mistakes in research design, aiding to restore confidence in research conclusions.

4.1. Ill-Defined Nomenclature and the Replication Crisis

A replication failure adds to the perception that there is a replication crisis; nevertheless, not all replication failure is symptomatic of a crisis. Some replication failure is both valid and desirable. Exploration of phenomena without failure is unlikely to identify the boundaries of the phenomena under examination. However, replication failure that is exclusively the result of ill-defined research design is misinterpreted by experts and non-experts alike to cast doubt on research conclusions.

The ill-defined nature of research design terms can be demonstrated as contributing to reduced confidence in psychological research. While I do not suggest that the lack of a shared nomenclature is the only cause of the replication crisis in psychological research, I do contend that it is a pivotal factor that must be addressed. This is demonstrated through the various ways in which replication methodologies are defined, and ill-defined, within the discourse. Although it is rarely explicitly expressed in the discourse, labelling a particular project as a replication of prior

research without explicitly identifying which of three replication methodologies was used is a case of understated transparency. Judgement about the validity of a replication research project's conclusions relies upon the answers to the following questions:

- Which replication methodology was used?
- Was the replication methodology used accurately named?
- Were the inferences attributed to the replication results appropriate for the replication methodology used?

Acceptable replication failure, and the inferences that can be validly drawn from it, depend upon the type of replication attempted.[153] Conflating the entire replication crisis with a failure of direct replications demonstrates both a misunderstanding of the arguments for declaring a crisis, as well as a misunderstanding of the intent behind different replication methodologies.

Stroebe and Strack focused their rebuttal of the replication crisis on arguing against direct replication as a solution for the crisis.[154] They proposed that the perception of a crisis has been fuelled by allegations concerning the prevalence of both fraud and QRPs in the discipline. They identified that those promoting direct replication as a panacea to the replication crisis demonstrate the mistaken assumption that research can be exactly replicated. Rather than implicitly privileging direct replications as the sole solution to the replication crisis, more attention needs to be paid to the existence of three different categories of replication research, and the strengths/weaknesses of each, as well as the role of replication failure within each category of replication research.[155] These categories are as follows:

- direct replications;
- systematic replications; and
- conceptual replications.

Direct replications are the category of research that replicates the research design of an original study as closely as is reasonably possible. In line with others within the discourse, Stefan Schmidt claimed that exact replication of research design is both

[153] Further discussion can be found in Section 5.2.
[154] Stroebe and Strack, 'The Alleged Crisis'.
[155] Further discussion can be found in Section 5.2.

impossible and undesirable.[156] Schmidt argued that instead, the inability for research to ever be exactly replicated is exactly what gives it its confirmatory power. According to Schmidt, exact research replication would only be useful to discuss the exactly specific conditions of that research design. Meanwhile, a common misunderstanding of exact replications assumes the capacity to duplicate source research identically. Due to this impossibility, the more accurate term *direct replication* is finding favour. Exact replication is acknowledged as unintentionally misleading, given that the conditions of research can never be exact for reasons such as:[157]

- the same participants cannot be involved in a state identical to before the first research project;
- if identical demographics are adhered to, then it is likely recruitment procedures need to be adapted; and
- if recruitment procedures are retained, then there is no guarantee demographics will be identical.

Instead, direct replications are designed to duplicate the conditions of prior research as identically as reasonable. Appropriate direct replication research follows the same research design decisions (where possible) as the original research, without any intentional iteration. Unlike direct replications, systematic replications make intentional changes to the original methodologies.

Systematic replication research makes intentional, incremental changes to the original research methodology. Systematic replications are similar to direct replications, whereby they are designed to preserve as much of the primary research methodologies as possible. The similarities between the two replication research methodologies mean that they are often conflated with each other.[158] However, systematic replication differs in the methodological decision to make incremental, reasoned, and intentional changes to the primary research design. Schmidt promoted this as the category that most appropriately describes most of the replication research being conducted, regardless of how the researchers referred to it. Similarly, Schmidt

[156] Schmidt, 'Shall We Really Do It Again?', 92.
[157] Excluding the theoretical existence of parallel universes.
[158] Nosek and Errington, 'What Is Replication?'; Machery, 'What Is a Replication?'; Schmidt, 'Shall We Really Do It Again?

maintained that conceptual replications are not always conceived of (or identified) as replications by the research team.

Direct replications and systematic replications are not always the appropriate replication methodology. Stroebe and Strack raised concerns about the overpromotion of direct (and assumedly, systematic) replications as the only correct way to perform replication research.[159] They contended that conceptual replications are important for testing the boundaries and external/ecological validity of theoretical constructs.

Conceptual replications preserve the underlying theory and/or hypothesis of the original research, while introducing a new research design. According to Schmidt, conceptual replication is the most forgiving form of replication research.[160] He reported that conceptual research is often not reported as replication research. Schmidt asserted that instead, some projects (that should be categorised as conceptual replications) are not intentionally designed nor treated as replication research. He argued that it is only after the research is published that the similarities with prior studies are drawn by others within the discourse. Nevertheless, Schmidt categorised all research that addresses the same phenomena as prior research, albeit in a novel way, as conceptual replication. Owing to the capacity of conceptual replications to probe the boundaries of a given theory, Harold Pashler and Christine Harris contended that this design is more valuable than direct (and assumedly systematic) replication research.[161]

Pashler and Harris rebutted the argument that to support the validity and generality of findings, conceptual replications are more important than direct replications.[162] It is accepted that if a theory is both valid and generalisable, then it will be observed under various research conditions; this is not disputed. However, as Pashler and Harris argued, if research conditions are never repeated, then it is easier for inaccurate conclusions to remain accepted. Therefore, they maintained that conceptual replications alone are not the answer to the replication crisis. While it is relatively uncontroversial to argue that replication failure is more appropriate in conceptual

[159] Stroebe and Strack, 'The Alleged Crisis', 60–1.
[160] Schmidt, 'Shall We Really Do It Again?', 96.
[161] Pashler and Harris, 'Is the Replicability Crisis Overblown?', 533–4.
[162] Pashler and Harris, 533–4.

replications, it is rarely addressed that what can be inferred from a replication failure differs for each of the three replication methodologies.[163]

Regardless of the methodology, replication failure identifies areas of research that require further evaluation. An argument that Stroebe and Strack, and Pashler and Harris appear to have agreed on is that the analysis of failures has identified some trends within psychological research that justify a lack of confidence in the research conducted. In their paper '*A Crisis of Confidence?'*, Pashler and Wagenmakers wrote that the evidence was insufficient to suggest the extent to which the issue of QRPs was prevalent within psychological research. However, they expressed concern that John Ioannidis was correct when he asserted that (across all of science) most published research findings are invalid and unreproducible.[164]

The common oversight regarding the explicit acknowledgment of the effect of a given replication methodology on the inferences that can validly be made from an observed replication failure demonstrates the effect of a lack of shared RDN.[165] That there is little to no transparency in how replication researchers orient themselves to cornerstone research concepts, creates an environment where it is necessary to continuously resolve contentious terms before the results of the research conclusions can be accurately considered. I agree with Stefan Schmidt that a shared RDN will reduce verbal disputes between researchers, foster shared meaning in research communities and expedite researchers' efforts to identify the correct research design decisions for the goals of the research they are conducting. Thus, I contend that rebuilding confidence in research design decisions will have a positive effect on the confidence placed in research conclusions.[166]

4.2. A Research Design Nomenclature as an Orienting Framework

A shared RDN will be useful for streamlining the effort needed to interpret research conclusions. The lack of a shared RDN has been offset by the capacity for commentary on research and research conclusions within the broader discourse. So long as those within the discourse are willing to interpret and argue the merits of a

[163] Further discussion can be found in Section 5.2.
[164] Ioannidis, 'Why Most Findings Are False'.
[165] Or even acknowledgement that there is more than one valid methodological way to approach replication, as discussed further in Section 5.2.
[166] Again, I do not argue that a shared RDN is the only step that needs to be taken towards restoring confidence. Instead, I argue that it is one (of a multiplicity) that needs to be taken.

given research project, the lack of a shared RDN can be offset; however, as in the case of the two addiction papers already discussed, this is not always the case.[167] When research is classified (or oriented) against terminology that can be interpreted ambiguously, effort must be expended to marry the differences between the use and the interpretation. This expenditure of effort can be alleviated (or avoided) by transparent reference to the nomenclature used by the researcher to inform the research design.

A shared nomenclature is a set of terms used to convey an agreed upon meaning and methodology. The lack of consensus around specific research design terminologies in the discourse around replication research makes the process of research communication unnecessarily cumbersome and contributes to the crisis of confidence in psychological research conclusions. For example, that Stefan Schmidt (among others) identified the term *exact replication* as a contentious term implies a lack of a shared meaning attribution that distracts from the research itself.[168] However, this is not a purely semantic distinction.

The lack of a shared nomenclature in relation to replication methodologies (and overall research design) contributes to the dispute on the validity of replication research conclusions. As will be discussed in further detail in the next chapter, the inferences that can be made from the failure of a direct replication differ from those that can be made from either a systematic or conceptual replication. When the incorrect term is used to describe a replication project, it is understandable when incorrect inferences are generated from the research conclusions. Despite this, when replication research outcomes fail to align with those of the original research, the appropriateness of the replication methodology used is rarely discussed in relation to the conclusions that are then drawn from the failure to replicate.[169] Referring to a particular research project as a direct replication, without reference to the definition of direct replication that oriented the researchers' design decisions, is less informative than could be accomplished with a simple citation.

[167] As already discussed in Section 1.3.
[168] Schmidt did not use the phrase contentious term; rather, his description of the various interpretations of the term aligns with Chalmer's definition of a contentious term; Chalmers, 'Verbal Disputes'; Schmidt, 'Shall We Really Do It Again?'
[169] Schmidt.

However, while I contend that the nomenclature that I synthesise in Chapter 5 is a useful foundation, it is not perfect and nor should it be. Part of my aim is to encourage a culture wherein researchers transparently address the definition of key terms with which they are working. This may be performed by referencing my (or other) definitions to demonstrate how the researcher oriented towards the topic, or by using my definition as a position to argue against. The nomenclature serves as a foundation from which to judge appropriate transparency, not as a way of proscribing research design.

A shared research nomenclature can provide useful touchstones for researchers to communicate more efficiently with their intended audiences. How research is conducted influences how research will be conducted in the future; researchers are influenced by the standards (explicitly or implicitly) used by other researchers. Stephen Vyse discussed how, in addition to engaging in p-hacking his own research, he was instrumental in helping other researchers to do the same.[170] Vyse was one of many engaging in a socially accepted practice. However, that p-hacking and other QRPs were (and still are) used (and defended) suggests that how research is being conducted has at least as much influence on other research projects as how research ought to be conducted.[171]

Similarly, how psychological research is communicated implies how psychological research ought to be conducted. The theory of *social learning* suggests that individuals' behaviours can be influenced by communication within related social groups.[172] That various QRPs have been argued as norms within research practice aligns with this interpretation of social learning.[173] Therefore, there is theoretical support to suggest that how psychological research is communicated will (positively or negatively) influence accepted psychological research standards. If a norm is detected in the discourse surrounding psychological research that implies a potentially inappropriate pattern of psychological phenomena, it is necessary to question the value of said norm.

[170] Vyse, 'P-Hacker Confessions'.
[171] John, 'Epistemic Trust'; Hunter, 'The Reproducibility "Crisis"'.
[172] Colman, *A Dictionary of Psychology*, s.v. 'social learning'.
[173] John, 'Epistemic Trust'; John, Loewenstein and Prelec, 'The Prevalence of Questionable Research Practices'.

As discussed, *exact replication* is a contentious term in replication research, one that is shifting (or has shifted) out of vogue.[174] In the discourse, individuals such as Stefan Schmidt have argued that the nomological impossibility implied by the use of the term *exact replication* is detrimental to research interpretation. In summary, because research cannot be exactly replicated, the inclusion of *exact* in the term is a misnomer.[175] Schmidt maintained that *direct replication* is more informative for orienting researcher (and reader) expectations as to the research design. That is, a direct replication project recreates the circumstances of the original research project as closely as is achievable and necessary without introducing any intentional changes to the research design. Any (even minor) intentional changes introduced into the research process invalidate labelling the project as a *direct replication*, and it should instead be referred to as a *systematic replication*.

This is not to argue that an RDN should necessarily be viewed as proscriptive as to the design decisions researchers can make. Rather, when a researcher communicates their research process to others, if they are oriented by a more idiosyncratic definition of a research design term and fail to make any reference to the way in which that term is defined, then, inevitably, researchers without the same frame of reference will misinterpret the research conclusions. As previously discussed, it is probable that some contentious terms are unavoidable. However, if a dispute can be anticipated and avoided with a single citation, then it should be easier for the focus to remain on the exploration of the research topic, rather than on the interpretation of the terminology used to describe the research. For example, the inferences that can be made from the failure of a *direct replication* differ from the inferences that can be made from either a *systematic replication* or a *conceptual replication*.[176] However, that difference is only immediately apparent if the research project is classified by the same criteria as it is then interpreted against.

In other words, the lack of a shared nomenclature is detrimental to psychological research replication practices. Schmidt claimed that a shared nomenclature provides a foundation for critically discussing replication research in psychological science.[177] Schmidt identifies the lack of consensus regarding the methodological application of

[174] Further discussion can be found Section 5.2.1.
[175] Without access to parallel realities.
[176] Further discussion can be found in Section 5.2.
[177] Schmidt, 'Shall We Really Do It Again?'

replication practices within psychological research as contributing to the replication crisis. He argued that the paucity of references to replication in basic textbooks, the inconsistent nomenclature in the discourse[178] and the dissonance between theoretical argument and research practice are detrimental to psychological research overall. In response to these concerns, Schmidt proposed definitions for targeted nomenclature relevant to replication research in psychology.

