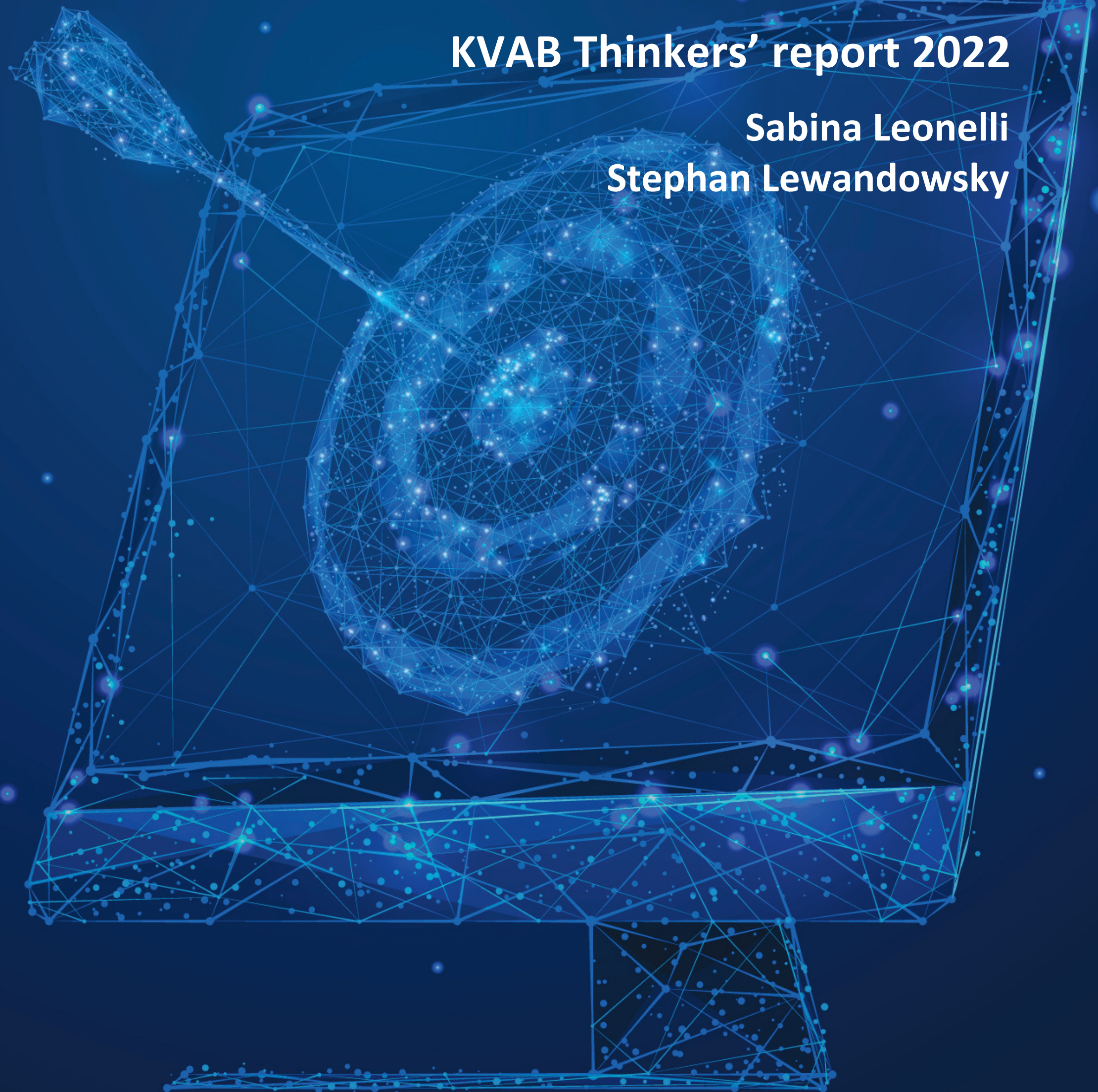
THE REPRODUCIBILITY OF RESEARCH IN FLANDERS: FACT FINDING AND RECOMMENDATIONS

KVAB Thinkers' report 2022

Sabina Leonelli
Stephan Lewandowsky



KVAB POSITION PAPERS
81 b
Royal Flemish Academy of Belgium
for Science and the Arts - 2022

THE REPRODUCIBILITY OF RESEARCH IN FLANDERS: FACT FINDING AND RECOMMENDATIONS

KVAB THINKERS' REPORT 2022

KVAB Press

KVAB POSITION PAPERS
81 b

Cover concept: Francis Strauven
Cover design: Charlotte Dua
Image: Shutterstock

The drawing of the Palace of the Academies is a reproduction of the original perspective made by Charles Vander Straeten in 1823. The logo of the KVAB was designed in 1947 by Jozef Cantré.

The KVAB Position Papers (Standpunten) are published by the Royal Flemish Academy of Belgium for Science and the Arts, Hertogsstraat 1, 1000 Brussel.
Tel. 00 32 2 550 23 23 – info@kvab.be – www.kvab.be

THE REPRODUCIBILITY OF RESEARCH IN FLANDERS: FACT FINDING AND RECOMMENDATIONS

KVAB THINKERS' REPORT 2022

Sabina Leonelli
Stephan Lewandowsky

D/2023/0455/02
ISBN 978 90 656 921 91

Recommended citation: Sabina Leonelli, Stephan Lewandowsky, *The reproducibility of research in Flanders: Fact finding and recommendations - KVAB Thinkers' report 2022*, KVAB Standpunt 81, 2022.

Printing office Universa

THE REPRODUCIBILITY OF RESEARCH IN FLANDERS: FACT FINDING AND RECOMMENDATIONS

KVAB THINKERS' REPORT 2022

TABLE OF CONTENTS

Introduction and Overview 11

Methods 13

PART 1. The Many Faces of Reproducibility 15
- BOX 1: Reproducibility May Not Require Replication 16
- Six forms of reproducibility 18
- Narrow and broad interpretations of reproducibility 20
- BOX 2: Pre-Registration 23

PART 2. Implementing reproducibility within academic institutions........ 27
- Implementation practices 27
- Tensions among stakeholders 29

PART 3. The Tension Between Resilience and Reproducibility........ 32
- BOX 3. Science communication 32
- BOX 4: Belfast Project and implications 36
- BOX 5 - Reproducibility and Transparency: Key Challenges 37

PART 4. The Tension Between Public and Private Interest 40
- BOX 6: Research Institutes at the Translational Cutting Edge 41

PART 5. Recommendations 42

Conclusions: How Does the Pursuit of ReproducibilityHelp Us Address the Science Crisis? 47

References 49

Executive summary

The report is produced in the context of a Thinkers' program of KVAB. It explores three aspects of reproducibility with the goal of providing recommendations to stakeholders about how to ensure reproducibility. Replication perse is only one aspect of reproducibility, and to thoroughly understand reproducibility requires casting a wider net that includes components such as transparency, research practices, and the role of theory in science. Second, the relationship between these components and reproducibility can be complex and counterintuitive. Third for science to yield robust, reproducible and credible knowledge, we must transform research institutions, evaluation and practices so that, ultimately, it is no longer necessary to talk about it. We will have achieved reproducibility when we no longer debate it. Reproducibility matters, but it takes different forms suited to different research situations, depending on the specific goals, methods, materials, and conditions of research. Given this necessary variation, reproducibility needs to be conceptualised and implemented in ways that are not too narrow, thereby inadvertently constraining research in ways that are unhelpful, nor too broad to be meaningful.

Extensive consultations with stakeholders provided the Thinkers with insights about current practices and possible futures in Flemish research, more specifically, Universities, Research Centers, Umbrella organizations, Pharmaceutical industry and Funders.

Positive actions and attention points in research institutes

- A recognition of the role of data stewards as integral parts of the research ecosystem, with many universities investing in permanent data steward positions to offer support in data collection, curation and management.
- A heightened awareness of the significance of integrity within research, resulting in a boost of ethics boards and integrity officers at various universities.
- Novel training resources for researchers and professional support staff to learn how to make their work more responsible and robust methodologically, BUT still a lack of recognition for the expertise and effort involved.

Findings for the funders

- A willingness to take account of integrity and Open Science practices by funders such as the FWO, where negative results are currently recognised as a legitimate output for research assessments, label of such considerations as "second axis" for research: BUT can be interpreted as a confirmation that questions of integrity, quality and reproducibility are not a priority.
- Difficulties to assess transdisciplinary research, which many institutions regard as paramount to confront global challenges in theory, but in practice remains very hard to assess.

- The recognition of Open Science as an important phenomenon and attempts to instigate a culture change within academic institutions. Such attempts can take substantive forms, such as modification of criteria for hiring or promoting academics so that they include efforts in data curation, reproduction of other studies and quality assessment.