The lack of a shared nomenclature is demonstrated in the lack of guidance around appropriate replication standards in psychological discourse. Schmidt's criticism is illustrated by Sherri L. Jackson's *Research Methods in Statistics*, a commonly used[179] textbook designed to instruct novice researchers on appropriate research methods in psychological research.[180] Although Jackson highlighted the importance of replication in psychological research, she did not include specific guidance on the ways that psychology researchers do approach (or should approach) replication. While Schmidt did not directly reference Jackson's textbook, it nevertheless demonstrates the accuracy of Schmidt's criticism against how replication is taught within the discipline of psychology.

Schmidt argued that there is little agreement in the discourse on what comprises replication.[181] He synthesised a detailed argument towards:

- identifying categories of research design variables useful in communicating research;
- identifying the functions of replication in research; and
- classifying the three categories of replication methods already mentioned.

Similar definitions can be viewed in the writings of thinkers such as Graeme Porte, Wolfgang Stroebe and Fritz Strack, and Daniel J. Simons, but I am unaware of any

[178] For example: Hendrick, 'Replications, Strict Replications, and Conceptual Replications: Are They Important?'; Leshner, 'Addiction Is a Brain Disease, and It Matters'; Levy, 'Addiction Is Not a Brain Disease (and It Matters)'; Simons, 'The Value of Direct Replication'; Abend, 'What Are Neural Correlates Neural Correlates Of?'; Nosek and Errington, 'What Is Replication?'; Machery, 'What Is a Replication?'; Stroebe and Strack, 'The Alleged Crisis and the Illusion of Exact Replication'; Schmidt, 'Shall We Really Do It Again?'
[179] By my understanding.
[180] Jackson, *Research Methods and Statistics*.
[181] Schmidt, 'Shall We Really Do It Again?', 92.

study that approaches the issue as directly (and in as much detail) as that of Schmidt.[182]

However, at times Schmidt's word choice has heavily relied upon experimental research terminology; this apparent privileging of experimental research is also found within the wider discourse.[183] When I unpack Schmidt's ideas further in the next chapter, I reinterpret his arguments and synthesise them with research definitions provided by Jackson (and others) to demonstrate their applicability to the broader psychological research discourse.

4.3. Appropriate Transparency and Intended Audience

I argue that there is a middle ground between understating and overstating research design detail, whereby it is both appropriate and necessary to share why a primary or replication research methodology is chosen for a research project.[184] The intended audience is a useful indicator for identifying the appropriate level of research design detail in communication; transparent reference to RDN aids in communication between different intended audiences. However, whereas not including any reference to the decision process is potentially a problem of understatement, it is possible to overstate this decision process as well. A shared nomenclature provides a guide about the minimum relevant information that should be communicated and an agreed upon terminology for communicating it—hence reducing verbal disputes by providing a formalised scaffold for researchers to build upon, or, to diverge from explicitly. It would be useful for researchers to begin by conveying which of the goals of psychological research, and/or goals of replication in psychological research, a project is intended to orient towards.[185]

Researchers do not need to be responsible for defining research concepts and requirements themselves. I propose that a reason for the perception that transparency is too difficult partly stems from the lack of a shared nomenclature. Insisting that researchers must be more transparent in their practices, without providing clear

[182] Graeme Porte, 'Who Needs Replication?' *CALICO Journal* 30, no. 1, (2013). https://doi.org/10.11139/cj.30.1.10-15; Stroebe and Strack, 'The Alleged Crisis'; Daniel J. Simons, 'The Value of Direct Replication', *Perspectives on Psychological Science* 9, no. 1 (2014). https://doi.org/10.1177/1745691613514755; Schmidt.
[183] Bakker and Wicherts, 'The (Mis)reporting of Statistical Results'; Stroebe and Strack.
[184] As discussed further in Chapter 5.
[185] Further discussion can be found in Chapters 5 and 6.

specifics regarding how that transparency should be actualised, leaves the onus unfairly on the researcher.

For example, it would be inappropriate to require a researcher focusing on the topic of priming to provide an in-text comprehensive defence for choosing a direct replication design. As Elliot argued, a certain level of design justification is appropriate.[186] Yet, it is unclear why the average priming researcher would be expected to possess the expertise necessary to appropriately justify the differences between the three replication methodologies. That there is sometimes an inappropriate conflation between direct and systematic replication methodologies justifies the argument that experts in given research topics are not necessarily experts in research design terminologies.

However, a lack of a shared framework for making key research design decisions in research communication encourages a state of understatement. John contended that a reason for the rise of agnotologists is the existence of a 'folk philosophy of science' that misrepresents appropriate research practices.[187] While John focused this 'folk philosophy' on various statistical data practices, the problem can also be understood in broader methodological assumptions.[188] For example, psychological research is sometimes discussed as to privilege experimental methodologies as synonymous with appropriate scientific methods. The incorrect implication is that the only valid research methodology is experimental. However, experimental research has a specific use that does not address other specific goals of valid research.[189] For example, the observed correlation of targeted symptoms in patients of a given illness is not experimental research, yet it is still valid (and important) research that aligns with the goals of psychological research. The conflation of all forms of valid psychological research as experimental research supports the claim that many important research design terms are being ill-defined by both experts and non-experts.

Researchers can leverage a shared nomenclature to succinctly convey relevant design decisions. I propose that if both Schmidt and John are correct, then a shared nomenclature is one tool that can be used to restore some confidence in psychological research results without being too difficult or distracting. Improving the level of

[186] Elliott, 'A Taxonomy of Transparency'.
[187] John, 'Epistemic Trust', 80–2.
[188] John, 80–2.
[189] Further discussion can be found in Chapter 5.

transparency regarding research design decisions does require additional communication by the researcher, but the additional level of transparency need not be unreasonably onerous.

However, there is no reason to assume that a shared nomenclature would act as a panacea for the entire replication crisis in psychological research. Although a lack of transparency in research practices is partially to blame for the replication crisis, a shared nomenclature alone is neither going to resolve the issue of appropriate transparency, nor the myriad other problems contributing to the lack of confidence in replication research.[190] Nevertheless, as demonstrated by the utility of the shared meaning of dependent and independent variables, an expanded RDN is one way the oft-promoted concept of transparency can be realised.

4.3.1. Experts Communicating to Non-Experts

When communicating to a blended audience, a shared RDN can be useful in promoting appropriate transparency without being exhaustively detailed. As argued by Stephen John, an overstatement of research particulars can negatively affect the way in which non-experts learn from experts, or even the way in which non-experts engage with research conclusions. Research experts form a small percentage of the audience for whom the research is intended. Therefore, it would be questionable (at best) to argue that researchers can assume to be exclusively communicating with an expert audience. An argument could be put forward that researchers should only be required to communicate to an expert audience; however, that would still fail to address the variability observed in particular contentious terms between experts.[191]

A framework for shared meaning, which is referenced within the research communication, would demonstrate transparency that is both appropriate and useful for experts, tangential experts and non-experts alike. For example, when discussing replication with a non-expert, a shared understanding of the distinction between *direct replication*, *systematic replication* and *conceptual replication* cannot be assumed. These are contentious terms between experts; therefore, it would be invalid to assume non-experts are oriented by the same meaning attributions as the researcher. Assumed meaning attribution of research design terms would be an example of understated transparency. Conversely, it would also be unreasonably

[190] Further discussion can be found in Section 2.3.
[191] Further discussion can be found in Section 1.3.

cumbersome if a researcher were expected to define all research design terms within a single research report. A defence of a given definition of *direct replication* in every research project described as such would be an example of overstated transparency. Instead, a simple citation indicating the definition against which the researcher oriented their design provides useful information without distracting from the focus of the research communication.

4.3.2. Topic Experts Communicating to Replication Experts[192]

When communicating to an audience of replication experts, a shared RDN will offset some of the problems identified as contributing to the replication crisis in psychological research. The specifics of a given research project may not necessarily be intentionally obfuscated from the broader community; rather, there has been a historical limitation on what could be communicated due to practical publishing restrictions. A full account of the research design variables necessary for an appropriate direct replication project are unlikely to fit the word count limitations of an academic journal. In addition, many of these variables are likely of little interest (or necessary information) for all but a subset of those for whom the journal communication is intended. However, that the level of detail needed by replication experts differs from that needed by most should not be taken as justification for its exclusion from all record.

The call for greater research accountability as an offset to the replication crisis is well supported in the discourse. Projects such as Psych File Drawer are an attempt (with varying degrees of uptake) to provide avenues for research—that otherwise might not reach publication—to be openly accessible to all.[193] The existence of these resources is useful for replication researchers (as well as those performing meta-analyses), but even with the removal of the constraints of publication formats, there is little consensus about how researchers should communicate their projects to support future replication attempts better.

[192] Whereas John only distinguished between non-experts and experts, I find it useful to distinguish between different types of experts; John, 'Epistemic Trust'. I refer to individuals with a research background from a field unrelated to a given topic as *tangential experts*. I refer to individuals with a research background in the specific field of research as *topic experts*. I refer to topic experts who also engage in replication research in the same field as *replication experts*.
[193] *PsychFileDrawer*, 2012, http://www.psychfiledrawer.org/

4.4. Chapter Summary

The primary conclusions of this chapter are as follows:

- The lack of a shared RDN introduces unnecessary strain on replication research interpretation and extrapolation.
- A shared RDN would aid researchers in identifying the appropriate level of transparency for their research communication audience.

In this chapter, I have argued that a shared RDN provides a foundation upon which researchers can judge the appropriate level of transparency for the intended audience of their research communication. I identified the deficiency in detail concerning the current RDN as related to replication research. For example, I highlight the often-obscured privileging of direct replication research as the only valid form of replication research. As I will highlight in the next chapter, the specific research design that justifies the application of the direct replication label (usefully and appropriately) limits what can be inferred from the results of the replication project. For example, strict adherence to the original research design will provide little insight into the ecological validity of the conclusions of the original research.[194] However, a robust, shared RDN is not exclusively useful for designing and interpreting replication research.

I contend that explicit engagement with a shared RDN can usefully orient all psychological research design and communication. The foundation levels of an appropriately transparent RDN require explicit identification of some of the likely a priori assumptions of different researchers. For example, an individual definition of scientific psychological research that does not account for falsifiability as a critical descriptive component of valid psychological research would likely result in a different research design than one that factored in falsifiability. As discussed in Chapter 1, uncritically assuming shared meaning between individuals leads to unnecessary disputes that can more easily be accounted for by breaking pivotal concepts into their constituent components.

Likewise, the implicit privileging of different research design decisions as the only valid way of performing scientific psychological research unnecessarily limits

[194] I discuss this in more detail in Section 5.1.4.

researchers. As I discuss further when unpacking *conceptual replications*,[195] replication researchers will find it useful to have a foundational orientation concerning the different ways in which researchers can explore a psychological phenomenon. For example, there is as much functional utility in psychological research that seeks to describe a phenomenon as there is in research that seeks to explain a phenomenon. One approach is not more scientific than the other; rather, each approach will perform different functions in the ongoing exploration of the phenomenon. This is a useful consideration for replication experts performing conceptual replication as well as researchers designing their own primary research.[196]

[195] Section 4.1 & Section 5.2.3
[196] In Chapter 5, I define primary research as the category of research that comes before replication research.

Chapter 5. A Shared Research Design Nomenclature for Judging Appropriate Transparency

In this chapter, I assume the following primary conclusions from the previous chapters:

- All researchers and their audiences do not necessarily use identical definitions of concepts related to psychological research (Chapter 1).
- It is useful to unpack pivotal concepts in psychological research to facilitate effective communication between researchers and their audience (Chapter 1).
- It is useful to interpret the replication crisis in psychological research as a crisis of confidence in research conclusions (Chapter 2).
- Transparency in research communication is a necessary step in restoring confidence in replication research conclusions (Chapter 2).
- Transparency is a pivotal term in psychological research that is conceived of differently by individual researchers, as well as their audience (Chapter 3).
- Appropriate transparency in any psychological research communication can only be judged in the context of the intended audience for that communication (Chapter 3).
- The lack of a shared RDN introduces unnecessary strain on replication research interpretation and extrapolation (Chapter 4).
- A shared RDN would aid researchers in identifying the appropriate level of transparency for their research communication audience (Chapter 4).

A trend in the discourse is to use terms such as *direct replication* with the assumption that all members of the audience will attribute the same meaning as the researcher, which Stefan Schmidt identified as problematic.[197] However, as has also been discussed, there are few to no resources for a shared RDN that capture a sufficient level of specificity for certain replication research terms. The lack of a shared RDN negatively affects a researcher's ability to select appropriate terminology for their communication with their intended audience. Likewise, if the original researcher has used a given contentious term in an atypical fashion, it reduces the audience's confidence that the researcher has conducted their research within a scientifically rigorous framework. The problem is not exclusive to traditionally replication focused

[197] Schmidt, 'Shall We Really Do It Again?'

terminology, and instead is a larger problem that contributes to a lack of confidence in research psychology practices.

Similarly, I contend that an individual's (explicit or implicit) definition of science will inform their interpretation of research concepts. When writing about psychological science, I interpret the language used by Stefan Schmidt as (intentionally or unintentionally) inferring a definition of scientific research that is synonymous with experimental research. Schmidt's privileging of experimental research excludes other scientific research methodologies from what can or should be discussed as science. Experimental research (by design) explores the scientific goal of explanation; however, explanation is not the only valid goal of science. Sherri L. Jackson's definition of science contends that scientific research orients towards one of three goals (with possible overlap into the other two): describe, predict and explain.[198] Additionally, Jackson argues that appropriate psychological research is: rational, empirical, falsifiable, replicable, and critical. However, it is also worth making explicit that I argue from the interpretation that if one of Jackson's criteria is not observed, if the opposite state is also not observed (as in the case of replication and replication failure), then psychological research can still be classified as transiently valid. By making my use of Jackson's definition explicit, I intend to provide an appropriately transparent orientation for the interpretation of my arguments that will be subtly different than if they were interpreted through the lens of experimental research alone.

From this, I argue that when research is communicated with the appropriate level of reference to a shared RDN, the field of research psychology will be one step closer to restoring confidence in research conclusions. By providing researchers with a common orientation to align with (or diverge from):

- researchers can evaluate the appropriate level of transparency necessary for their projects; and
- audience members can evaluate research conclusions against a common understanding of research decisions.

In previous chapters, I discussed that a shared RDN is one way to achieve appropriate transparency in psychological research design and communication. In this chapter, I

[198] Jackson, *Research Methods and Statistics*, 15–6.

present an RDN as a framework for researchers to leverage towards appropriate transparency in research communication. The need for a realised (as opposed to theoretical) shared RDN is supported by the discussed observations that a lack of common orientation for specific terms within the literature has proven a barrier to effective communication between experts and non-experts, as well as between topic experts and replication experts.

However, the nomenclature discussed later in this section is not intended to promote a universal research norm. Rather, the proposed RDN is intended to serve as a framework and common reference against which to position specific research. As I have built upon many of the definitions presented by Stefan Schmidt, Schmidt himself has built upon definitions presented by Clyde Hendrick.[199] As Schmidt wove in additional specificity to Hendrick's definitions, I do the same with Schmidt's. My goal in doing this is to provide as robust a set of working definitions as possible (within the limitations of this thesis) to demonstrate some key considerations that should factor into incorporating appropriate transparency in research communication.

By synthesising a set of foundational definitions for an RDN from the broader discourse, I intend to demonstrate how appropriate transparency can be achieved in psychology research communication. Both overstated and understated transparency have already been discussed as problematic for rebuilding confidence in psychology research conclusions. However, without a common point of reference, the desirable state of appropriate transparency is more difficult for researchers to identify. As Schmidt claimed, a shared RDN removes some of the unnecessary burden from researchers when judging what is appropriate for inclusion in their research design and the subsequent communication of said research.

However, this nomenclature is as useful for researchers who disagree with any of the given definitions as it is for researchers who agree with them. If a researcher accepts the definitions presented in this chapter, then the framework under which they use a contentious term can be communicated with a single reference.[200] Similarly, if a researcher rejects this research terminology, they can focus on the individual points of difference, while referencing this terminology to provide the remainder of the

[199] Schmidt, 'Shall We Really Do It Again?'; Clyde Hendrick, 'Replications, Strict Replications, and Conceptual Replications: Are They Important?*' Journal of Social Behavior and Personality* 5, no. 4 (1990).
[200] As discussed further in Sections 5.1–5.3.

definition. Both scenarios provide a useful orientation for their audience to interpret the researchers' work in the specific context intended by the researcher.

I propose a RDN that would be useful for orienting both researcher as well as research audiences as to the characteristics of various (potentially) contentious terms. I do this via unpacking different categorical research design concepts that build upon each other to provide an orienting framework for research design and communication:

1. I distinguish primary research in psychology as that which comes before replication research in psychology.
2. I make explicit the goals of primary scientific research.
3. I consider a selection of primary research methods oriented towards different primary research goals.
4. I define the goals of replication research that are additive to the goals of primary research.
5. I categorise the different replication research methods that are additive to the primary research methods.
6. I promote a set of classes of research design variables.

Each of the sections of the nomenclature interweave in a related network with each section being informed by and informing the others. To highlight the links between each section, Figure 1. depicts some of the main connections:

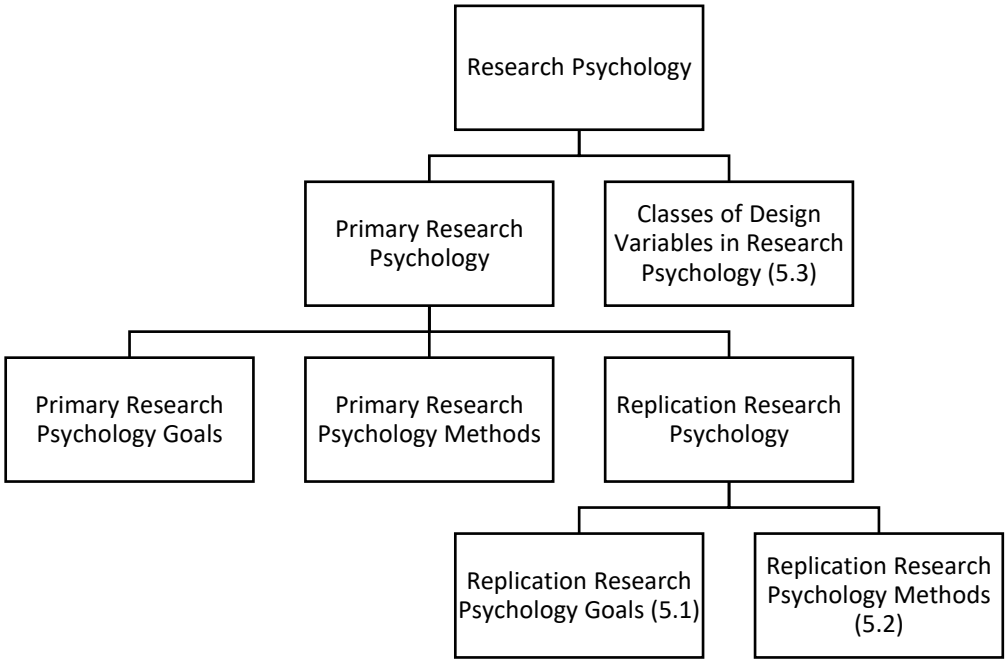

Figure 1. *Chart depicting the main connections between discrete sections of the proposed research design nomenclature.*

However, although this RDN is synthesised from extant literature, it is not exhaustive or universally applicable. As discussed in relation to appropriate transparency, all the concepts herein need to be judged against the contextual specifics of each individual research project, as well as the ways it will then be communicated to a specific research audience. While I do argue that all the classes of research design variables should be considered, and the variables judged relevant recorded in a manner easily communicated to replication researchers, it would be a vice of overstated transparency to include this detail in an executive summary of the research. Having a shared RDN provides an orienting framework intended to lessen the epistemological burden on researchers; however, researchers are still expected to apply their own practical judgement in identifying the characteristics relevant to their context.

In discussing research design, it is useful to identify the ways in which the assumptions, goals and methodological decisions that inform replication research differ from the original. No single term is available to distinguish the specifics of research that comes before replication research; in the context of a given replication project, the prior research is referred to as using words such as earlier, original and prior. However, when discussing the way in which replication differs from that which came before, having a single term to reference will aid in the clarity of discussion.

Primary research in psychology is the term I have given to all psychology research that is sufficiently unique to not be evaluated as a replication of any prior research paradigm. Sherri L. Jackson argued that research psychology aligns with three simple goals: describe, predict and explain. While these three goals are realised in both primary research in psychology and replication research in psychology, replication research in psychology has additional goals that do not necessarily inform primary research in psychology. *Replication research in psychology* is any research that replicates (whether by design or incidentally) one or more *primary research* projects and/or their findings/results.[201]

Although it may initially appear unnecessary to regress to the level of explicitly identifying the characteristics of psychological research that justify it as scientific, that argument assumes that all readers have a shared definition of the characteristics

[201] As discussed further in Section 5.2.3.

of science.[202] According to Jackson, the first goal of psychological science is to describe psychological phenomena. Jackson characterised this goal as the systematic observation and detailing of the patterns of activity and cognition of humans and other animals.[203] The second goal is to predict the outcomes of psychological phenomena. Jackson argued that after researchers have observed a series of patterns between specific stimuli, the environment and psychological phenomena, they can attempt to predict when a psychological phenomenon will be observed.[204] The final goal is to explain psychological phenomena. Jackson described how once a pattern is confirmed, an explanation is suggested and then tested.[205] She argued that the validity of an explanation is best tested through the systematic manipulation of the proposed causal factor.

Different types of primary research methods in psychology can be categorised as aligning more with one of the three goals of psychological science. Jackson used these goals to summarise and highlight the strengths and weaknesses of various example research methodologies, as shown in Table 1:

Table 1. Example Primary Research Methods Grouped by Orienting Primary Research Goal

Primary Goal	Example Research Methodologies
Describe	• observational (including naturalistic and laboratory) • case study • survey
Predict	• correlational methods • quasi-experimental
Explain	• experimental

The goals of psychological science (as a discipline) build upon each other: patterns of psychological phenomena are observed, observations are used to predict psychological phenomena and accurate predictions are used to generate explanations. However, although replication research in psychology also embodies these goals, this

[202] In addition, in his book *What is this Thing Called Science?*, Alan Chalmers was reticent to provide an ahistorical definition of science. However, unpacking this discussion further would be outside the scope of this thesis. Therefore, to be appropriately transparent with my intended audience I judge it necessary to briefly include the descriptive characteristics that inform my interpretation of scientific research psychology below; Alan F. Chalmers, *What Is This Thing Called Science?* (St Lucia: University of Queensland Press, 2013): 147–60.
[203] Jackson, *Research Methods and Statistics*, 15.
[204] Jackson, 15.
[205] Jackson, 16.

category of research introduces additional goals that are not necessary for primary research design.

5.1. Replication Research Goals in Psychology

The goals of replication research are additional to the goals of primary research, in that primary research goals will factor into replication research, but replication research goals do not necessarily factor into primary research. Replication research shares the goals of primary research in that it is often used to describe, predict and explain psychological phenomena. However, as Stefan Schmidt argued, certain factors that inform replication research design are not as immediately relevant to primary research design.

I have developed the following six goals of replication from the interpretation of both Stefan Schmidt, as well as the broader, psychological discourse. Schmidt proposed a similar set of factors that he referred to as the five *functions of replication.* I have built upon Schmidt's functions to create the following goals by introducing a new goal (Control for Practical Error), as well as by revising the language Schmidt used to avoid privileging experimental research.

I argue that there are six goals specific to replication research that do not necessarily inform primary psychology research design. When creating his similar list, Schmidt relied heavily upon the language of experimental research to justify his summarisation of the five functions of replication.[206] However, rephrasing his definitions makes it is possible to highlight the broader applicability of Schmidt's arguments to a wider range of methods of psychological research. I have reordered and rephrased Schmidt's five functions of replication as the final five of the six goals of replication research in the list that follows (and have listed Schmidt's corresponding function in parentheses under the related goal).[207] In addition, the first listed goal has no direct analogue in Schmidt's functions; however, it has been identified as a factor that replication research is effective at identifying.[208]

i. Control for Practical Error;
ii. Control Internal Validity;
    - (2: To control for artefacts)

[206] I use this list as the foundation for my goals; Schmidt, 'Shall We Really Do It Again?', 93.
[207] Schmidt, 93.
[208] John, Loewenstein and Prelec, 'The Prevalence of Questionable Research Practices'.

iii. Control External Validity;
    - (1: To control for sampling error)

iv. Control Ecological Validity;
    - (4: To generalise results to a larger or to a different population)

v. Investigate Primary Research Theory;
    - (5: To verify the underlying hypothesis of the earlier experiment)

vi. Control for Fraud.
    - (3: To control for fraud)

A researcher need not design a primary research project with direct reference to these goals—because a necessary characteristic of psychological research is that it is replicable—but it would be useful for primary researchers to consider how each of the six replication goals might be affected by their initial research design. Similarly, when considering how (or if) their design explicitly addresses each of these goals, cases where replication research conclusions are being misinterpreted will be more obvious.[209] The additional level of definition for each of the six goals of replication will also be useful when discussing how each class of research design variable relates to the replicability of primary research.

5.1.1. Control for Practical Error

Practical error is an unavoidable risk in all forms of research. Two aspects of practical error need to be considered: human error and process error. First, researchers are fallible and thus their performance is unlikely to be consistently infallible. Second, sometimes elements such as environmental factors, faulty instrumentation and methodological conventions can introduce error into research results.

Thus, some problematic data in psychological research may be the result of fraud, whereas others will be the result of simple, practical error.[210] However, differences in results between primary research and replication research should not be taken to automatically assign fault to the primary research project. Practical error can occur in both primary research and/or replication research. If practical error is suspected, both primary and replication projects should be subjected to the same rigorous investigation.

[209] As in the case of a single replication failure being used to discredit an entire body of research.
[210] Fraud is discussed further in Section 5.1.6.

Errors in research conclusions may be initially unrealised and upon further analysis found to be the result of issues such as a slow stopwatch, inaccurate behavioural inventory and mathematical rounding. When the results of different projects contradict each other, one of the first steps should be an audit for all forms of practical error; for example:

- A researcher may incorrectly transcribe results during data entry and/or data cleaning.
- A primary research project report may not reference critical design variables necessary for the accurate replication of the study.
- A replication researcher may misapply certain statistical principles.
- During analysis of the primary research data, a critical number may be rounded up that is then rounded down in the analysis of the replication research data, resulting in enough shift between results to appear incomparable.