Tensions among stakeholders

- Tension between publicly funded researchers, mainly in universities, who are expected to be fully transparent on the one hand, and researchers working in industry on the other, who receive scrutiny from regulatory bodies but are not required to undergo scrutiny by academic peers.
- Tensions between researchers at different levels of seniority, mainly because the additional cost and effort involved in conducting research to a higher criterion of reproducibility is borne by junior researchers who are more likely to have acquired relevant skills.
- Key perceived problem in the pursuit of reproducibility the institutional reward and incentive structures are lagging behind and continue to revolve around conventional metrics.
- Even when institutions complement performance metrics with qualitative evaluation, which provides rewards for reproducibility efforts, the emphasis continues to be on novelty over replication or quality.
- Reviewing is a labour-intensive activity but is voluntary and almost completely unrewarded.

Science communication

- Misconception of the so-called "deficit model"; i.e. the public is lacking knowledge that science communicators can provide, thus remedying the public's information deficit.
- For publics to be able to assess the value of scientific claims, it is imperative to understand science as a self-correcting process of inquiry, rather than as a set of infallible truths.
- View of scientific communication as 'information provision' does not provide any avenue for non-researchers to enter in a dialogue and to contribute their own expertise (citizen science).
- Majority of public challenges to data by regulated industries, lobbyists, and trade organizations.
- The blanket use of "open data" when research involves sensitive information is problematic and in conflict with the goals of diversity and inclusivity.
- The trade-offs between transparency and privacy and between reproducibility, resilience, trust, and respect for privacy, must be examined frankly and thoroughly.

– the stakes are higher for commercially exposed research, and particularly research devoted to transforming scientific insights into commodities for global markets, characterized by huge pressure to reproduce results and use very precise and standardised quality checks.
– Both publicly and privately funded researchers are exposed to incentives and demands towards producing outputs as fast as possible, with little time devoted to exploring implications and to documenting and properly scrutinizing research processes.
– For privately funded researchers there are severe obstacles to the free sharing - and wide-ranging scrutiny - of data.
– For publicly funded research there is a tendency to privilege secrecy over open collaboration, especially in crowded domains such as biomedicine; and to overly rely on automated systems
– For Research Institutes at the Translational Cutting Edge there is no institutional body in charge of overseeing whether their transnational research carried is socially responsible.

Preface

The Academy's Standpunten Series (Position Papers) contributes to a scientifically validated debate on current social and artistic topics. The authors, members and workgroups of the Academy write under their own name, independently and in full intellectual freedom. The quality of the published studies is guaranteed by the approval of one or several of the Academy's classes. This position paper was approved for publication by the meetings of the Class of Technical Sciences on 17 September 2022 and the Class of Humanities on 19 November 2022."

The organization of the Thinkers' program 2022 by KVAB in cooperation with the KAGB Royal Academy for Medicine of Belgium and the Flemish Young Academy

Patrick Onghena (KMW), Joos Vandewalle (KTW) and Inez Dua (KVAB staff)

Each year KVAB organizes two Thinkers' programs at the initiative of one of its classes and/or reflection groups. The program "Reproducibility and replicability in research" was proposed by the permanent reflection group on Responsible Research and Innovation (RRI) and science ethics at its online meeting in March 2021 in a full pandemic situation.

Context: In a typical Thinkers' program a central role is assigned to one or two leading international experts. The proposed theme of "Reproducibility and replicability in research" is linked to a well-received KVAB position paper "Replicability in the empirical social, behavioral, and educational sciences" (in Dutch) written by one of our academy members, Patrick Onghena, and based on his 2018 academy presentation. Moreover many reports from academies have recently been published on this important theme: the Dutch Academy KNAW, Royal Netherlands Academy of Arts and Sciences (2018), "Replication studies: Improving reproducibility in the empirical sciences", the US Academies "Reproducibility and Replicability in Science" (2019), UK The Academy of Medical Sciences (2015) "Reproducibility and reliability of biomedical research: improving research practice" London, UK: The Academy of Medical Sciences, and the EU overview report "Reproducibility of scientific results in the EU Scoping report" November 2020.

Theme: The proposed theme "Reproducibility and replicability in research" is quite broad, but it is certainly intended for the entire collection of scientific disciplines, each with their own specificity. On October 18, 2018, Nature published a special issue about Challenges in irreproducible research: "Science moves forward by corroboration – when researchers verify others' results. Science advances faster when people waste less time pursuing false leads. No research paper can ever be considered to be the final word, but there are too many that do not stand up to further study."

Ultimately, the confidence of society and the wider public in science and the societal value of scientific results is partly based on the reproducibility and replicability of the results. The problem presents itself in various forms in the various scientific domains. Transparency and openness of the data, the methods of processing, better statistical processing and openness of the software can certainly contribute to this. In addition, successful approaches from hard-hit domains such as psychology can inspire other domains. In order to limit the scope the steering group had to discuss a focus in interaction with the Thinkers. Central themes are certainly the 'approach to more transparency, improvement of the quality of research, good practices, ethical questions and public trust in research' and not so much the repressive methods of poor or hardly replicable research.

Activities: The proposal was accepted by the Board of Directors of KVAB in May 2021, and a starting note was worked out in July 2021. Then the cooperation with the KAGB and the Belgian Young Academy was secured. In July 2021 a Steering Group was composed with members of the KVAB, the Young Academy, the KAGB (Royal Academy for Medicine of Belgium) and several external experts (see Annex 3). The role of the Steering Committee was to ensure proper underpinning of the Thinkers' activities and to provide the necessary input. So, no direct steering was expected, but rather support. The two Thinkers were selected in the Summer of 2021: *Sabina Leonelli*, University of Exeter, Turing Institute, Global Young Academy, Open Science Group, philosopher of science and *Stephan Lewandowsky*, University of Bristol, Chair in Cognitive Psychology (see Annex 2 for their cv). The Thinkers were given ample freedom and remain completely independent in writing their report with recommendations. It was the intention that the Thinkers work together with the steering group and numerous partners and stakeholders on the chosen theme and make a significant contribution to the further scientific strategy of Flanders by developing a long-term vision and in this way contributing to policy making. Through numerous activities concrete observations, ideas, and recommendations can be developed.

The steering group had a first online meeting on October 15, 2021 followed by an online kick-off meeting with the Thinkers on November 29 2021, during which the further planning of the Stakeholder meetings in open or closed workshops and visits was prepared. Relevant questions were prepared by the Thinkers and shared with the Steering Group. These were given in advance to the stakeholders and discussed with the Thinkers in the Spring of 2022 in several online and physical meetings. A well-attended closing symposium was held June, 1 2022, with the findings of the Thinkers. The presentations of the Thinkers have been made accessible on YouTube.
https://kvab.be/nl/activiteiten/final-symposium-reproducibility-and-replicability-science
During the symposium the public and other speakers had several interesting interactions with the Thinkers. Based on these two-way communications the Thinkers have drafted their report.

It is a pleasure to thank all those who made this program successful, the Steering Group, the stakeholders, the public and the KVAB staff. In particular we would like to thank the two devoted and active Thinkers, Sabina and Stephan for an excellent report. They succeeded in bringing together and integrating the information from the initial interviews and the discussions during the June symposium in a thought-provoking way. It is an excellent basis for the Flemish scientific community to expand the discussion and keep the focus on improving and defending science.

Introduction and Overview

Reproducibility is fundamental to science as a way to verify the credibility of results and procedures. Findings that cannot be replicated and scientific procedures that cannot be reproduced may fail to contribute to knowledge and, at worst, waste researchers' time when they pursue a blind alley based on unreliable results. As a *Nature* editorial put it on October 18, 2018: "Science moves forward by corroboration – when researchers verify others' results." Reproducibility has also long been regarded as a pillar of the demarcation between science and pseudoscience. Horoscopes, homoeopathy, the hollow earth theory and other pseudoscientific claims do not offer any reproducible insights but are, at best, offering entertaining narratives (Zwaan, Etz, Lucas, & Donnellan, 2017). Moreover, the confidence of society, policy makers, and the wider public in science and scientific results is linked to the perceived reproducibility and replicability of the results. Scientific advice may legitimately change with emerging knowledge, especially during a rapidly-evolving crisis such as the onset of the COVID-19 pandemic, and it is important that the public understands that scientific knowledge comes with uncertainty bounds (Joslyn & LeClerc, 2016); nevertheless, scientific knowledge is only trustworthy insofar as it is the outcome of reliable and well-scrutinised processes of inquiry, and reproducibility is an important criterion to evaluate the quality and legitimacy of scientific claims.

It must therefore be of key concern that numerous experimental findings in a variety of disciplines have recently turned out not to be reproducible. In psychology, a large-scale effort to replicate 100 published findings found that the effect sizes in the replications were, on average, only half those of the original publications and only 25% and 50% of findings replicated, respectively, in social psychology and cognitive psychology (Open Science Collaboration, 2015). Although much of the discussion surrounding reproducibility has focused on psychology, numerous other disciplines have encountered similar problems. For example, in a parallel attempt to establish replicability in economics, less than half of the results of 60 published articles were found to be reproducible (Chang & Li, 2015); and attempts to independently verify published biomedical findings found significant errors in the data reporting and analysis (Allison et al., 2016). Moreover, scientific publications often do not contain sufficient detail (about methods, data and codes, for instance) to support attempts to reproduce the results, making it impossible to verify the reliability of those studies in this way (National Academies of Sciences, Engineering, and Medicine, 2019).