5.1.2. Control Internal Validity

Replication is useful to identify a research project that lacks internal validity. Schmidt explained the problem as one wherein the interactions of variables within a research project have unintended effects.[211] In a more expanded definition than that provided by Schmidt, Andrew M. Colman explained *internal validity* as the accuracy of the conclusions of a project being the result of the research method, namely, '[t]he extent to which the conclusions of an empirical investigation are true within the limits of the research methods and subjects or participants used'.[212] In addition, Colman defined an *artefact* as an unintentional by-product of the research process. These definitions somewhat align with and expand upon Schmidt's explanation.[213]

Hence, *internal validity* can be compromised when research variables interact in a manner unintended by researchers. This can be related back to the already discussed example of priming research; a research project is designed to demonstrate changes in *participant behaviour* as correlating with perception of *age-related words*. The primary research project supports this observation, yet replication research demonstrates that (in addition to the changes observed by the original research) there

[211] Schmidt, 'Shall We Really Do It Again?', 94.
[212] Colman, *A Dictionary of Psychology*, s.v. 'internal validity'.
[213] Colman, *A Dictionary of Psychology*, s.v. 'artefact'.

is also a correlation observed between *participant behaviour* and *experimenter knowledge*. If *experimenter knowledge* is not accounted for within the primary research, then the internal validity of both projects must be questioned. Perhaps the relationship described by the primary research is more complicated than being exclusively an interaction between *participant behaviour* and *age-related words*; or perhaps the observation in the replication research that *experimenter knowledge* unintentionally influences participant behaviours is a misdirection; or perhaps there are many other possible interpretations. Further investigation of both projects is required to justify the results of either. However, not all valid psychological research is conducted using experimental methods.

For research methodologies that do not directly deal with the interaction between variables, internal validity will be explored through the conclusions drawn from the data collected. For example, the accuracy of assumptions that a series of questions or behaviours aligns with and/or demonstrates a psychological phenomenon may be incorrect. Primary researchers may interpret the results to suggest one thing, whereas replication researchers may draw different conclusions from similar data. Again, the mistake should not automatically be assumed in favour of either research project. When a discrepancy of this nature is encountered, both projects should be equally scrutinised.

The opposed conclusions of different research may be the result of a lack of internal validity in either of the projects. For instance:

- A confounding variable may not have been controlled for in the primary research.
- A test used in the replication research may have been a poor measure of the concept being explored.
- The replication researchers' own bias may have (incorrectly) informed the way in which they coded the research observations.

5.1.3. Control External Validity

Patterns observed during a research project could be unique to that specific research paradigm. Any research involving statistical analysis deals with the probability of the broader accuracy of results. Thus, while data may be collected perfectly, there is always the possibility that the conclusions are the result of chance alone. Similarly,

the results of a research project may be specific to an individual (or set of individuals) that are therefore not an accurate representation of the psychological phenomena in general. The applicability of research conclusions outside the specifics of the individual research paradigm is referred to as the external validity of the results. Colman defined *external validity* as '[t]he extent to which the conclusions of an empirical investigation remain true when different research methods and research participants or subjects are used'.[214] This aligns with Schmidt's argument that replication can control for sampling error. However, because Schmidt's definition relies upon a representation of specific statistical terminology, I judge it more appropriate to reinterpret Schmidt's argument using Colman's definition.[215]

The opposed conclusions of different studies may be the result of a lack of external validity in either of the projects' observations. For example:

- A primary research project describing typical visual perception may have inadvertently included participants with atypical vision.
- A replication research project predicting the typical effect of language on behaviour may have been inadvertently skewed by one or more of the research methods used.
- A replication research project explaining the typical mating behaviours of chimpanzees may unknowingly focus on subjects with an atypical social structure.

5.1.4. Control Ecological Validity

Ecological validity is concerned with the applicability of research observations outside of a research setting. The results of psychological research may be intrinsically linked to the process of research itself. Whereas external validity tests the applicability of research results under different research conditions, ecological validity is concerned with the applicability of research results in a more naturalistic setting. Schmidt's fourth function ('To generalize results to a larger or to a different population'[216]) can be interpreted as overlapping with some of the concerns covered under external validity. However, it can also be interpreted to include the applicability of research conclusions outside of a research situation. Colman defined

[214] Colman, *A Dictionary of Psychology*, s.v. 'external validity'.
[215] Schmidt's use is particularly idiosyncratic within the discourse.
[216] Schmidt, 'Shall We Really Do It Again?', 93.

ecological validity as how (or if) the results of research apply to situations outside of research conditions; '[t]he confidence with which the conclusions of an empirical investigation can be generalised to naturally occurring situations in which the phenomenon under investigation occurs'.[217] In his fourth function, Schmidt argued that the replication of research can highlight the applicability of research conclusions as pertaining to different population sizes and/or compositions.[218] Therefore, I conclude that the similarities in definitions justify the rephrasing of Schmidt's fourth function while maintaining his point.

The conclusions drawn from all research projects may be argued to be biased by the act of the research itself. For instance:

- A primary research project describing the honesty of human children in a laboratory setting may not accurately describe the honesty of human children in an educational setting.
- A replication research project predicting the mating habits of giant pandas in the wild may not be reflective of the mating habits of giant pandas in captivity.
- A replication research project wherein participants are instructed to perform a given task may not produce results comparable to research that observed individuals voluntarily choosing to engage in the same behaviour.

5.1.5. Investigate Primary Research Theory

Replication research is important for exploring the underlying theory informing individual research. The language used by Schmidt implies that research hypotheses and/or theories can be verified.[219] However, since verification is not reasonably achievable, I rephrase Schmidt's argument to be consistent with frameworks that leverage falsification as a useful tool.

Via his fifth function, Schmidt inaccurately asserted that the replication of results can be used to verify a hypothesis. However, in the context of Schmidt's writing (as well as the broader discourse), I propose that Schmidt did not use *verify* such that it would

[217] Colman, *A Dictionary of Psychology*, s.v. 'ecological validity'.
[218] Schmidt, 'Shall We Really Do It Again?', 93.
[219] Schmidt, 94.

imply a necessary uniformity of nature (the problem of induction).[220] Rather, (in context) I interpret Schmidt as using the term verify to express a position of increased justification. Likewise, Schmidt's exclusive use of *hypothesis* (and other terms) inappropriately privileges experimental research as the only valid method of psychological research. Therefore, (to avoid further confusion) I have rephrased Schmidt's function to be broader and less controversial.

Replication failure is an important tool for exploring the validity of theories. By common definition, a psychological theory is a proposition that explains the underlying principles of an observed psychological phenomenon. Therefore, it may be incorrectly surmised that a theory only directly relates to experimental research. However, this would be inaccurate because although a theory may be supported by experimental research (focused on the goal of explanation), it is equally likely to have been informed by research focused on observation and/or description.[221]

Replication research is useful for identifying the boundaries of various psychological phenomena. A replication research project may identify issues such as the following:

- Different levels of experimenter knowledge concerning the research theory correlate with different research outcomes.
- Different surveys concerning the same phenomenon result in different outcomes from the same participants.
- Different beliefs in personal ability to effect change influence the likelihood of an individual acting to prevent others from a perceived harm.

5.1.6. Control for Fraud

Replication is a useful tool for challenging the veracity of previous research conclusions by considering whether the patterns observed in the primary research misalign with those observed in the replication research; however, a priori assumptions of fraud against a research project are rare. Accusations of research fraud are a weighty topic that needs to be discussed with care and consideration, but nevertheless needs to be discussed.

---

[220] For a discussion on verification and the problem of induction, see Leah Henderson, 'The Problem of Induction', In *The Stanford Encyclopedia of Philosophy*, Spring 2020 ed., ed. Edward N. Zalta (Stanford: Metaphysics Research Lab, Stanford University, 2020), https://plato.stanford.edu/archives/spr2020/entries/induction-problem/

[221] As illustrated in Table 1.

Replication research is proposed as an important tool for identifying intentionally deceptive research results. When a replication attempt fails, one cause may be that results were fabricated. Researchers have been found to fabricate either parts of, or entire, research projects.[222] It does not follow that a replication failure necessarily denotes a fraudulent research project, but further analysis of both research projects may uncover deceptive research behaviours. However, although replication failure can uncover research fraud, the original research may not necessarily be fraudulent. If fraud is suspected, both original and replication projects should be treated to the same rigorous investigation.

Yet, given that replication research has played little to no part in uncovering some of the largest cases of fraud in the history of research psychology, the practical actualisation of this goal may be called into question. [223] However, the potential for replication to play a role in identifying potential fraud remains, as does the potential for replication to be used as a supplementary tool for investigating suspected cases of fraud, regardless of the uncomfortable discourse around the issue.

5.2. Replication Research Methods in Psychology

As has been discussed, in research psychology literature, it is often not immediately apparent which methodological approach a replication project has adopted. The three methodological approaches are:

i. direct replication;
ii. systematic replication; and
iii. conceptual replication.

However, the differences in methodological approach to replication research are not always recognised within the literature.[224] Therefore, further unpacking of the characteristics of the different replication approaches is useful for interpreting the appropriate inferences that can be made from replication failure and/or replication success.

[222] Vyse, 'P-Hacker Confessions'; Wicherts, 'Psychology Must Learn a Lesson'.
[223] Wicherts.
[224] Schmidt, 'Shall We Really Do It Again?', 90–1.

Schmidt's definitions of each of the three methods are spread throughout his paper.[225] He reserved certain subsections for discussing each method in focused detail but also made key points in other subsections. Thus, the following descriptions are synthesised from Schmidt's entire paper as well as other definitions found in the wider discourse.[226]

Although Schmidt did not argue for a necessarily linear relationship between direct, systematic and conceptual replication, the differences between each method are linear:

- Direct replication preserves as much of the original design as possible.
- Systematic replication observes the effect of planned changes to the research design.
- Conceptual replication introduces new research design to investigate the same phenomena.

However, as Schmidt stressed, all three of these methods can be used to categorise replication research after it has already been completed.[227] Schmidt proposed that it is rare for replication researchers to consider these types of methodological distinctions before the research is conducted. That is not to suggest that it is never included as a design decision; rather, Schmidt asserted that it is not a necessary predesign decision. The concepts can still be usefully applied to interpret research once it has already been completed.

Moreover, Schmidt argued that each of the three different methods are better suited for addressing specific replication goals, as summarised in Table 2.

[225] Schmidt.
[226] Nosek and Errington, 'What Is Replication?'; Machery, 'What Is a Replication?'; Stroebe and Strack, 'The Alleged Crisis and the Illusion of Exact Replication'; Simons, 'The Value of Direct Replication'; Schmidt, 'Shall We Really Do It Again?'; Hendrick, 'Replications, Strict Replications, and Conceptual Replications: Are They Important?'; Fiona Fidler and John Wilcox, 'Reproducibility of Scientific Results', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2021 (Metaphysics Research Lab, Stanford University, 2021), https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/.
[227] Schmidt, 97.

Table 2. Replication Research Goals Grouped by Orienting Replication Research Goal(s)

Replication Method	Replication Goals
Direct Replication	Control for Practical Error Control External Validity Control for Fraud
Systematic Replication	Control Internal Validity Control External Validity
Conceptual Replication	Control Internal Validity Control External Validity Control Ecological Validity Investigate Primary Research Theory

5.2.1. Direct Replication Methods

Schmidt defined direct replication as the category of research that replicates the design of a primary study as closely as is reasonably possible.[228] He claimed that *exact replication* (as it is sometimes referred to) of research design is both impossible and undesirable.[229] Instead, Schmidt agreed with Harry Collins that the inability for research to ever be exactly replicated is what gives it its confirmatory power.[230] Exact research replication would only be useful to investigate the exact conditions of that specific research design.

The specific research history of participants necessitates that exact replication of research is nomologically impossible. According to Schmidt, for research replication to be exact, all the design variables would need to be unchanged.[231] If even one is changed, then the replication project is not exact. As it is impossible for identical participants to be re-engaged in the same state they were in when first participating, *exact replication* is nomologically impossible. This is a rather narrow definition of *exact* that nevertheless highlights Schmidt's reasoning for supporting the use of *direct* (instead of *exact*) to describe this category of replication research.

Control for Practical Error. Direct replications are better suited for identifying practical error in data collection, interpretation and analysis. Human error and process error are more likely to be identified when the same research design is being utilised by the replication project.

[228] Schmidt, 95–96.
[229] Schmidt, 92.
[230] Harry Collins, *Changing Order: Replication and Induction in Scientific Practice* (University of Chicago Press, 1992); Schmidt, 92.
[231] Schmidt, 92–93; research design variables are discussed further in Section 5.3.

Control External Validity. Direct replications are useful for testing the external validity of original research conclusions using a different group of participants. Even if nothing else about a research project is changed, the participants will differ, and thus, it is possible to identify whether the same observations are made under directly related conditions.[232]

Control for Fraud. Intentional fraud is generally found in the details of the research data more so than in the artefacts or theoretical framework that informs the research. A false theory without false data would not be considered fraud—it would simply be a bad theory. However, regardless of the validity of a theory itself, providing false data intentionally is always a case of fraud. Thus, because they are designed to target the data with as little alteration to the research process as reasonable, direct replications are positioned better to identify fraudulent research than systematic replications and/or conceptual replications.

5.2.2. Systematic Replication Methods

Systematic replication builds upon primary research by making reasoned and intentional changes to the research design variables.[233] Although Schmidt promoted this as the category that most appropriately describes the majority of replication research, he argued that it is often insufficiently systematic. Schmidt defended this with reference to the problem of an insufficient consensus regarding appropriate replication methods. Schmidt's assessment is supported by an example I discussed previously—Sherri L. Jackson's textbook on research psychology methods.[234] In this guide, she argued that replication is a necessary criterion for science, yet provided minimal (to no) guidance on how to perform replication.

Control Internal Validity. Since systematic replications are designed to introduce reasoned changes to the original research design, replication researchers can choose their iterations to specifically explore the internal validity of the original research.

Control External Validity. Systematic replications allow researchers to explore the applicability of the observations under moderate iterations of research variables. This allows for testing the boundaries of the original research conditions and identifying

[232] Either different participants; or (excluding the possibility of parallel universes) the same participants having already engaged in the research, as well as all other experiences they have subsequently had, exist in a different state from when they were involved in the primary research.
[233] Schmidt, 'Shall We Really Do It Again?', 96–97.
[234] Jackson, *Research Methods and Statistics*.

the factors that may or may not contribute to the expression of the targeted psychological phenomena.

5.2.3. Conceptual Replication Methods

Conceptual replication preserves the underlying theory and/or hypothesis of the original research, while introducing a new research design. According to Schmidt, conceptual replication is the most forgiving form of replication research.[235] He reported that conceptual research is often not reported as replication research, proposing that some (appropriately categorised) conceptual replications were not intentionally designed as such.[236] Schmidt argued that, instead, it is only after the research is published that the similarities with prior studies are drawn by others within the discourse. Nevertheless, Schmidt categorised all research that addresses the same phenomena as prior research did, albeit in a novel way, as conceptual replication.

Control Internal Validity. Conceptual replications are intended to explore the same phenomena with drastically different research designs, and hence, the replication researcher has more scope to intentionally explore the internal validity of the original research by introducing radically different conditions upon which to test the conclusions.

Control External Validity. Conceptual replications can be used to explore the validity of the original research conclusions under conditions that bear little to no resemblance to the original research design. As with systematic replication, this method can be used to explore the factors of the target phenomenon, to identify whether the observed expression of the psychological phenomenon is intrinsically linked to the original research design.

Control Ecological Validity. Conceptual replications are more suited for controlling ecological validity than either direct or systematic replications. If experimental research is not immediately concerned with exploring phenomena in naturally occurring situations, then, because direct and systematic replication methods would both require adherence to the original research method, a different primary research

[235] Schmidt, 'Shall We Really Do It Again?', 96.
[236] Schmidt, 96.

method must be used to explore the ecological validity of the original research conclusions.

Investigate Primary Research Theory. Conceptual replications are better suited to exploring the underlying theory of primary research. Research that is designed to explore the same theory, but with dramatically different design principles, is still a method of research replication. Schmidt concluded that significant changes in the way in which a theory is explored (primary research method) provides meaningful insight into the generalisability of the theoretical underpinnings.[237] The greater the diversity of methodological investigation into a theoretical claim, the clearer the boundaries of said theory should become.

5.3. Classes of Design Variables in Research Psychology

The four classes of design variables I define in this section draw heavily from the classes of contextual variables proposed by Stefan Schmidt. Schmidt detailed four classes of contextual variables that research designers should consider explicitly for facilitating replication attempts between research projects. By contextual variables, Schmidt referred to the specific details regarding the research project design. However, as Schmidt also used the word *contextual* in one of his four classes of variables, and I have used *contextual* heavily in my discussion of transparency, I have judged it as more informative to rephrase the overarching categorisation of these classes of variables as design variables.

Research design variables are either the conditions imposed regarding, or choices made by a researcher that inform, a research project. These variables are distinct from research topic variables in that they underpin and describe the decision points made in how the research project is actualised. For example, if exploring the topic of anger, there will be a variable measure used to quantify changes in anger; depending upon the context, the measurable quantity of anger may be either the dependent or independent variable; thus, anger is a research topic variable. However, the test chosen by the researcher to measure anger is a research design variable.

Different research design variables are important information to convey to different audiences. In discussing the classes of research design variables, I commonly refer to their utility in reference to replication research, because this is the style of research

[237] Schmidt.

communication that requires the highest level of detail. However, some of the design variables discussed will still be relevant to different audiences. Nonetheless, a universal list is impractical since the relevance of each variable is contextual. Primary researchers will need to exercise practical judgement in identifying the variables that demonstrate the appropriate level of transparency for the intended audience of their research communication.

Schmidt agreed with Clyde Hendrick that certain elements of research design itself (unintentionally) influence the results of the research project.[238] Moreover, Schmidt identified an implicit assumption in the discourse that design variables are irrelevant to research conclusions.[239] However, Schmidt argued (and I agree) that this assumption is incorrect because the impact of said variables can be observed in the broader discourse.

Hendrick argued for eight classes of design variables, but Schmidt rephrased Hendrick's classes in four ways to develop four classes of design variables:

i. Schmidt agreed with Hendrick's first class.
ii. Schmidt converted six of Hendrick's classes into subclasses of Schmidt's second class.
iii. Schmidt introduced a new third class.
iv. Schmidt expanded upon Hendrick's eighth class to form Schmidt's fourth class.

However, as with Schmidt's definitions for the functions of replication research, Schmidt's defence of his classes relied heavily upon the language of experimental psychology. [240] Therefore, I have rephrased Schmidt's classes and subclasses to align with my interpretation of the broader discourse.

Without changing the order, I have revised Schmidt's definitions to make them compatible with the broader psychological discourse. This has been done to highlight the relevance of Schmidt's definitions outside of the limitations of experimental research. Schmidt's corresponding description is included after each class and subclass in Table 3.

[238] Schmidt, 93; Hendrick, 'Replications: Are They Important?'
[239] Schmidt, 'Shall We Really Do It Again?', 93.
[240] I have reinterpreted these into my goals of replication research.

Table 3. Classes of Research Design Variables compared with Schmidt's nomenclature.

Label	Title	Schmidt's Nomenclature
Class 1	Participant Research Interpretation	Primary information focus.
Class 1a	Immaterial Information Aspects	Immaterial informational aspects.
Class 1b	Material Realisation	Material realisation.
Class 2	Contextual Background	Contextual background of the experiment.
Class 2a	Characteristics of the Participants	Participant characteristics.
Class 2b	Specific Research History of the Participants	Specific research history of the participants.
Class 2c	Cultural and Historical Context of the Study	Cultural and historical context in which the study is embedded.
Class 2d	General Physical Setting of the Research	General physical setting of the research.
Class 2e	Control Agent	Control agent.
Class 2f	Task Details	Specific task variables.
Class 3	Participant Selection Procedures	Procedures for the selection and allocation of the participants.
Class 4	Operationalisation	Procedures for the constitution of the dependent variable.

In Schmidt's detailing of the characteristics of each class, some overlap is observed; a single type of variable can be categorised under multiple classes. For example, a variable concerning the type of materials presented to research participants may be classified under both Class 1 and Class 2. However, Schmidt said that each of the classes of variables interacts differently with the various goals of replication.[241] Therefore, the different effects of a single design variable may require it to be considered through the lens of multiple classes.

Irrelevant Design Variables

Although I contend that the classes of design variables are useful for orienting researcher thinking, I do not propose that transparency around every design variable will be universally appropriate. Sometimes, design variables are only found to be relevant (or irrelevant) after they are changed between studies. Despite Schmidt identifying four comprehensive classes of design variables, he recognised that some may be irrelevant to specific research projects.[242] Schmidt maintained that there are two ways in which the relevance of variables can be judged once research is repeated:

- If the results of a replication study align with the results of the original study, any design variables that were changed between studies may be **irrelevant** to the research focus.

[241] Schmidt, 'Shall We Really Do It Again?', 93–4.
[242] Schmidt, 97–8.

- If the results of a replication study fail to align with those of the original study, any design variables that were changed between studies may be **relevant** to the research focus.

Schmidt did not propose that researchers must account for every conceivable design variable from the outset.[243] Rather, he proposed that post-hoc analysis of linked primary and replication research may illuminate which design variables inform the results interpretation of those specific research designs/topics.

5.3.1. Class 1: Participant Research Interpretation

Design variables that influence the participant's (explicit and implicit) interpretation of the research intent may influence research outcomes (referred to by Schmidt as the *stimulus complex*). In agreement with Hendrick, Schmidt argued for this class to identify the variables that inform the participant's cognitive construction related to the research's independent variable.[244] To generalise Schmidt's description of this class, it is necessary to unpack some of his terms.

Since Schmidt did not provide a clear definition for his use of the term *complex*, I use Andrew M. Colman's *A Dictionary of Psychology* to interpret Schmidt. Colman provided three definitions of a *complex*, the first two of which provide a framework that appears to align with Schmidt's use of the term.[245] First, Colman defined *complex* as a structure composed of connected parts.[246] Colman's second definition relates to the concept of *complex* in psychoanalysis as a cognitive construct that influences thoughts, emotions and behaviour.[247] In context, I interpret Schmidt's reference to a *stimulus complex* as the way in which the participant's (explicit and implicit) interpretation of the research intent is (unintentionally) influenced by the research design, which may have a subsequent impact upon the observations made of said participant.

As discussed, Schmidt's arguments that privilege experimental psychology terminology can be rephrased so that they apply to all research psychology (not only experimental research psychology). Schmidt's definition refers to the *stimulus*

[243] Schmidt, 96.
[244] Hendrick, 'Replications: Are They Important?', 44; Schmidt, 'Shall We Really Do It Again?', 93.
[245] Colman, *A Dictionary of Psychology*, s.v. 'complex'.
[246] '1. An organized structure made of interconnected units'; Colman.
[247] '2. In *psychoanalysis, an organized collection of ideas, emotions, impulses, and memories that share a common emotional tone and that have been excluded either partly or entirely from consciousness but continue to influence a person's thoughts, emotions, and behaviour'; Colman.

complex that surrounds the *independent variable* as the feature that informs any interaction between the *independent* and *dependent variables*. If it is interpreted that in experimental psychology the interaction between the *independent* and *dependent variables* is another way of describing the core focus of the research project, then research focus can be substituted for independent variable in Schmidt's definition.

Thus, there is a class of design variables that informs a participant's interpretation of the research focus and therefore may inform the observations made of said participant. Schmidt further categorised this class of variables into two subclasses:[248]

- **Class 1a: Immaterial Information Aspects**—the immaterial way in which the research focus is conveyed (e.g. instructions and events).
- **Class 1b: Material Realisation**—the material actualisation of the Immaterial Information Aspects.

For example, adjectives used to describe a research survey may influence a participant's response to the survey itself. Different adjectives used to describe the same survey may result in statistically different results.

5.3.2. Class 2: Contextual Background

The context of the research project influences research outcomes. Schmidt reclassified six of Hendrick's classes as subclasses within his own contextual class of research design variables.[249] Schmidt's contextual class is used to identify the importance of transparent detailing of all individuals involved in the research, as well as the consideration of environmental and task variables. The specific subclasses are:

- **Class 2a: Characteristics of the Participants**—specific demographics of all participants, such as gender identity, age, ethnicity, cultural upbringing and education level.
- **Class 2b: Specific Research History of the Participants**—such as the details of a participant's experience with research projects and their motivation for participating.
- **Class 2c: Cultural and Historical Context of the Study**—details pertaining to language, cultural beliefs, attitudes, expectations and the historical context of the region in which the study is conducted.

[248] Schmidt, 'Shall We Really Do It Again?', 93.
[249] Schmidt, 'Shall We Really Do It Again?', 93; Hendrick, 'Replications: Are They Important?', 43–4.

- **Class 2d: General Physical Setting of the Research**—the specific environmental conditions under which the study was conducted, such as lighting, ambient noise and temperature.
- **Class 2e: Control Agent**—the specific demographics of the researcher.[250]
- **Class 2f: Task Details**—the minute details of the research task, such as the font of text and the colour of interactive materials.

For example, individuals that receive course credit to participate in research (motivation) may demonstrate statistically different results than those participating out of a sense of altruism.

5.3.3. Class 3: Participant Selection/Allocation Procedures

Schmidt briefly proposed that procedures for the selection and allocation of participants can affect the research outcomes. Schmidt maintained that the methodology behind how participants are selected and allocated may influence research outcomes.[251] This is Schmidt's first addition to the classes identified by Hendrick.

For example, consider a scenario wherein only a limited number of research participants can be accommodated, the number of volunteers exceeds this number and all volunteers act with the same motivation. Schmidt's argument is that the way the researchers select from this pool of volunteers may influence the research outcomes. In addition, if the research divides participants into more than one group, the way in which they are allocated between groups may also have an effect.

5.3.4. Class 4: Operationalisation

The way in which psychological concepts are defined, measured and manipulated may influence research outcomes. Schmidt concurred with Hendrick that the way data are produced and presented may influence the research outcomes.[252] Schmidt extended this assertion to include materials and methods related to the dependent variable in an experiment. However, when comparing Schmidt's fourth class with the research design fundamentals presented by Sherri L. Jackson, this class of variables can be subsumed under Jackson's explanation of operational definitions.

[250] Individuals may be implicitly biased towards different research behaviours/interpretations; Schmidt, 'Shall We Really Do It Again?', 93.
[251] Schmidt, 93.
[252] Schmidt, 93; Hendrick, 'Replications: Are They Important?', 44.

Jackson explained that the operational definition of a research variable describes the activities a researcher implements to measure and manipulate said variable.[253] To do so, the researcher must operate from a specific definition of the concept that supports the appropriateness of the way in which they operationalise the concept. Thus, Jackson's explanation contains the details of Schmidt's fourth class while also allowing it to be extended to include the assumptions of all research variables.

For example, there are multiple psychological concepts of intelligence.[254] Class 4 variables will not only indicate which theory of intelligence a researcher is working within but also, which methods and measures will be used to observe psychological phenomena judged to be expressions within the boundaries of the given theory. The justification for the relevance of said psychological phenomena (among other details) are also Class 4 design variables.

5.4. Chapter Summary

The primary conclusions of this chapter are as follows:

- An RDN should be as explicit as possible without being unnecessarily proscriptive.
- An RDN is most useful when it retains sufficient flexibility for experts to apply practical judgement regarding appropriate transparency.