The emergence of the "replication crisis" has triggered intense discussion both within the scientific community and among the public and stakeholders. A wide diversity of viewpoints has emerged, and although the issue is far from settled at the time of this writing, three consensual positions appear to have emerged:

1. Replication per se is only one aspect of reproducibility, and to thoroughly understand reproducibility requires casting a wider net that includes components such as transparency, research practices, and the role of theory in science.
2. The relationship between these components and reproducibility can be complex and counterintuitive.
3. In order for science to yield robust, reproducible and credible knowledge, we must transform research institutions, evaluation and practices so that, ultimately, it is no longer necessary to talk about reproducibility. We will have achieved reproducibility when we no longer debate it.

The remainder of this report explores those three aspects of reproducibility with the goal of providing recommendations to stakeholders about how to ensure reproducibility. Based on the brief we received, our intent was to generate a broad set of recommendations that are relevant across disciplines and open the door to further discussion rather than prescribe specific actions.

Methods

The report was informed by several rounds of intensive meetings with stakeholders around Flanders. The following stakeholders contributed to these discussions:

- KU Leuven
- University of Gent
- VUB
- University of Antwerp
- University of Hasselt
- Research Centers (representatives from Imec, VITO, Flanders Make, VLIZ, and ILVO)
- Umbrella organizations (Young Academy, ReproducibiliTea journal club, Open Science Belgium)
- Pharmaceutical industry (representatives from Janssen Pharmaceutica, part of Johnson & Johnson, and GSK)
- Funders (representatives from FWO, VLAIO)

Although the discussions with stakeholders were flexible and open-ended, they were seeded with a list of questions circulated ahead of each meeting. The following questions were presented to stakeholders:

- How do you interpret the notions of reproducibility and replicability? Do you see them as relevant to current practices in your organization/projects (if yes, how; if not, why not)?
- Is reproducibility always a necessary ingredient of reliable and robust knowledge production?
- What is the relation between reproducibility and confirmation bias (if any)?
- What do you think about the relation between reproducibility and Open Science?
- What institutional structures and/or incentives are in place to support reproducibility?
- Do you foresee any negative consequences from implementing these structures?
- Can you point to good practices in reproducibility and/or role models (both at the level of individuals and institutions)?
- What discipline(s) do you work with? Do you use qualitative, quantitative or mixed methods?
- What, if any, is the role of theory in creating reproducible research?
- Do communications with the public and the press take into account reproducibility issues?
- How do we trade off reproducibility / replicability against timeliness in a crisis situation, such as the current pandemic?
- Is there a need for special training for young researchers, PhD students, Master students?
- What incentives do you need for such training to be practicable?

- Is disciplinary research more or less reproducible than interdisciplinary research, and why?
- Can you describe one or more concrete scenarios that exemplify your thinking on this?
- Do you think artificial intelligence (AI) could support or hinder reproducibility efforts, and if so, how?
- Can you give examples of famous findings in your field of study that later turned out to be nonreproducible or nonreplicable?

The stakeholder discussions were extensively minuted and discussed by the authors, and the input is reflected throughout this report. Where appropriate, we highlight specific concerns raised by stakeholders throughout the report.

PART 1. The Many Faces of Reproducibility

Reproducibility can be broadly defined as an overarching scientific value, often evoked in discussions of what constitutes 'best practice' or 'valid research', which denotes the extent to which consistent results are obtained when a piece of research is repeated. The idea of reproducibility is closely associated with debates around Open Science, due to the common emphasis on the sharing of information that may be used to verify whether or not a piece of research is reproducible.

A key challenge immediately emerging in discussions of reproducibility across all groups of stakeholders is that of diversity. Science is a highly fragmented enterprise, with countless subfields deploying methods and tools that differ considerably from each other, since each is finely tailored to specific materials, problems, and goals. Key approaches to research may include qualitative as well as quantitative methods; experimental approaches as well as observational studies and computer simulations; and theory-oriented as well as data-intensive modes of investigation. This scientific diversity is enhanced by various factors including: the characteristics of the entities that researchers investigate, with studies of human populations differing significantly from studies of biological samples or nano particles; the differences among research cultures, resources and objectives underpinning inquiry at different locations; the incentives and priorities set by institutions that support research; and the division of labour adopted in different scientific settings, with specific and varying responsibilities assigned to early career researchers, senior academics, support staff, librarians, technicians, data scientists and data stewards. As philosopher Hans Radder noted already in 1996, given such overwhelming diversity, "to clarify the notion of reproducibility we need to address the following question: reproducibility of what and by whom?" (Radder, 1996, p. 16).

It is generally recognised that there is a difference between efforts to reproduce research processes and efforts to reproduce the outcomes of such processes. The U.S. National Academy of Sciences (NAS) has recently referred to this difference by distinguishing between replicability and reproducibility, with the former defined as "obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data" (National Academies of Sciences, Engineering, and Medicine, 2019), and the latter as "obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis" (National Academies of Sciences, Engineering, and Medicine, 2019; see also commentary in Onghena 2020). While acknowledging the value of this basic distinction, in what follows we point to a more complex and diversified way in which scientific domains view the idea of reproducibility (see Box 1).

BOX 1: Reproducibility May Not Require Replication

Concern about reproducibility and robust science is usually tightly coupled with the presumed need for replication. “Making replication mainstream” (Zwaan et al., 2017) is a stated goal of many scholars who are concerned about reproducibility and the current “replication crisis” that has embroiled many disciplines. However, there has been relatively little exploration of the circumstances under which replications are or are not advisable. Should every study be replicated? If so, by whom? If a decision about replication is to be made, what should drive it? Should we always replicate before we test a new prediction? It turns out that the devil is very much in the details when it comes to decisions about replication. Sometimes those details are strikingly counterintuitive. We consider two recent examples.

The first example addressed the question whether researchers should always replicate a finding before they publish it (Lewandowsky & Oberauer, 2020). To answer the question, Lewandowsky and Oberauer modeled the aggregate performance of a hypothetical scientific community, focusing on the costs and benefits of knowledge generation under different replication regimes. The take-home message of that modeling study was that the publication of potentially non-replicable studies minimizes cost and maximizes efficiency of knowledge gain for the scientific community overall. This counterintuitive conclusion was the inevitable consequence of one crucial assumption; namely, that the scientific community does not find all published results to be of equal interest. Scientists publish their findings because they find them interesting and exciting, and they publish in the hope that the scientific community will share that excitement and will build on the published findings with more interesting research. In reality, those hopes and expectations are rarely fulfilled: most published papers disappear without a trace and only a few findings strike a sweet spot and everyone starts talking about them—at conferences, on Twitter, or by citing the findings in the literature.

The moment one realizes this lopsided distribution of interest in the community, the question of who should replicate findings, and when, yields counterintuitive answers. The model of Lewandowsky and Oberauer compared two replication regimes: Under the “private” regime, all findings are replicated by the author before publication to guard against subsequent replication failures. Findings that did not replicate were never published. Under the alternative “public” regime, all significant findings were published, even though their replicability was uncertain. Attempts to replicate those published findings were limited to those (few) phenomena the scientific community considered to be interesting. One pervasive finding emerged from comparison of those two regimes: The private regime incurred a considerably greater cost, in terms of total number of experiments conducted, than the public regime to uncover the same amount of reproducible new knowledge. The greater cost arose because many pre-publication replications involved findings that other scientists ultimately did not care about. And every wasted replication is wasted effort that could have been put towards a more productive purpose. Low replicability of published findings may therefore be beneficial for scientific productivity overall because less effort is wasted on irrelevant replications.

The second example involves the interplay between data and theory. If a finding has been successfully predicted by a theory, should it be replicated to further enhance our confidence in the theory, or would we be better off testing a new prediction by the theory? The answer is again nuanced and depends on how tightly coupled the predictions are to a theory (Oberauer & Lewandowsky, 2019). When theories are highly constrained and specified in great detail (e.g., in a computer simulation), then it turns out that replication of a successfully-predicted finding is less informative than a test of a novel prediction. Using the mathematics of Bayesian inference, Oberauer and Lewandowsky showed that our confidence in a theory should increase more after a successful test of a second, different prediction than after replication of the first successful test. The reverse was true for less specific theories whose predictions were not tightly coupled to the axioms of the theory–as is the case in most verbally formulated theories. For the discovery-oriented research that is typical for those theories, replications enhance our confidence more than tests of new predictions.

These two examples illustrate that there is no simple one-to-one mapping between reproducibility and robust science on the one hand, and replication on the other. Solving the replication crisis requires a cultural shift and requires that well-targeted replications become mainstream. But robotic replication before and after publication is not the best solution.