In this chapter, I have developed a series of interrelated conceptual definitions that are useful for identifying the level of appropriate transparency when designing and communicating psychological research—an RDN. I have argued that it is more important that a researcher use the RDN than that they use the nomenclature I have proposed; however, the nomenclature I have developed is a useful example, if nothing else. As part of this interpretation, I have judged that if an a priori assumption has the potential to reshape an RDN, then it needs to be described as part of the nomenclature. Consequently, I divided the nomenclature into seven interrelated concepts.

I argued that a researcher's definition of science will inform their interpretation of research concepts. Therefore, I included Jackson's definition of the characteristics of valid psychological research. However, I also clarified my interpretation of Jackson's

[253] Jackson, *Research Methods and Statistics*, 58–59.
[254] Colman, *A Dictionary of Psychology*, s.v. 'intelligence'.

criteria as being that if one of the criteria is not observed, as long as the opposite state is also not observed (as in the case of replication and replication failure), then psychological research can still be classified as transiently valid.

I also proposed that it is useful to distinguish between primary and replication research. Although it may not be immediately apparent why a formal distinction between the two is necessary, when discussing the different goals and methods relevant between primary and replication research, the utility of an identifiable term becomes apparent. Specifically, although some research goals and research methods are used by both primary and replication research, these are better understood under the broader concept of primary research. In addition, certain research goals and research methods are exclusive to replication research. Primary research goals and methods inform valid replication research; however, replication research goals and methods do not necessarily inform valid primary research.[255]

Last, I reinterpreted Schmidt's arguments about the categories of research design variables that are useful for communicating primary research to replication experts. In alignment with Schmidt, I argued that there are classes of design variables that are useful for researchers to consider when documenting their research for communication to potential replication experts. Although identification of these variables is primarily useful for replication research design, without primary researchers factoring these variables into their research design (and then communicating them appropriately to replication experts), replication researchers will be forced to use their own practical judgement. Although it is not inherently negative that a replication researcher must exercise their own judgement when designing a replication project, it is possible that any differences between projects may unintentionally affect the research outcomes. When primary researchers report as much design detail as reasonable, less unnecessary repetition and/or invalid inferences can be expected among replication projects.

[255] It would be useful for all primary researchers to aim for their research to be replicable, and thus, they would find it useful to consider the goals of replication, but it is not necessary to the design of primary research that the goals of replication be considered a part of the primary research design paradigm.

Chapter 6. Patterns and Examples of Appropriate Transparency

In this chapter, I assume the following primary conclusions from the previous chapters:

- All researchers and their audiences do not necessarily use identical definitions of concepts related to psychological research (Chapter 1).
- It is useful to unpack pivotal concepts in psychological research to facilitate effective communication between researchers and their audience (Chapter 1).
- It is useful to interpret the replication crisis in psychological research as a crisis of confidence in research conclusions (Chapter 2).
- Transparency in research communication is a necessary step in restoring confidence in replication research conclusions (Chapter 2).
- Transparency is a pivotal term in psychological research that is conceived of differently by individual researchers, as well as their audience (Chapter 3).
- Appropriate transparency in any psychological research communication can only be judged in the context of the intended audience for that communication (Chapter 3).
- The lack of a shared RDN introduces unnecessary strain on replication research interpretation and extrapolation (Chapter 4).
- A shared RDN would aid researchers in identifying the appropriate level of transparency for their research communication audience (Chapter 4).
- An RDN should be as explicit as possible without being unnecessarily proscriptive (Chapter 5).
- An RDN is most useful when it retains sufficient flexibility for experts to apply practical judgement regarding appropriate transparency (Chapter 5).

I divide this chapter into two different sections of examples that demonstrate appropriate transparency for an intended audience. I contend that appropriate transparency is important at both the research design as well as the research communication stages. However, the most straightforward way to judge whether a researcher has been appropriately transparent is by auditing their research communication. Appropriate transparency will manifest differently depending upon the intended audience of a research communication; therefore, to judge appropriate

transparency in research design it is useful to consider which intended audience requires the highest level of transparency: the replication expert.

In the first section, I focus on the level of appropriate transparency between experts and replication experts. As reiterated throughout this thesis, appropriate transparency for a replication expert is achieved through exhaustive levels of research design specificity. This level of detail would demonstrate *overstated transparency* for most intended audiences; however, for a replication expert, the more detail regarding the concepts presented in the RDN, the less likely a replication failure due to issues with the research design. When replication failure is observed and issues with the research design can be eliminated as contributing factors, then the focus of analysis can remain on the psychological phenomenon being explored. Thus, a reduction in unnecessary replication failures will be one step towards restoring confidence in the conclusions of psychological research. However, another step towards restoring confidence is the appropriate transparency in communication between experts and non-experts.

In the second section, I focus on the level of appropriate transparency between experts and non-experts. Replication experts need to be appropriately transparent when communicating with both other experts and non-experts; yet, the same level of transparency is useful for all research communication, not simply replication research communication. As Stephen John claimed, overstated transparency between expert and non-experts has been observed to encourage agnotological thinking.[256] However, unlike John, I do not agree that following from this, no utility can be identified in being transparent with non-experts. Rather, I contend that reference to a shared replication RDN allows experts to demonstrate the appropriateness of their research design decisions, without unnecessarily distracting from summary information that they have judged as most useful for a non-expert audience. As John pointed out, excluding any reference to research design decisions is likely to unnecessarily alienate a portion of the non-expert audience; however, I reject John's false dichotomy wherein the only other option is complete unfiltered transparency. Instead, I propose that a shared RDN can be leveraged to provide context to research communication; then, the onus is on the non-expert to decide whether it is necessary to follow-up on this reference, or whether the reference is sufficient to engender confidence that the research has followed accepted norms.

[256] John does not use the term overstated transparency.

6.1. Patterns of Transparency Appropriate for an Audience of Replication Experts

In this section, I present three examples of information from the RDN that highlight patterns useful for an audience of replication experts. I have aligned with the arguments of Alfred N. Whitehead, Gottlob Frege, David Chalmers, and Gabriel Abend (and others) that suggest assuming identical meaning between individuals can be problematic.[257] I have built upon this position to demonstrate how this problem has been observed in the discourse about the replication crisis. To avoid continuing this mistake, I argue that it is useful for primary researchers to communicate using the necessary definitions of an RDN to provide meaningful orientation to potential replication experts.[258] It is apparent that a primary researcher may not be able to anticipate when, or if, their research will be the subject of a replication project; yet, that replication is a cornerstone of psychological research implies that primary researchers should endeavour to aid replication of their research to the extent that is practical.[259] Thus, primary researchers need to communicate the research design variable details that they can anticipate. However, as already discussed, this does not require primary researchers to consider and communicate every conceivable variable.

Building upon the RDN, in each example, I summarise a hypothetical research scenario, as well as relevant decisions in line with the RDN; I provide a brief example of selected research design variables that a researcher may consider relevant to record. Then, using the same framework as the primary research example, I build upon the primary scenario to demonstrate one example from each of the three replication methods, as well as the changes that would be expected. These examples are by no means intended to be exhaustive; rather, these examples are intended to be illustrative.

If researchers were to record every variable that may or may not affect their research outcomes, the research process would become impractical. For example, I am unaware of any research that would suggest that the make and model of a monitor affects research that uses an observational methodology to covertly describe the

[257] Chalmers, 'Verbal Disputes'; Whitehead, *Science and the Modern World*; Frege, 'Sense and Reference'; Abend, 'What Are Neural Correlates?'
[258] Primary researchers can use either the RDN presented in this thesis or another similarly explicit source of definitions.
[259] As was discussed in Chapter 5.

behaviour of research subjects. Therefore, without reason to believe it is relevant, communication of this information to replication researchers would be an example of overstated transparency. However, due to the understanding that different make and model monitors generate colour differently, the inclusion of the specifics of the monitor would be appropriate if the research were related to colour perception in any way. Hence, which information is relevant in a research communication depends upon the research context.

Seemingly innocuous research design decisions, when informed by different contexts, can become less innocuous. Appropriate transparency requires the application of a researcher's own practical judgement, but also the acceptance that sometimes researchers will get it wrong. If a researcher is suspected of overstated or understated transparency, then that in and of itself may orient the decision to replicate a primary research project. I have stated that I am unaware of any research that would suggest the make and model of a monitor as having any impact on research using an observational methodology to covertly describe the behaviour of research subjects. However, if I had reason to suspect this claim, then it would be appropriate for me to perform experimental research wherein the make and model of a monitor used for observation is the independent variable being manipulated. This could be performed as a systematic replication of existing primary research, or it could be explored within a new primary research project.

Even given the definitions from the RDN, the interpretation of research design variables is not straightforward. I argue that rather than a weakness, this is a strength that encourages the application of researchers' practical judgement. Some examples of research design variables that I present below:

- are open to interpretation under multiple classes within the research design variable nomenclature;
- can be broken down further into more specific sub-variables; and/or
- cannot be affected by researchers.

This is intended to illustrate the characteristics of certain variables and highlight patterns for individual researchers to consider. For example, I am not an expert in the mating behaviours of chimpanzees; if I were to attempt to provide an exhaustive list of specific research design variables relevant to research into chimpanzee mating

behaviours, I would make mistakes that would obscure the point I am trying to convey.

Based upon their own practical judgement, researchers need to proactively consider the research design variables that may be necessary for replication experts to recreate their research. I recognise that this process will affect researchers' workload, for the reasons already discussed, but it is not an unnecessary burden.

6.1.1. Replication Expert: Primary Research

Primary Research Scenario. A research team is interested in the mating behaviour of chimpanzees in their natural habitats.

Primary Research Goal. Describe.

Primary Research Method. Naturalistic Observation.

Selected Research Design Variables:[260]

- Research team (Class 1a);
- Researcher background (Class 2e);
- Unique subject identifier (Class 2b);
- Subject species (Class 2a);
- Hidden cameras' make and model (Class 4);
- Monitors' make and model (Class 4);
- Time of year (Class 2d);
- Daily weather (Class 2d);
- Mating time of day (Class 4);
- Mating length (Class 4);
- Mating interruptions (Class 4);
- Sex of mating initiator (Class 2b);
- Age of mating initiator (Class 2b);
- Sex of mating participant (Class 2b); and
- Age of mating participant (Class 2b).

[260] The results of the variable itself differ from the variable as a design decision. For example, that there is a *research team* is a design variable, and the members of that team would be the result attached to the variable.

6.1.2. Replication Expert: Direct Replication Research

Replication Research Scenario. A research team suspects that the uninfluenced mating behaviours of chimpanzees in natural habitats in a prior research project was the result of a mistake in the data analysis. However, as the original research data are not available for reanalysis, the research is being replicated to explore whether the original conclusions are supported.

Replication Research Goal. Control for Practical Error.

Replication Research Method. Direct Replication.

Unchanged Primary Research Goal. Describe.

Unchanged Primary Research Method. Naturalistic Observation.

Unchanged Research Design Variables:

- Research team (Class 1a);
- Researcher background (Class 2e);
- Subject species (Class 2a);
- Hidden cameras' make and model (Class 4);
- Monitors' make and model (Class 4);
- Time of year (Class 2d);
- Daily weather (Class 2d);
- Mating time of day (Class 4);
- Mating length (Class 4);
- Mating interruptions (Class 4);
- Sex of mating initiator (Class 2b);
- Age of mating initiator (Class 2b);
- Sex of mating participant (Class 2b); and
- Age of mating participant (Class 2b).

Changed Research Design Variables:

- Research team (Class 2f);
- Unique subject identifier (Class 2b);

6.1.3. Replication Expert: Systematic Replication Research

Replication Research Scenario. A research team suspects that the descriptive details of a prior research project exploring the mating behaviour of chimpanzees in their natural habitats was influenced by the equipment used in the research.

Replication Research Goal. Control for Internal Validity.

Replication Research Method. Systematic Replication.

Unchanged Primary Research Goal. Describe.

Unchanged Primary Research Method. Naturalistic Observation.

Unchanged Research Design Variables:

- Researcher background (Class 2e);
- Unique subject identifier (Class 2b);
- Subject species (Class 2a);
- Monitors' make and model (Class 4);
- Time of year (Class 2d);
- Daily weather (Class 2d);
- Mating time of day (Class 4);
- Mating length (Class 4);
- Mating interruptions (Class 4);
- Sex of mating initiator (Class 2b);
- Age of mating initiator (Class 2b);
- Sex of mating participant (Class 2b); and
- Age of mating participant (Class 2b).

Changed Research Design Variables:

- Research team (Class 1a);
- Hidden cameras' make and model (Class 4).

6.1.4. Replication Expert: Conceptual Replication Research

Replication Research Scenario. After observing the mating behaviours of chimpanzees in their natural habitats, the same research team theorises that the specific mating behaviours observed are being influenced by the palm wine regularly consumed by the subjects.

Replication Research Goal. Control for External Validity.

Replication Research Method. Conceptual.

Changed Primary Research Goal. Explain.

Changed Primary Research Method. Experimental.

Unchanged Research Design Variables:

- Research team (Class 1a);
- Researcher background (Class 2e);
- Subject species (Class 2a);
- Hidden cameras' make and model (Class 4);
- Monitors' make and model (Class 4);
- Time of year (Class 2d);
- Daily weather (Class 2d);
- Mating time of day (Class 4);
- Mating length (Class 4);
- Mating interruptions (Class 4);
- Sex of mating initiator (Class 2b);
- Age of mating initiator (Class 2b);
- Sex of mating participant (Class 2b); and

Age of mating participant (Class 2b).**Introduced Research Design Variables:**

- Subjects' access to naturally fermenting palm wine (Class 4).