It is not enough to differentiate between the computational idea of reproducing every single part of a study, including its outcomes, and the emphasis on reproducing results, while admitting variations in methods, more popular within the experimental sciences. Across research domains, and often even within them, there are different assumptions made around: whether and how much researchers can control research conditions; what variation may indicate about the phenomena being investigated, which can be an error or an attempt at fraud but also a novel discovery or an artefact of the research design; which methods are best suited to infer knowledge from data, and particularly the degree to which the research depends on statistical and computational tools (which are sometimes simply not suited to the goals at hand, as in the case of interpreting qualitative data); which research goals are most important and to which extent they can be specified in advance of actually carrying out the research; what constitutes a robust evidence base, and whether there are established criteria to make that assessment; and whether or not researchers' subjective judgement can be trustworthy, particularly in situations where the research cannot be highly standardised (see Table 1).

Table 1. Factors affecting researchers' understandings of reproducibility

Assumed degree of control over research conditions
Understanding of variation
Dependence on statistics and computation as inferential tools
Precision of the research goals
Stability of background knowledge and evidence base
Dependence on researchers' expert judgement

Given this diversity, we propose a taxonomy of six forms of reproducibility that is closely associated to the meaning that reproducibility takes under different research conditions (see also Leonelli, 2018).

Six forms of reproducibility

1. Computational reproducibility

Most commonly found in computational research and digital experiments, computational reproducibility focuses on running the same data through a given set of algorithms over again. Discrepancies among the outcomes are taken as a sign that there are problems in the computational apparatus used to analyse data, and are thus used to identify and resolve mistakes and bugs in data analysis and programming. A key precondition for this approach is the availability of reusable code and data, through which analysis can be repeated.

2. Direct experimental reproducibility

Most commonly found in highly controlled experiments aiming to test a well-specified hypothesis, such as for instance clinical trials, direct experimental reproducibility consists of the ability to obtain the same results through the repeated application of the same research methods/processes. A key precondition for this approach is the presence and public availability of standardised protocols, materials and settings, which have been carefully developed and calibrated to yield very specific outcomes with great precision, and through which the right degree of experimental control can be achieved.

Indirect Reproducibility

When it is acknowledged that considerable parts of an experimental set-up elude the control of experimenters (a situation we will refer to as "semi-standardised experiments"), researchers may appeal to indirect reproducibility: that is, verifying whether results obtained from performing different experiments are sufficiently similar. Comparing data produced on the same phenomenon and/or hypothesis by different experimental set-ups is often seen as a useful validation tool to see whether results produced under variable circumstances converge or not, thereby helping to verify the robustness of the given findings (Radder, 1996; Derksen & Morawski, 2022). This approach to reproducibility is typical of exploratory research, where experimental settings cannot be highly standardised since researchers are still defining the methods and/or the targets of their study. It also characterises studies of phenomena that change all the time in response to their wider environment, such as evolving organisms or social behaviours. Examples include pre-clinical research (Lowe, Leonelli, & Davies, 2019), exploratory experiments on model organisms (Ankeny & Leonelli, 2020) or studies of developmental processes (Love, 2020; Weber, 2022), and some psychological experiments whose broader context is specific to a given place or historical period (Felt, 2019; Derksen & Morawski, 2022).

4. Scoping reproducibility

Again in reference to semi-standardised experiments, a different take on reproducibility is the idea of scoping reproducibility, that is the repetition of experiments aimed at gauging whether or not results obtain under different conditions (in other words, the scope of the results: for instance, whether a given finding on the expression pattern of a particular genetic pathway holds only for the specific strain of mice on which the experiment was first carried out, or whether the same patterns are found when using other mice strains or even other organisms altogether). Scoping reproducibility consists in taking differences in the outcomes of repeated experiments not as evidence of error, but rather as evidence that findings may be restricted to specific experimental conditions (Leonelli, 2018). What researchers are interested in, in this case, is probing the differences between the circumstances under which certain results obtain and the circumstances where they do not, which in itself may teach something significant about the phenomena at hand.

5. Hypothetical reproducibility

In cases where research methods and settings are not highly standardised, and therefore difficult to replicate without significant alterations to the research design, the idea of hypothetical or conceptual reproducibility is often invoked. This is the effort to obtain outcomes that match those *predicted as implications of previous findings,* thereby indirectly confirming the reliability of the previous findings (Romero, 2017). This approach to reproducibility is grounded on the existence of robust chains of reasoning underpinning the research at hand: an hypothesis and/or prediction is made about the implications of a given set of findings, and when such hypothesis is empirically tested, it is taken as validation of the findings on which it was constructed.

6. Reproducible expertise

Particularly when exercising qualitative research methods within a highly volatile and dynamic environment, the changing nature of the target makes it very difficult to think about reproducing research results or even research methods, since those have to be constantly adapted to the changing situation of inquiry. However, there is an expectation that the expertise employed to conduct the study is reproducible: in other words, that any skilled researcher working with the same methods and the same type of materials at a particular time and place would produce similar results and pick out, if not the same data, at least the same overarching patterns and insights (Leonelli, 2018). Fieldwork in ethology or ethnographic research in anthropology, for instance, is heavily reliant on the idea of reproducible expertise, with methodologies developed specifically to cope with the impossibility of directly reproducing findings (such as vetted access and cross-samples research, or the governance of how research will be conducted; e.g. the centralisation of research in locations where many researchers can work together, check each other's work and ensure its reliability for those with no access to the same instruments / sources).

Narrow and broad interpretations of reproducibility

The six different forms of reproducibility we presented here are idealised and do not aim to be comprehensive of all forms of reproducibility present within contemporary research. Their usefulness lies in exemplifying the degree of diversity among approaches to best research practice, and the significance of finding ways to acknowledge and support such diversity in any initiative undertaken in support of reproducibility.

So what do we learn from considering multiple interpretations of reproducibility? Perhaps most crucially, we should recognize that reproducibility matters, but that it takes different forms suited to different research situations. Differences do not just emerge at the level of disciplines, but can be much more granular, depending on the specific goals, methods, materials, and conditions of research of the research group in question. Given this necessary variation, reproducibility needs to be conceptualised and implemented in ways that are not too narrow, thereby inadvertently constraining research in ways that are unhelpful, nor too broad to be meaningful.

Consider first the problems with backing too **narrow** an interpretation of reproducibility. This is sometimes done by institutional research policies, whereby universities focus on computational or direct reproducibility as the main and sometimes only way to understand this concept. When this happens, highly controlled experiments with pre-specified goals come to exemplify “best practice” and “rigorous research”, doing no justice to other research methods – such as qualitative traditions focused on analysis of situatedness, and exploratory quantitative research (e.g., data mining). And yet, a narrow interpretation of reproducibility is simply inappropriate in the case of many research settings, and particularly those characterised by lack of standardisation (perhaps because the phenomenon being studied is unique or changing all the time) and/or reliance on expert knowledge by the researchers involved. To illustrate, an ethnographic study of how indigenous people on a Polynesian island rebuild their physical and social structures after a devastating flood cannot be replicated in any meaningful sense at a later point.

By devaluing such research situations, the narrow approach to reproducibility devalues the significant role played by expertise and embodied knowledge in data production, processing and assessment, as well as the significance of social context in defining specific research settings. The narrow approach also tends to emphasise a false dichotomy between “hermeneutic” and/or “subjective” research methods, on the one hand, and “quantitative” and/or “objective” methods on the other, typically supporting the latter. This distinction is however unhelpful in assessing best practices across different types of research settings, and particularly basic research situations where researchers are developing the goals, methods, tools and materials for their work as they go along. Most unhelpfully, the narrow interpretation of reproducibility does not help to distinguish between unintentional mistakes, cheating and fraud, and variation due to differences in research conditions; nor does it help to separate genuine attempts to question and verify accepted “facts” from the malicious questioning of scientific results grounded in vested interests (the so-called “weaponization” of reproducibility, see below).

Now consider the problems encountered when adopting an overly **broad** interpretation of reproducibility. To illustrate, an overly broad interpretation might arise when requirements of replication are waived in favour of more flexible “conceptual replications” (i.e., other, usually unanticipated manifestations of a result are re-interpreted as being “conceptually” identical or similar, thus ostensibly constituting a “replication”). First of all, an overly broad conception of reproducibility opens up research to inappropriate scrutiny as well as possible abuse. While applying a strict understanding of direct reproducibility to ethnographic research makes no sense, it is equally problematic to justify the failure to reproduce well-specified experimental conditions by appealing to conceptual reproducibility and simply shifting to a different method. It is perhaps unsurprising that there is evidence that internal “conceptual replications” – that is, conceptual replications reported by the same original authors – do not make subsequent direct replication by other researchers more likely (Kunert, 2016).

Even something as seemingly sensible as refining one's research target during a study, in response to what one is learning about that target as one investigates, can be abused. Such strategies can be used to dissimulate the fact that a given research finding is too idiosyncratic to be credible, and open up a slippery slope for justifying post hoc whatever result is obtained.; This may be especially true in cases where the same research group is involved in conducting a study and carrying out conceptual replications, since that group has a clear instance, and this has been argued in particular when those conceptual replications are carried out by the same research group (which has a vested interest in confirming its own results), - leading some to argue that conceptual replication cannot and should not replace direct reproducibility (Kunert, 2016).

Secondly, a broad conception of reproducibility is commonly associated with notions such as generalisability and transparency of a given piece of research, and this association is problematic. As discussed in the case of scoping reproducibility, the generalisability of research is not the same as the quality of research: a given set of findings may be highly reliable when applied to one specific domain, and not useful at all when applied to another. Insights on how cell walls filter cell signalling to facilitate plant responses to pathogens, for instance, can be highly reliable and extremely significant to producing pesticides for crops; but they will not be generalisable to cell signalling in animals, and this lack of generalisability has nothing to do with their quality as research findings. Similar confusion can arise when associating reproducibility to the idea of transparency. While sharing information about an experiment undoubtedly makes it easier to reproduce it and independently verify its quality, sometimes too much transparency can make verification difficult or downright impossible (for instance when obfuscating the key elements of a research method by providing too much non-essential detail, which makes it hard for others to understand and re-use the method at hand). Precisely *who* should have access to information about a given experiment is also a key issue for reproducibility: is it relevant that anybody interested in replicating a study may access information on how to do it, or is it acceptable to restrict that privilege to trusted groups? This question becomes crucial in the case of commercially sensitive research, where typically reproducibility does not entail transparency, but rather the guarantee that regulators may verify the integrity and validity of a given experiment on behalf of prospective consumers.

Third, there is a tendency to think that by making research reproducible, one will also resolve issues surrounding the reliability of research results at scale, the exposure of research findings to transdisciplinary scrutiny, and the ease with which research can be translated into products. This is, however, a misunderstanding of what reproducibility can do for research. Reproducibility can be a helpful requirement to check the reliability of results under specified research conditions. It does not help to assess whether a given insight can hold when applied to a larger or different target (for instance when attempting to scale up and optimize technical solutions, such as a given piece of software, for a much wider pool of users than originally proposed, such as a large patient pool); it does not imply that a given piece of research will be exposed to feedback by a wide variety of stakeholders, which would provide robust scrutiny; and it does not ensure that the insights will correctly inform the development of commodities to bring to the market. Implementing optimization, transdisciplinarity, and translation arguably requires many more resources and different forms of labour than mere reproducibility.

The importance of walking the line between an overly narrow and an overly broad interpretation of reproducibility becomes clear when considering the possible use of pre-registration as a strategy to foster reproducibility of research results. Preregistration refers to the idea that experimenters specify a priori, before data collection commences, important details of the methodology and analysis. Preregistering one's intent guards against a number of inadvertent questionable research practices (e.g., "p-hacking" by continuing the experiment until a desired result has been obtained) and has been cited as an important aspect of reproducible science (Strømland, 2019). In our view, pre-registration can be an excellent tool when it is used as a way to record methodological reasoning and decisions at specific stages of the research process, so that the history of a given research project can be easily reconstructed, evaluated and reproduced. There should not be, however, an expectation that researchers stick to the plans made when pre-registering a study: well-justified variations in a research plan are often required as the project proceeds and researchers learn more about the objects of their study, so failure to stick with a plan does not necessarily indicate problematic and unreliable research (see Box 2).

BOX 2: Pre-Registration

THE "PREREGISTRATION REVOLUTION"

The ability to predict outcomes is fundamental to science. Although scientists may formulate hypotheses on the basis of unguided observations of natural events, much rigorous research is characterized by testing hypotheses about what will occur in an experiment. Testing predictions is widely considered vital to establish evidence for explanatory claims (Nosek, Ebersole, DeHaven, & Mellor, 2018). Predictive tests differ from "postdictions", that is attempts to explain – sometimes unexpected – findings after the data have been obtained (post hoc).

Problems arise when prediction and postdiction are confused or conflated. For example, if an unexpected finding is presented as though it had been predicted before the data were obtained, such post-hoc reasoning may lead to over-inflated confidence in the interpretation and may raise unwarranted expectations about the replicability of a finding. This Hypothesising After the Results are Known (HARKing; Kerr, 1998) is widely considered to be a questionable research practice that has contributed to the replication crisis.

Preregistration provides a solution to this problem because it requires that researchers stipulate their hypotheses ahead of time, before data have been collected or analyzed. Preregistrations are archived online and date stamped and, once submitted, can no longer be edited. Preregistration of hypotheses prevents inadvertent or furtive HARKing because readers (and reviewers) can check the author's interpretation of results against the preregistration.

POSITIVE ASPECTS OF PREREGISTRATION

Preregistration offers numerous advantages and safeguards. For example, in addition to guarding against HARKing, a detailed preregistration of method and analysis plan guards against several other questionable research practices such as eliminating aberrant observations in order to favourably affect the results of an analysis, or adjusting a sampling plan on the basis of partial results (e.g., keep testing participants until a result is significant). In general, preregistration curtails researchers' "degrees of freedom" that might otherwise unduly shape the generation and interpretation of data (Oberauer & Lewandowsky, 2019).

A particularly powerful extension of preregistration are so-called "registered reports" which involve peer review of method and analysis plan, and a provisional decision on publication, before the data are collected. The outcome then does not affect the publication decision, which guards against publication bias in favor of significant results. In addition, registered reports provide reviewer feedback to the researcher when it arguably matters most, namely before the study is conducted and while changes to the methodology can still be made.

Compiling such reports provides a useful, explicit picture of the assumptions made at the start of research, as well as providing an incentive to publish negative results in case the study does not succeed in eliciting positive findings (since the researchers already went through peer assessment proving that their research design was significant and credible, which in turn makes negative findings into a significant result in and of itself; see https://journal.trialanderror.org/).

COMMON MISUNDERSTANDINGS OF PREREGISTRATION

1. Preregistration is not a research blueprint

Several misunderstandings about preregistrations have been articulated that are important to correct. Perhaps most important among them is the misconception that a preregistration is an unyielding blueprint for research that must be observed at all costs and that prevents researchers from exploring unexpected aspects of their data. This concern is misplaced because preregistrations do not prevent or prohibit departures from an analysis plan. Instead, preregistrations merely stipulate that exploratory unplanned analyses are identified as such in the report of a study. In no way does this constrain creativity or exploration; it simply requires clear differentiation between planned analyses and others that were exploratory.

2. Preregistration is not a tool for research quality assessment

Preregistration guards against using the same piece of evidence to both generate and then test a hypothesis, which can arise when researchers create post-hoc explanations for unexpected results using a theory that was tested in the same experiment. This reasoning necessarily precludes falsification of a theory, thus luring researchers into placing inflated confidence in the theory (Rubin, 2017). Preregistration prevents the use of results in formulating a hypothesis by enforcing a strict temporal sequence: the hypothesis must be formulated and recorded before the data are known.

However, temporal sequencing is only one way in which it can be ensured that hypotheses are derived from the theory without being informed by the same data (because the data do not exist yet). However, temporal sequencing per se is not necessary irrelevant to ensureing that the results do not unduly influence the hypotheses are derived from a theory without being informed by the same data. The same end can be achieved by ensuring that existing data do not contribute to formulation of a hypothesis. A famous example involves Einstein's General Theory of Relativity, which is commonly credited with predicting the observed precession of Mercury's perihelion, even though those observations predated the General Theory of Relativity by several decades. Because those observations played no role in developing the theory, relativity predicted the precession of Mercury's perihelion and the success of that prediction yields support for the theory "just as fully as if those facts had come to light only after the formulation of the theory" (Worrall, 2014, p. 55).

Likewise, if a hypothesis is derived after the results of an experiment are known, but the derivation relies entirely on ante hoc (a priori) theory and evidence, then its legitimacy is identical to that of a hypothesis that was derived before the data were known (Oberauer & Lewandowsky, 2019; Rubin & Donkin, 2022). Moreover, preregistration is not required to prevent HARKing under those circumstances, provided they are adequately documented in the research report.

There is, however, one important condition for the legitimacy of hypotheses that are derived after the data are known, and that relates to the precision and conceptual clarity of the theory. In psychology, many theoretical constructs are fuzzy and lack conceptual rigor (Bringmann, Elmer, & Eronen, 2022). This can make it difficult to ascertain whether a hypothesis was truly derived from a theory without having been informed by the data it purports to predict. To avoid this problem requires that theories be formulated at a computational or quantitative level with all auxiliary assumptions being spelled out (Oberauer & Lewandowsky, 2019).

PART 2. Implementing reproducibility within academic institutions

Whatever form reproducibility ultimately takes, it requires institutional support and embedding in the research culture to be successful. The extensive consultations with stakeholders provided us with insights about current practices and possible futures in Flemish research.

Implementation practices

Several important changes to the ways in which research is institutionalised, rewarded and funded have emerged from our consultation with Flemish stakeholders. Among them, the most notable are:

- A recognition of <u>the role of data stewards</u> as integral parts of the research ecosystem, with many universities investing in permanent data steward positions to offer support in data collection, curation and management. Having discussed the role with several data stewards who participated in our consultation session, we noted the positive impact of these professional figures on the academic landscape, both in terms of supporting research and in terms of helping to mediate conflicts and tensions among stakeholders (see section below). At the same time, we also noted that these roles were quickly overwhelmed by the workload and there is still a lack of recognition for the expertise involved in data curation, with such roles often delegated to the status of a technician rather than a researcher, and promotion/progression trajectory being delimited accordingly.