6.2. Minimal Transparency Appropriate for a Non-Expert Audience

Although the most obvious benefit of a shared RDN is found in communication between experts and replication experts, since the replication crisis is usefully interpreted as a crisis of confidence in research conclusions, appropriate transparency between experts and non-experts can also usefully leverage an explicitly identified set of research concepts. In this section, I present a series of examples demonstrating a contextually appropriate minimum standard of research design detail useful for a non-expert audience.

Building upon the definitions of primary research goals and replication goals from the RDN, in each example, I summarise a hypothetical research scenario, as well as

relevant decisions in line with the RDN. I provide a brief example of how these decisions could be communicated in-text without the need for a researcher to redefine each term. Then, using the same framework as the primary research example, I build upon the primary scenario to demonstrate one example from each of the three replication methods.

I reject the implication that transparency in communication between experts and non-experts will inevitably lead to non-expert confusion. Stephen John argued against transparency with non-experts in research communication by highlighting the ways in which non-experts misinterpret research data and methodologies.[261] I agree with John that it can be a concern when too much transparency is included in communications intended for a non-expert audience (overstated transparency). However, where John's arguments and my own diverge is that John appears to use this argument to support entirely rejecting transparency in research communication. Yet, perhaps John intended his argument to apply exclusively towards communication created for a non-expert audience. Even so, I argue that making no attempt at transparency between experts and non-experts is harmful. If no attempt at transparency is made, then the lack of confidence in research conclusions expressed by both non-experts and other experts is reasonable.[262]

Moreover, that non-experts may initially lack the shared language necessary to correctly interpret research design decisions does not necessitate that relevant research design decisions cannot be communicated usefully to them. As I have argued, one factor that contributes to non-expert (and expert) confusion about research design is that it is not communicated using a set of standardised terms. However, by including key research design terms with reference to the wider discourse, experts can be appropriately transparent with non-experts (and other experts) without needing to exhaustively define the terms themselves. However, how *transparency* is being defined is as important as the concepts it is intended to make explicit.

The following examples are demonstrative of one kind of contextually appropriate transparency.[263] I suggest that failure to address the goals of research and/or

[261] John, 'Epistemic Trust'.
[262] Kahneman, 'A Proposal'.
[263] Only the information within the Replication Research Communication boxes would be provided to the non-expert audience.

replication is often demonstrable of understated transparency (deficient level of explicit definition). However, (depending upon context) it is unlikely that more is needed than a simple reference to the researcher's chosen definition of pivotal concepts. The examples (in boxes) in Sections 6.2.1 to 6.2.4 illustrate how different replication research design concepts can be included in-text to provide appropriate transparency without distracting from the level of summarisation useful for non-expert audiences.

6.2.1. Non-Expert: Primary Research

The sentence/s I propose as demonstrating contextually appropriate transparency in goal-oriented research decisions is underlined in the following examples. In addition, I have used an asterisk to denote where it would be appropriate to include a reference to the research design definition.

Primary Research Communication. To explore the uninfluenced mating behaviours of chimpanzees in natural habitats, the research team chose to use hidden cameras installed in a national park to reduce the likelihood of an observer effect influencing the chimpanzee behaviours. Since the goal of this research project was to describe* the behaviour without predictive or explanatory biases, a naturalistic observation* method was chosen.

Primary Research Scenario. A research team is interested in the mating behaviour of chimpanzees in their natural habitats.

Primary Research Goal. Describe.

Primary Research Method. Naturalistic Observation.

6.2.2. Non-Expert: Direct Replication Research

Replication Research Communication. In the results section of Gaffe's (2019) report into the uninfluenced mating behaviours of chimpanzees in natural habitats, the results reported appear incongruent with the physical limitations of chimpanzees as observed by Meticulous (1923). Unfortunately, as the original research data were not available for reanalysis, the research project must be repeated to further explore the original research conclusion. Because the goal of this replication project was to control for practical error* in the primary research* analysis, a direct replication* method was chosen.

Replication Research Scenario. A research team suspects that the uninfluenced mating behaviours of chimpanzees in natural habitats in a prior research project were the result of a mistake in the data analysis. However, as the original research data are not available for reanalysis, the research is being replicated to explore whether the original conclusions are supported.

Replication Research Goal. Control for Practical Error.

Replication Research Method. Direct Replication.

Unchanged Primary Research Goal. Describe.

Unchanged Primary Research Method. Naturalistic Observation.

6.2.3. Non-Expert: Systematic Replication Research

Replication Research Communication. The research team suspected that noise generated by the hidden cameras that Miser (2020) used to explore the mating behaviours of chimpanzees in natural habitats had a transformative effect on the behaviours observed. For this replication, Miser's (2020) original research design was adhered to with the sole exception of a different model of hidden camera being deployed. Since the goal of this research project was to control for internal validity* in the primary research* conclusions, a systematic replication* method was chosen.

Replication Research Scenario. A research team suspects that the descriptive details of a prior research project exploring the mating behaviour of chimpanzees in their natural habitats were influenced by the equipment used in the research.

Replication Research Goal. Control for Internal Validity.

Replication Research Method. Systematic Replication.

Unchanged Primary Research Goal. Describe.

Unchanged Primary Research Method. Naturalistic Observation.

6.2.4. Non-Expert: Conceptual Replication Research

Replication Research Communication. The research team suspected that the conclusions of Bawdy (2019) concerning the mating behaviour of chimpanzees in their natural habitats were influenced by the subjects' diet. For this replication, subjects were moved to a laboratory environment and randomly divided into two groups. Group A was provided with a diet identical to that recorded by Bawdy (2019), whereas Group B was provided with dietary supplements chosen by a

specialist veterinarian. Since the goal of this research project was to both control for ecological validity* within the primary research* project, and provide a possible explanation for the behaviours observed, a conceptual replication* method was chosen wherein the primary research observational* design was replaced with an experimental* approach.

Replication Research Scenario. After observing the mating behaviours of chimpanzees in their natural habitats, the same research team theorises that the specific mating behaviours observed are being influenced by the palm wine regularly consumed by the subjects.

Replication Research Goal. Control for Ecological Validity.

Replication Research Method. Conceptual.

Changed Primary Research Goal. Explain.

Changed Primary Research Method. Experimental.

6.3. Chapter Summary

The primary conclusions of this chapter are:

- A shared RDN requires extra effort but also provides common ground for primary researchers and replication researchers when designing projects and communicating efficiently with each other.
- A shared RDN can be leveraged to demonstrate appropriate transparency for non-experts with minimal expenditure of effort by the researcher.

In this chapter, I have provided two sections of examples demonstrating appropriate transparency for different intended audiences. In the first section, I highlighted the types of consideration from the RDN that can aid researchers in actualising appropriate transparency in their research design and communication. While intentionally not exhaustive, the examples I have provided are illustrative of some information useful for efficient communication with replication experts. In the second section, I demonstrated that appropriate transparency with non-experts can be achieved through identification of contentious terminology that is often the source of disagreement between individuals. By referring to key design concepts, and then referencing the definitions that best describe how the researcher is using said

concepts, researchers can begin to restore confidence in their research design without needing to justify the decisions to a distracting level of detail.

Chapter 7. Conclusion

In this chapter, I assume the following primary conclusions from the previous chapters:

- All researchers and their audiences do not necessarily use identical definitions of concepts related to psychological research (Chapter 1).
- It is useful to unpack pivotal concepts in psychological research to facilitate effective communication between researchers and their audience (Chapter 1).
- It is useful to interpret the replication crisis in psychological research as a crisis of confidence in research conclusions (Chapter 2).
- Transparency in research communication is a necessary step in restoring confidence in replication research conclusions (Chapter 2).
- Transparency is a pivotal term in psychological research that is conceived of differently by individual researchers, as well as their audience (Chapter 3).
- Appropriate transparency in any psychological research communication can only be judged in the context of the intended audience for that communication (Chapter 3).
- The lack of a shared RDN introduces unnecessary strain on replication research interpretation and extrapolation (Chapter 4).
- A shared RDN would aid researchers in identifying the appropriate level of transparency for their research communication audience (Chapter 4).
- An RDN should be as explicit as possible without being unnecessarily proscriptive (Chapter 5).
- An RDN is most useful when it retains sufficient flexibility for experts to apply practical judgement regarding appropriate transparency (Chapter 5).
- A shared RDN requires extra effort but also provides common ground for primary researchers and replication researchers when designing projects and communicating efficiently with each other (Chapter 6).
- A shared RDN can be leveraged to demonstrate appropriate transparency for non-experts with minimal expenditure of effort by the researcher (Chapter 6).

In this thesis, I have argued that when practically defined, transparency is a useful concept for combating the replication crisis as found in psychological research. It is common to see transparency proposed as a solution for the replication crisis, yet little

attention is given to the characteristics that can be used to identify whether transparency has been achieved. To this end, there is utility in providing researchers with clearer guidelines regarding how transparency can be practically actualised without becoming an unnecessary burden. One way to achieve this transparency is via identification of the categories of objects of research design. I support this position in the first five chapters in the following ways:

In Chapter 1, I outlined the three premises that orient my exploration of meaning attribution as it can be applied to the concept of transparency in psychological research: all human thought is abstract and no human thought can exhaustively represent the world (Whitehead); not all human thoughts attached to concepts share identical meaning attribution between individual humans (Chalmers and Frege); obscured differences in meaning attribution between individual humans is an observable issue within psychological research discourse (Abend).

In Chapter 2, I explored how the replication crisis in psychological research is better framed as a crisis of confidence in research conclusions that is contributed to by inappropriate inferences made because of replication failure. I identified transparency in research communication as a commonly proposed (yet ill-defined) solution for the replication crisis.

In Chapter 3, I defined appropriate transparency in research communication as a desirable state (virtue) contextually situated between opposing undesirable states (vices). Too little research design information can exemplify understated transparency (vice of deficiency) for some research communication audiences. Too much research design information can exemplify overstated transparency (vice of excess) for some research communication audiences.

In Chapter 4, I argued the utility of a shared RDN towards identifying the appropriate level of research design transparency in targeted research communication.

In Chapter 5, I synthesised an RDN from the discourse. I intended to provide a framework for researchers to leverage[264] when both designing research and identifying the appropriate level of transparency when communicating to targeted audiences.

[264] For or against.

In Chapter 6, I provided examples of some research design variables that would be contextually appropriate for communication to replication experts and one way in which relevant research design decisions can be communicated at a level appropriate for non-expert audiences.

In this chapter, I conclude my exploration of the replication crisis by highlighting the benefit of my interdisciplinary approach to the broader discourse. I do this in the following ways:

1. I discuss how different parts of the RDN interact to support appropriate transparency.
2. I propose the utility of my thesis in relation to both philosophy and research and discuss ways in which the RDN can be expanded upon further.

7.1. An À La Carte Menu of Definitions

Appropriate transparency can be achieved through the application of practical judgement by individual researchers to identify the definitions from the RDN that are necessary for their intended audience. Even with an RDN to identify the terms necessary for *appropriate transparency* in research communication, a universally applicable list of characteristics cannot be provided to researchers. As I have argued, much of the detail necessary for a replication researcher is undesirable for a non-expert reader. For example, as demonstrated by Doyen et al.'s priming replication,[265] it may be relevant for a replication expert to know the make and model of the stopwatch used by a primary researcher, but this information serves little purpose for a non-expert.

Appropriate transparency is a key foundation for the effective communication of research. As Stephen John contended, there is justification for believing that too much *transparency* may unnecessarily distract non-expert audiences.[266] However, if no information is provided to support the research design decisions of researchers, then little is being done to restore confidence that appropriate research design decisions are informing research conclusions. For example, if a replication attempt fails to align with the conclusions of the *primary research* upon which it was based, without transparent communication concerning the method of replication used, less

[265] Doyen et al., 'Behavioral Priming'
[266] John, 'Epistemic Trust'.

confidence can be placed in the replication's conclusion that the *primary research* conclusion was incorrect. Not all replication methodologies support the same inferences. Research design transparency in *primary research* is also a contributing factor to appropriate replication research design.

I found no dispute in the psychological research discourse concerning the benefit of transparency towards improving replication research. Even though John argued against transparency in research communication, that his arguments centre on non-expert misinterpretation of research norms does not suggest he argued against transparency for the benefit of replication.[267] Nevertheless, John's argument also ignores the implications of reduced transparency when communicating between experts. The RDN I promote is aimed at providing an example of the type of detail necessary for communication between original research teams (primary researchers) and replication research teams (replication researchers). However, there may be some objection that the nomenclature I provide is an example of my own definition for *overstated transparency*. For example, an objection to the synthesised RDN may claim that it is unnecessary to define the characteristics of psychological science as part of said nomenclature, and that this foundational level of definition is unnecessary and therefore overstated.[268]

It is not sufficient for practical science to be exclusively rational, empirical, falsifiable and critical, if it is not also replicable.[269] Replication is a critical characteristic of science, which implies the importance of replication concerns to all levels of science. If primary research is not designed to encourage replication, then it is not being designed to encourage appropriate science. If a researcher disagrees with this assessment, that replication is not a necessary characteristic of science, then that is an important orientation towards their thinking that is useful to be communicated to other researchers. A researcher may have an argument that they could present as to why they disagree with replication as a necessary component of their research project. However, if that argument is not explicitly addressed, then any research design decisions that were made because of this position will be obscured from the interpretation of their research. I have argued that as these characteristics inform the

[267] John.
[268] As such, I limited the detail to which foundational definitions of science were included in this thesis.
[269] Jackson, *Research Methods and Statistics*.

entire nomenclature, failure to make this orientation explicit when discussing the nomenclature, risks demonstrating *understated transparency*. Alternatively, if I had gone into much more detail on psychology as a science, then I concede that I would risk demonstrating *overstated transparency*. Instead, I hope I have achieved balance and demonstrated *appropriate transparency*.