- A heightened awareness of <u>the significance of integrity within research</u>, resulting in a boost of ethics boards and integrity officers at various universities. These developments are very welcome, provided they are sufficiently streamlined to guard against over-burdening researchers with additional workload. However, we detected similar issues to those experienced by data stewards with regards to the recognition, and potential undervaluing, of such roles within institutions.

- Novel <u>training resources</u> for researchers and professional support staff to learn how to make their work more responsible and robust methodologically. These were welcomed by researchers and institutions alike, as well as by private collaborators, though it was also noted that training programmes constitute a limited measure given that they place the burden of complying with reproducibility standards squarely on researchers' shoulders and away from institutional infrastructure. Training should therefore be complemented by <u>workload allocation</u> that recognises the time and effort that researchers and support staff need to spend in getting and implementing such training, as well as the alternative outputs that may derive from such effort (such as datasets, infrastructures, and other resources that are shared with the broader community to great impact, and yet are not typically considered noteworthy as marks of excellence by academic employers and funders).

- A willingness to take account of integrity and Open Science practices by funders such as the FWO, where negative results are currently recognised as a legitimate output for research assessments, and considerations around research integrity (such as, whether the research being considered is conducted responsibly and in collaboration with the right stakeholders) now constitute a so-called "second axis" for research evaluation. These are very encouraging signs, as funders have enormous power to change the system of incentives by nudging researchers and their institutions to pay more attention to these issues. At the same time, discussions among stakeholders revealed unease among researchers about the label of such considerations as "second axis" for research: It was remarked that such label can be interpreted as a confirmation that questions of integrity, quality and reproducibility are not a priority for research assessment, and will always play a secondary role compared to number and venue of publications planned or resulting from a given project. The FWO expressed strong awareness of these issues and concerns, and highlighted the difficulties in implementing such novel criteria of assessment in the Flemish context, where much of the assessment is carried out by international expert panels who have highly diverse perspectives on what should be considered valuable in research. Of particular note are difficulties in assessing transdisciplinary research, which many institutions regard as paramount to confront global challenges in theory, but in practice remains very hard to assess through traditional research metrics and discipline-focused review panels.

- The recognition of Open Science as an important phenomenon and attempts to instigate a culture change within academic institutions. Such attempts can take substantive forms, such as modification of criteria for hiring or promoting academics so that they include efforts in data curation, reproduction of other studies and quality assessment. Some universities are attempting to bring in such consideration at the professorial level, for instance through the adoption of narrative CVs and self-declaration statements around Open Science compliance, though it is as yet unclear whether bringing in such additional criteria is making a difference in who is hired as or promoted to Professor. These initiatives were welcomed by researchers as extremely important in shifting the academic system of credit towards taking better account of good research practice. Another, less demanding change consists of recognising good scientific work through one-off awards, such as Open Science prizes and 'badge' systems for recognition of OS activities. These awards were treated with more ambivalence by researchers, who feared that these may be facile forms of recognition that would make no difference at all as long as they are not taken into account by hiring and promotion committees, and by funding bodies when awarding grants. These forms of recognition are also at risk of being easily 'gamed', thus giving rise to" 'open-washing" (i.e., the cosmetic dressing up of research practices as confirming to Open Science mandates, when in fact little has changed and researchers have simply learned to "look" like they are compliant; for instance, when a researcher makes her data "open" by loading them onto a database, but does not actually bother to format those data for re-use by others, does not provide appropriate metadata and/or only loads data that she thinks may not be very significant).

Alongside institutional initiatives, we found a wealth of bottom-up activities and training programmes instigated by researchers on the ground. Particularly deserving of mention here are initiatives coming from early career researchers, such as the ReproducibiliTea.[1] ReproducibiliTea started as a small journal club at the University of Oxford that was dedicated to discussing papers relating to reproducibility. It is now an international network of journal clubs spread across more than 100 institutions in 25 different countries. In Flanders, a main node of the network is at the KU Leuven.[2] The goal of ReproducibiliTea is to create a safe haven for interdisciplinary discussion about science in general, and more specifically about open science and reproducibility. Participants read relevant articles ahead of each meeting and discuss those issues in break-out sessions, with a plenary introduction and concluding reflection. The meetings are informal and interdisciplinary, and coordinated by PhD students and postdoctoral researchers, who commit their time to these initiatives because they believe them to be a valuable contribution to their own training and to the research excellence of their host institution.

These bottom-up initiatives have made a real difference by offering training and opportunities for discussion among researchers, as well as a forum to bring in external expertise when needed. At KU Leuven, conduits to the administration exist through the open science task force of the University. We should note that such initiatives do not always seem to be well-supported and recognised by the institutions who host them, which is a shame (see recommendations). It remains an open question whether the PhD students and postdoc devoting effort to these initiatives will find their work for Open Science rewarded in their future careers.

Tensions among stakeholders

The call for reproducibility has introduced a variety of tensions between different stakeholders and different communities which need to be explored and resolved. Perhaps foremost among them is the tension between publicly funded researchers,

---

1 https://reproducibilitea.org/
2 https://kuleuvenopenscienceday.pubpub.org/pub/1479tyu8/release/1

mainly in universities, who are expected to be fully transparent on the one hand, and researchers working in industry on the other, who receive scrutiny from regulatory bodies but are not required to undergo scrutiny by academic peers, particularly when it comes to their upstream research efforts.

Additional tensions have been identified between researchers at different levels of seniority, mainly because the additional cost and effort involved in conducting research to a higher criterion of reproducibility is borne by junior researchers who are more likely to have acquired relevant skills. Data curation, for instance, is increasingly acknowledged as a highly skilled task, which deserves recognition in the form of co-authorship of papers and academic status (European Commission 2018, Open Science Policy Platform 2020). The current hierarchy of researchers, where Professors who have ideas and look for funding sit at the top and postdoctoral researchers who conduct experiments and process the data sit at the bottom, does not reflect the expertise and seniority required in order to intersect effectively with the complex international data ecosystem (Leonelli et al., 2016; Leonelli et al., 2017).

Relatedly, there are significant tensions between work considered to be 'academic' and work considered to be 'technical' or constituting 'support' for research. The role of professional support staff often remains unclear. Often those who support reproducibility efforts and are not themselves academic researchers are undervalued, despite the high level of skill required. For instance, it is great to see the emergence of data stewards as a separate and well-recognised professional role, and to see that the technical and professional skills required for such a demanding role are recognised. However, relation of such professionals to research (e.g., are they expected to be authors?) remains unresolved, and promotion/progression for such specialised staff is often underspecified or limited. Even less clear is the relation of staff supporting ethics and communication strategies to reproducibility efforts. This staff, when interviewed, showed clear awareness of the key role they can and should play in reproducibility efforts alongside researchers; but this awareness was not reflected in institutional arrangements and, often, in interviews with researchers who regarded such support as extraneous to "proper" research. These barriers are not helping.

Our consultations identified a key perceived problem in the pursuit of reproducibility, which is that <u>institutional reward and incentive structures are lagging behind</u> and continue to revolve around conventional metrics (i.e., quantity of publications and citations with no direct representation of reproducibility indices; see critiques in European Commission 2018). This leads to an emphasis on rapidly obtainable outputs over the short term with little recognition of the longer ramping up times required to build a laboratory and research culture that focuses on reproducibility. We observed that even when institutions complement performance metrics with qualitative evaluation, which provides a possible avenue towards rewarding reproducibility efforts, the emphasis continues to be on novelty over replication or quality. Although the funders are aware of the problem, they find it challenging to shift emphasis during assessment. The problem is particularly acute for the research centres which who indicated that while they take reproducibility very seriously, due to their direct link to industry, they find it challenging to match the levels of transparency that are available to university-based researchers. are more focused on producing. One activity that is crucial to achieving reproducibility but that remains largely invisible is reviewing of journal submissions or grant applications. Reviewing is a labour-intensive activity but is entirely voluntary and almost completely unrewarded.

Last but not least, a strong and recurrent theme that permeated most discussions with stakeholders involved questions of trust. How can a research culture be built that relies on open and frank discussion, when everybody is monitoring everybody else for signs of ‘bad faith’ or non-compliance with reproducibility procedures? We pick up on this broader issue in the next section.

PART 3. The Tension Between Resilience and Reproducibility

Science is rarely conducted in a vacuum and the interaction between the public and the scientific community can take many forms (see Box 3). When scientists discover a planet in our Milky Way that is made of diamonds (Bailes et al. 2011), public fascination and applause are virtually assured. However, when scientists discover that large-scale combustion of fossil fuels causes climate change, or that a lethal airborne virus requires mask wearing and social distancing to be controlled, then the public and political response may be less favorable. In many cases, public or political disapproval can translate into attacks on individual scientists or groups of scientists that may compromise the scientific process or its outcome (Brysse, Oreskes, O'Reilly, & Oppenheimer, 2013; Freudenburg & Muselli, 2010; Lewandowsky, Oreskes, Risbey, Newell, & Smithson, 2015). To illustrate, there is overwhelming evidence that the science of climate change has been subject to well-funded and well-organized opposition that has not only successfully delayed policy action, but has also undermined the public's perception of the strength of the scientific consensus (Brulle, 2013, 2018; Brulle, Hall, Loy, & Schell-Smith, 2021; Dunlap & McCright, 2011; Lewandowsky, 2021; Oreskes & Conway, 2010).

BOX 3. Science communication

The public rarely receives scientific information directly from scientific sources. Instead, a variety of interlocutors are usually involved in translating scientific findings and results into media reports or other sources that are accessible to the public. Science communication, at its best, offers the opportunity for the public to both gather insight into and contribute to the process of scientific knowledge production. Local organizations and non-academic groups and institutions (e.g., musea) have an important role to play when it comes to communicating research in accessible ways that emphasise the reasons for trusting specific scientific results and foster participation in research for those who are willing to become more engaged.

When science communication is expanded to involve the community into the scientific process itself, the outcome may be highly productive and may fulfil community needs that might otherwise remain unmet. A well-known example involves the town of Pickering in Yorkshire which had repeatedly experienced devastating floods in the early 2000s. The government-proposed initial solutions were met with strong resistance by the local community, which ultimately succeeded in becoming directly involved in devising alternative solutions that found broad community support and protected the historical town (Garvey & Paavola, 2021; Whatmore & Landström, 2011).

However, not all science communication is successful, and there are several common pitfalls. Perhaps the most common misconception involves the so-called "deficit model"; that is, the idea that the public is lacking knowledge that science communicators can provide, thus remedying the public's information deficit (Hornsey, 2020). Although information provision can make a positive difference (and indeed ultimately it must; otherwise all education would be doomed to fail), it can give rise to several problems. First, it is encouraging a simplistic view of science as a source of ultimate truth, which makes scientific progress (including u-turns, changes of understanding and mistakes) look like a betrayal of trust. Rather, the strength of science lies in its fallibility and in its ability to take nothing for granted: for publics to be able to assess the value of scientific claims, it is imperative to support an understanding of science as a self-correcting process of inquiry, rather than as a set of infallible truths (Oreskes, 2019). Second, the view of scientific communication as 'information provision' does not provide any avenue for non-researchers to enter in a dialogue with scientific research, for instance by contributing their own expertise (as often the case with citizen science projects). This lack of avenues for engagement can have severe implications, as seen in the context of scientific responses to the COVID-19 pandemic, where the few projects who involved citizens as participants yielded much more reliable results than projects who excluded citizen participation from the get-go (Leonelli, 2021). Third, the view of scientific communication as 'information provision' provides no path for a dialogue between researchers and publics, which makes any conflict look like an insurmountable disagreement. Indeed, very often the obstacles surrounding the communication of science arise from conflicts between people's deeply-held worldviews and scientific evidence (Lewandowsky & Oberauer, 2016). People who smoke may be motivated to reject evidence about the ill health effects of tobacco. People who endorse a laissez faire approach to free-market economics are motivated to reject evidence for climate change, and the interference in the markets that any mitigation will require. Science communication under those circumstances is difficult and requires careful analysis and suitable messengers and messages. (See Lewandowsky, 2021, for a review of communication tools.)

Any consideration of reproducibility must therefore also consider the *resilience* of the scientific enterprise against politically motivated interference or other types of vested interests. It turns out that the relationship between transparency and reproducibility on the one hand, and resilience on the other, is complicated, nuanced, and sometimes counter-intuitive. For example, freedom-of-information requests launched under the banner of transparency, typically targeting e-mail correspondence between scientists, have been a frequent tool of political operatives who sought to undermine public trust in climate science (Lewandowsky & Bishop, 2016; Lewandowsky, Mann, Brown, & Friedman, 2016).

Similarly, in the U.S., much political effort has been expended on increasing the "transparency" of science-based institutions through legislation or regulation. Those efforts have been characterized as a "Trojan Horse" by members of the public health community (Levy & Johns, 2016) because they allow vested interests to delay policy initiatives by challenging data that federal agencies rely on. For example, the campaign to legislate "sound science" in the 1990s, which sought to enact data access and data quality laws, was initiated by Philip Morris in an effort to dispute the link between second hand smoking and lung cancer (Baba, Cook, McGarity, & Bero, 2005). The resultant Data Access Act and Data Quality Act have demonstrably delayed health-related regulations by permitting corporate interest groups to challenge the underlying data. The vast majority of all public challenges to data were by regulated industries, lobbyists, and trade organizations (Levy & Johns, 2016).

The consequences of access to data by vested interests can be illustrated with research on the link between smoking and coronary heart disease (CHD). The tobacco industry, through its "Council for Tobacco Research", began to fund a longitudinal study into CHD after federal funding was discontinued 22 years into the project. The tobacco industry provided funding for the primary purpose of gaining full control of the data (including the period of federal funding) for reanalyses (Cataldo, Bero, & Malone, 2010). Perhaps unsurprisingly, these reanalyses ascribed the increased risk of CHD in smokers to "constitutional factors" such as ethnicity.

Further legislative and regulatory efforts under the banner of transparency and reproducibility have continued almost to this date, culminating in the U.S. Environmental Protection Agency's (EPA) proposal, under the Trump administration, to exclude any studies whose raw, individual-level data are not publicly available from consideration in the design of environmental standards or regulations (Schwartz, 2018). This requirement for transparency is problematic because much of the evidence involving environmental pollutants involves large cohort studies that include sensitive medical information and pollutant exposure data based on participants' place of residence. In consequence, those data are readily de-anonymized and participants' privacy cannot be ensured if the data are made publicly available (Schwartz, 2018). Because preservation of privacy is mandated by other legislation and ethical codes, the proposed transparency requirement would effectively exclude a large portion of epidemiological data from consideration by the EPA. The proposed rule was vacated in court on procedural grounds and is unlikely to be reinstated by the Biden administration (Lash21). A similar legislative effort, the Secret Science Reform Act (SSRA) that neatly paralleled the proposed EPA regulation was sponsored by a Congressman (Lamar Smith, R-TX) with a long history of opposition to climate science and close ties to the fossil fuel industry (Levy & Johns, 2016).

In summary, some political efforts that on the surface seem laudable – who would not want greater transparency and reproducibility? -- turn out to be problematic on further inspection and may achieve the opposite of their ostensible purpose. There are inescapable tensions between transparency and other important concerns, such as privacy, and there is no one-size-fits-all resolution of that tension that necessarily benefits science (Leonelli, 2023). Box 4 discusses one particular case to illustrate the dilemma.

Emphasis on reproducibility also creates multiple complex tensions with efforts to increase diversity and inclusivity of scientific research. Although this often escapes notice, reproducibility relies on diversity and inclusion. For example, non-Hispanic whites of European ancestry only comprise 61% of the U.S. population but they comprise more than 90% of participants in clinical trials (Ma, Gutiérrez, Frausto, & Al-Delaimy, 2021; Mak, Law, Alvidrez, & Pérez-Stable, 2007). This lacking inclusivity raises the possibility that much medical knowledge may not be reproducible in all sectors of the population. In confirmation, the side effects of 5-Fluorouracil, a common cancer drug, are now known to occur at higher rates in under-represented populations. Because the original clinical trials involved predominantly white participants, these side effects in minoritized groups were initially overlooked (Yates, Byrne, Donahue, McCarty, & Mathews, 2020). Clearly, therefore, to be fully reproducible, science must strive for inclusivity in samples of participants in clinical trials.

The need for inclusivity in participant samples, however, creates other tensions with reproducibility. Ethnic minorities have historically been discriminated against in the health care system and suffer subtle (or not so subtle) discrimination to this date. For example, African-Americans are systematically undertreated for pain relative to white patients, based at least in part on widespread but baseless beliefs about biological differences between races (Hoffman, Trawalter, Axt, & Oliver, 2016). Moreover, in light of a history of mistreatment of indigenous populations by western countries (Lowes & Montero, 2021) and continued contemporary misuse of medical institutions in developing countries (e.g., by the CIA in its hunt for Osama bin Laden; Reardon, 2011), it is unsurprising that minoritized people would be reluctant to participate in research by the same institutions that caused them harm or deceived them in the past. It was as recently as 2021 that the American Psychological Association (APA) issued an "Apology to People of Color for APA's Role in Promoting, Perpetuating, and Failing to Challenge Racism, Racial Discrimination, and Human Hierarchy in U.S."[1] Against this background, the blanket use of "open data" when research involves sensitive information is problematic and in conflict with the goals of diversity and inclusivity. Black or Hispanic participants may be willing to participate in an experiment on working memory or attention, but they may be less likely to participate if they knew their data might be re-analyzed to examine associations between cognitive ability and

1 https://www.apa.org/about/policy/racism-apology

race or ethnicity–and yet, an "open data" transparency regime would not prevent any such reanalysis (Fox Tree, Lleras, Thomas, & Watson, 2022, Lewandowsky & Bishop. 2016). There is indeed ample evidence of data about individuals and communities (from social media, medical services, quantified self-technologies and/or administrative sources) being used in ways that discriminate against minorities and vulnerable members of society (Leonelli et al., 2017; Taylor, 2017; Noble, 2018; Eubanks, 2018; Zuboff, 2019).