7.2. Final and Future Considerations

My thesis has demonstrated one way in which philosophical foundations can be used to synthesise actionable scientific practice. As separately written about by Gabriel Abend and Sebastian De Haro, science and philosophy are often two avenues for exploring the same topic; however, there remains a tension between the two disciplines.[270] By beginning with a philosophical foundation upon which I justified practical advice for researchers to actualise the common (yet nebulous) instruction that replication is important, I demonstrated one way in which the concepts embedded within either philosophy or science can be leveraged to inform concepts within the other discipline.[271] Considerations of the meanings attributed to contentious terms is more traditionally philosophy than science. What variables are necessary to account for in research is more traditionally science than philosophy. However, I demonstrate that there is utility in synthesising the ideas from both disciplines. Despite Stephen Hawking's objections to the contrary, science has not killed philosophy; [272] rather, I have aligned my arguments with the position that misunderstanding the strengths of both the discipline of science and the discipline of philosophy unnecessarily detracts from the overall analysis of a given topic.

By creating an RDN synthesised from disparate avenues within psychological and philosophical discourse, I provide a single position for psychology researchers to either align with or argue against. While I have exercised all due care in producing a series of definitions that account for a significant part of the discourse around each of the terms I unpack, it would be unconscionable to insist that the definitions I have provided are infallible. This is obvious, and possibly, in pointing it out I am providing an example of *overstated transparency*. However, I propose that it is nevertheless a

[270] Abend, 'What Are Neural Correlates?'; De Haro, 'Science and Philosophy'.
[271] Telling researchers to be more transparent without first defining transparency provides no practical process for them to improve.
[272] Stephen Hawking and Leonard Mlodinow, *The Grand Design* (Westminster: Random House Publishing Group, 2010): 5.

useful orientation for considering the two ways in which the RDN is useful: the RDN is as useful for researchers who agree with it, just as it is for those who disagree with it.

If a researcher agrees with the definitions I have synthesised, then they have a single point of reference for a wide range of terms, the meaning of which is useful to convey to their intended audience. For example, researchers who agree with the definition of *direct replication* provided can focus their limited resources on performing their target research, rather than defending the way in which they use a given term.

If a researcher disagrees with one (or more) of the definitions I have synthesised, then they have a single point of reference against which they can defend their own use of the term. For example, researchers who disagree with the definition of *direct replication* provided can focus on defining the concept they associate with the term differently from the way I have synthesised it from the discourse. However, the utility of the nomenclature is dependent upon the acceptance of my argument that idiosyncratic researcher meaning is one of the factors related to transparency in research practice.

As I identify, previous researchers have argued for transparency in research as one avenue to account for the replication crisis in psychological research.[274] Further, they have argued that transparency in relation to research is an ill-defined instruction.[275] However, I am unaware of other arguments that frame transparency as an epistemological virtue with interrelated epistemological vices.

Strong arguments for and against transparency in research have been put forward.[276] Yet, I have argued that both sides focus their premises so narrowly as to detract from the broader implications of their arguments. With the idea to represent both sides of the debate for transparency in the context within which each argument is appropriate, my thesis uses a vice/virtue/vice trichotomy as a bridge between epistemology and research practice.

Rather than simply positioning transparency as a desirable state for researchers to orient towards, I have identified the utility of considering transparency as a

[274] Figueiredo Filho et al., 'Seven Reasons Why'; Rosen, *Research Transparency and Replication at MDRC*.
[275] Elliott, 'A Taxonomy of Transparency'.
[276] Figueiredo Filho et al., 'Seven Reasons Why'; Rosen, *Research Transparency and Replication at MDRC*; Elliott; John, 'Epistemic Trust'.

contextually judged point on a spectrum between states of excess and deficiency. I consider the circumstances that justify each position differently from the circumstances that justify the opposite position. Under this interpretation of virtue epistemology, understanding the *vice of deficiency* and the *vice of excess* is as equally important as understanding the *virtue*. To the best of my knowledge, this is not a concept discussed in relation to psychological research. In my exploration of the discourse, transparency in relation to psychological research was either argued for or argued against; I am unaware of any attempts to position the desirability of transparency in research as dependent upon the intended audience for the research communication. In addition, the implications of my arguments propose certain avenues for future research.

The replication crisis in psychological research (as with every other field of research) is a problem that cannot be improved with a single response. Each improvement to research practices and communication is one more step towards rebuilding confidence in research conclusions. Once enough steps are taken, the label of crisis should no longer be applicable. To this end, I propose that the RDN I have synthesised, and any other communal research design terminologies, need to be subjected to descriptive, predictive and explanatory research.

Although I have positioned this thesis as an interdisciplinary approach to the topic that leverages both philosophical and scientific concepts, it could also be positioned as metaresearch. That is, an RDN could serve as the foundation for using scientific methodologies to study research practices. With an RDN serving to justify the rational characteristic of Jackson's definition of valid research, I propose that the next step would be one or more studies that critically analyse the empirical and falsifiable inferences stemming from these definitions. Then, the primary research should be subjected to vigorous replication attempts. I hold that synthesised definitions, such as those presented here, are a reasonable starting point, but ideally, an RDN should also be subjected to evidence-based analysis, and hence itself become part of the practice of science.

Bibliography

Abend, Gabriel. "What Are Neural Correlates Neural Correlates Of?" *BioSocieties* 12, no. 3 (2017): 415–38. https://doi.org/10.1057/s41292-016-0019-y.

Almassi, Ben. "Climate Change, Epistemic Trust, and Expert Trustworthiness". *Ethics and the Environment* 17, no. 2 (2012): 29–49. https://doi.org/10.2979/ethicsenviro.17.2.29.

Bakker, M. and J.M. Wicherts. "The (Mis)reporting of Statistical Results in Psychology Journals". *Behavior Research Methods* 43, no. 3 (2011): 666–78. https://doi.org/10.3758/s13428-011-0089-5.

Bargh, John A., Mark Chen and Lara Burrows. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action". *Journal of Personality and Social Psychology* 71, no. 2 (1996): 230–44. https://doi.org/10.1037/0022-3514.71.2.230.

Boyd, Craig A, and Kevin Timpe. *The Virtues: A Very Short Introduction*. Very Short Introductions. Oxford: Oxford University Press, 2021.

Chalmers, Alan. *What Is This Thing Called Science?* St Lucia: University of Queensland Press, 2013.

Chalmers, David J. "Verbal Disputes". *The Philosophical Review* 120, no. 4 (2011): 515–66. https://doi.org/10.1215/00318108-1334478.

Collins, Harry. *Changing Order: Replication and Induction in Scientific Practice*. University of Chicago Press, 1992.

Cooper, Rachel. "On Deciding to Have a Lobotomy: Either Lobotomies Were Justified or Decisions Under Risk Should Not Always Seek to Maximise Expected Utility". *Medicine, Health Care and Philosophy* 17, no. 1 (2014): 143–54. https://doi.org/10.1007/s11019-013-9519-8.

Crüwell, Sophia, Johnny van Doorn, Alexander Etz, Matthew C. Makel, Hannah Moshontz, Jesse C. Niebaum, Amy Orben, Sam Parsons and Michael Schulte-Mecklenbeck. "Seven Easy Steps to Open Science: An Annotated Reading List". *Zeitschrift Für Psychologie* 227, no. 4 (2019): 237–48. https://doi.org/10.1027/2151-2604/a000387.

De Haro, Sebastian. "Science and Philosophy: A Love-Hate Relationship". *Foundations of Science* 25, no. 2 (2020): 297–314. https://doi.org/10.1007/s10699-019-09619-2.

Doyen, Stéphane, Olivier Klein, Cora-Lise Pichon and Axel Cleeremans. "Behavioral Priming: It's All in the Mind, but Whose Mind?" *PloS One* 7, no. 1 (2012): e29081. https://doi.org/10.1371/journal.pone.0029081.

Elliott, Kevin C. "A Taxonomy of Transparency in Science". *Canadian Journal of Philosophy*, 2020, 1–14. https://doi.org/10.1017/can.2020.21.

Fidler, Fiona, and John Wilcox. 'Reproducibility of Scientific Results'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/.

Figueiredo Filho, Dalson, Rodrigo Lins, Amanda Domingos, Nicole Janz and Lucas Silva. "Seven Reasons Why: A User's Guide to Transparency and Reproducibility". *Brazilian Political Science Review* 13, no. 2 (2019). https://doi.org/10.1590/1981-3821201900020001.

Frege, Gottlob. "Sense and Reference". *The Philosophical Review* 57, no. 3 (1948): 209–230.

Hawking, Stephen and Leonard Mlodinow. *The Grand Design.* Westminster: Random House Publishing Group, 2010.

Hendrick, Clyde. "Replications, Strict Replications, and Conceptual Replications: Are They Important?" *Journal of Social Behavior and Personality* 5, no. 4 (1990): 41.

Hunter, Philip. "The Reproducibility "Crisis"". *EMBO Reports* 18, no. 9 (2017): 1493–6. https://doi.org/10.15252/embr.201744876.

Ioannidis, John P. A. "Why Most Published Research Findings Are False". *PLoS Medicine* 2, no. 8 (2005). https://doi.org/10.1371/journal.pmed.0020124.

Ioannidis, John P. A. "Why Science Is Not Necessarily Self-Correcting". *Perspectives on Psychological Science* 7, no. 6 (2012): 645–54. https://doi.org/10.1177/1745691612464056.

Jackson, Sherri L. *Research Methods and Statistics: A Critical Thinking Approach*, 5th ed. Boston, Massachusetts: Cengage Learning, 2016.

John, Leslie K, George Loewenstein and Drazen Prelec. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling". *Psychological Science* 23, no. 5 (2012): 524–32. https://doi.org/10.1177/0956797611430953.

John, Stephen. “Epistemic Trust and the Ethics of Science Communication: Against Transparency, Openness, Sincerity and Honesty”. *Social Epistemology* 32, no. 2 (2018): 75–87. https://doi.org/10.1080/02691728.2017.1410864.

Kahneman, Daniel. “A Proposal to Deal with Questions About Priming Effects”. September 26, 2012. Letter. From *Nature*. Accessed 22 June 2021. https://www.nature.com/news/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf.

Korzybski, Alfred. *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. New York: Institute of General Semantics, 1958.

Leshner, Alan I. “Addiction Is a Brain Disease, and It Matters”. *Science* 278, no. 5335 (1997): 45–7. https://doi.org/10.1126/science.278.5335.45.

Levy, Neil. “Addiction Is Not a Brain Disease (and It Matters)”. *Frontiers in Psychiatry* 4 (2013): 24. https://doi.org/10.3389/fpsyt.2013.00024.

Lilienfeld, Scott O. “Psychology’s Replication Crisis and the Grant Culture: Righting the Ship”. *Perspectives on Psychological Science* 12, no. 4 (2017): 660–4. https://doi.org/10.1177/1745691616687745.

Machery, Edouard. ‘What Is a Replication?’ *Philosophy of Science* 87, no. 4 (October 2020): 545–67. https://doi.org/10.1086/709701.

Nosek, Brian A., and Timothy M. Errington. ‘What Is Replication?’ *PLOS Biology* 18, no. 3 (27 March 2020): e3000691. https://doi.org/10.1371/journal.pbio.3000691.

Open Science Collaboration. “Estimating the Reproducibility of Psychological Science”. *Science* 349, no. 6251 (2015). https://doi.org/10.1126/science.aac4716.

Pashler, Harold and Christine R. Harris. “Is the Replicability Crisis Overblown? Three Arguments Examined”. *Perspectives on Psychological Science* 7, no. 6 (2012): 531–6. https://doi.org/10.1177/1745691612463401.

Pashler, Harold and Eric–Jan Wagenmakers. “Editors’ Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?” *Perspectives on Psychological Science* 7, no. 6 (2012): 528–30. https://doi.org/10.1177/1745691612465253.

Porte, Graeme. “Who Needs Replication?” *CALICO Journal* 30, no. 1 (2013): 10–5. https://doi.org/10.11139/cj.30.1.10-15.

PsychFileDrawer. 2012. http://www.psychfiledrawer.org/.

Rosen, Rachel. *"*Research Transparency and Replication at MDRC". *MDRC*. January, 2018. Accessed June 22, 2021. https://www.mdrc.org/publication/research-transparency-and-replication-mdrc.

Schmidt, Stefan. "Shall We Really Do It Again? The Powerful Concept of Replication Is Neglected in the Social Sciences". *Review of General Psychology* 13, no. 2 (2009): 90–100. https://doi.org/10.1037/a0015108.

Simons, Daniel J. "The Value of Direct Replication". *Perspectives on Psychological Science* 9, no. 1 (2014): 76–80. https://doi.org/10.1177/1745691613514755.

Stroebe, Wolfgang and Fritz Strack. "The Alleged Crisis and the Illusion of Exact Replication". *Perspectives on Psychological Science* 9, no. 1 (2014): 59–71. https://doi.org/10.1177/1745691613514450.

Stroebe, Wolfgang, Tom Postmes and Russell Spears. "Scientific Misconduct and the Myth of Self-Correction in Science". *Perspectives on Psychological Science* 7, no. 6 (2012): 670–88. https://doi.org/10.1177/1745691612460687.

Vazire, Simine, and Alex O Holcombe. 'Where Are the Self-Correcting Mechanisms in Science?' *Review of General Psychology*, 2020.

Vyse, Stuart. "P-Hacker Confessions: Daryl Bem and Me". *The Skeptical Inquirer* 41, no. 5 (2017): 25.

Whitehead, Alfred North. *Science and the Modern World*. New York: Macmillan, 1954.

Wicherts, J. M. "Psychology Must Learn a Lesson from Fraud Case". *Nature (London)* 480, no. 7375 (2011): 7. https://doi.org/10.1038/480007a.