In summary, reproducibility is partly contingent on inclusivity. But achieving inclusivity requires particular care with safeguarding participants' privacy and anonymity, which may in turn create tensions with the transparency that many consider to be essential for reproducibility. Again, there are no easy solutions to this tension between competing aspects of reproducibility.

BOX 4: Belfast Project and implications

A striking concrete example of the double-edged nature of transparency, and how it can be in conflict with commitments to privacy, involves The Belfast Project, an oral history of "The Troubles" in Northern Ireland that was hosted by Boston College in the U.S. The Troubles refer to the violent conflict in Northern Ireland between republicans (i.e., those who wanted Northern Ireland to become part of the Republic of Ireland) and loyalists (i.e., those who wanted Northern Ireland to remain part of the United Kingdom) that erupted in the 1960s and continued until 1998, causing the death of some 3,500 people.

The Belfast Project is a collection of interviews with former republican and loyalist paramilitaries with first-hand knowledge of bombings, kidnapping, and murders committed by both sides in the conflict. Participants were promised confidentiality and that their taped interviews would not be released until after their death.

In 2011, the US Attorney General issued subpoenas for the data to Boston College, the custodian of the project, on behalf of the Police Service of Northern Ireland. After a prolonged court battle, and over the objections of the researchers, the Police Service of Northern Ireland succeeded in obtaining the tapes.

A short time later, Gerry Adams, the leader of the republican party Sinn Fein, was held for questioning for four days in connection with evidence contained in the tapes. The release of the Boston College tapes also put an end to another planned oral-history project that had relied on promising participants confidentiality, putting at risk efforts to create a better historical archive of The Troubles (McDonald, 2014).

The tension between transparency and replicability on the one hand, and the reality of research in violent settings, is not limited to Northern Ireland but is a defining attribute of field research conducted under difficult circumstances (Thaler, 2021).

BOX 5 - Reproducibility and Transparency: Key Challenges

There are good reasons to be careful in associating the demand for reproducibility with a simplistic desire for more transparency in research. Transparency is not always desirable, and needs to be managed intelligently, as argued already in 2012 by the Royal Society report "Science as an Open Enterprise" (Boulton et al., 2012). Among the reasons for such caution, consider the following three concerns:

1. Privacy and surveillance: it is already clear how particularly in the case of research on human subjects, transparency can lead to problematic surveillance, breaches of privacy and abuse of personal information, such as experienced by patients whose medical insurance costs rise after they consulted a doctor (Ebeling, 2016; Dennis et al., 2019; Tempini & Leonelli, 2018). This can also happen in cases of research on landscapes and climate, since those investigations can include information about communities living in the region and constitute a veritable tool for surveillance of those populations (Willamson & Leonelli, 2022). Clear examples of such issues are the use of high-resolution satellite images, which can be used to capture the spread of plant pathogens and document human right abuses or war crimes - but can also easily be deployed to plan hostile attacks or exploitative bioprospecting of specific regions.

2. Trust. It is impossible, in theory and in practice, to convey every single aspect of a scientific procedure in a comprehensive way: what researchers need to do when deciding how to describe an experiment, for instance, is to select which aspects of their methods and settings are most salient to reproducing that experiment, so that other researchers can focus on those aspects when probing the experiment's results. Needing to select which aspects are described and shared brings issues of trust into the reproducibility debate, since researchers need to understand and trust each other to capture the most salient aspects of their practice in ways that are intelligible to their peers. This may prove very difficult in situations where researchers do not share the same background, and are therefore unable to understand the reasoning behind specific sharing choices (Nguyen, 2021). In the absence of such trust and reciprocal understanding,

sharing information - even in the most transparent of ways, such as by putting information online - may well prove useless to those attempting to evaluate and reproduce results (Leonelli, 2023).

3. Diversity. The considerations made in Section 1 around the multiple meanings and forms of reproducibility within contemporary research have implications for understandings of transparency. Research on scientists' attitudes to data sharing found that many researchers are highly supportive of Open Data and reproducibility policies and yet – at the very same time – some felt threatened or undermined by OS initiatives (Levin et al., 2017; Bezuidenhout et al., 2017; Leonelli, 2022; State of Open Data reports, Science, 2022). This was particularly the case for researchers working in less visible and well-funded environments, some of whom experienced severe difficulties in using open data repositories and reproducibility protocols. Some of these tools reflect the interests, resources and cultural assumptions of those who develop them (who are largely based in rich institutions from the Global North). Researchers also reported difficulties in accessing high-end technologies and software often presupposed by data sharing initiatives, particularly when trying to acquire or develop OS resources and skills that are directly relevant to their own specific priorities and settings. There is arguably an ongoing mismatch between high-level, universal principles associated with OSto OS, such as "openness", "transparency" and "reproducibility", and the different ways in which those principles are interpreted depending on the research environment at hand (Leonelli, 2023). This needs to be taken into account when stipulating policies around reproducibility and transparency in research.

A further, often-overlooked tension is that the goals of reproducibility and transparency can come into conflict with cultural inclusivity and diversity of the research community. There is concern that scientists from marginalized backgrounds (e.g., women of color) may be unintentionally harmed by the goals of transparency and openness. There are several reasons for this that were articulated by Fox Tree et al. (2022). First, there is evidence from the U.S. that Black or Brown researchers, despite equal qualifications, are less likely to be funded by federal agencies (Ginther et al., 2011). This pattern of dismissal and overlooking creates a disincentive for data sharing for fear that a researcher may not be given due credit for their work and their data. A second reason is that researchers from historically disadvantaged backgrounds may preferentially study marginalized communities that have a long history of being distrustful of mainstream science (often for very good but painful reasons). We already noted earlier that sharing of data from research involving such communities is particularly fraught with risks, given the possibility of deanonymization (which is easier for minoritized participants than White participants) or the possibility of re-analysis of data in pursuit of questionable political goals (e.g., to compare cognitive performance between different ethnic groups; Lewandowsky & Bishop. 2016). In addition to the consequences for participants already highlighted above, the added burden of responsibility on researchers from minoritized backgrounds must not be underestimated. We did not find widespread awareness of this problem among stakeholders.

In summary, reproducibility and transparency are not simple unipolar concepts. They require nuanced analysis and, very often, non-trivial trade-offs between competing demands. Transparency may be at odds with expectations of privacy and reproducibility may compromise the resilience of the scientific community. In order to create a research culture that combines reproducibility with resilience, trust, and respect for privacy, those trade-offs must be examined frankly and thoroughly.

PART 4. The Tension Between Public and Private Interest

The concerns around transparency and lack of open scrutiny discussed in the previous section are not limited to the private sector, and are in fact endemic both to the incentive structures and to the competitive regimes in place in publicly-funded research.

On the one hand, both publicly and privately funded researchers have reason to be extremely careful about the validity of their findings. Arguably, the stakes are higher for commercially exposed research, and particularly research devoted to transforming scientific insights into commodities to be made available on global markets. In that domain, researchers are under huge pressure to reproduce results and use very precise and standardised quality checks. There are of course regular inspections by regulators, which private companies need to comply with. Even beyond those standard checks, however, the financial and reputational costs of releasing products to the public that do not function as promised, or - even worse - bring harm on potential users (as in the case of a drug that turns out to have deadly side-effects), are enormous, and there is a strong commitment to spotting possible mistakes in the research before they result in defective products. Indeed, representatives of the pharmaceutical industry that reported on their work within this exercise emphasised the strict training and protocols in place to foster and police good research practice, including specific guidelines on reproducing findings and policies on detecting and rectifying errors in both hypothetical and real-life situations.

On the other hand, however, both publicly and privately funded researchers are exposed to incentives and demands that skew their activities towards producing outputs as fast as possible, with little time devoted to exploring implications and to documenting and properly scrutinizing research processes. For privately funded researchers, aside from the enormous pressure to produce actionable results quickly (especially given the costs of carrying out large experiments such as clinical trials), there are severe obstacles to the free sharing - and wide-ranging scrutiny - of data. This is chiefly due to commercial sensitivity, with companies trying to protect their investments in research from being scooped by competitors; though there are also concerns with privacy (for instance when dealing with medical data) and with the possible weaponization of data (in cases of findings that threaten powerful industrial complexes, such as the fossil fuel industry in the case of climate change).

In the case of researchers working for the public sector, the commercial sensitivity of results may not be as strongly emphasised, but reluctance to make results and methods widely available may emerge from the incentive to compete with other groups towards discovering new technologies and making novel claims. The competitive nature of academic research, where contributions are valued as novel and important depending on whether they are the first to achieve a given discovery, means that there are few incentives to spend time to investigate and report the human and environmental costs and significance of a given discovery (Open Science Policy Platform, 2020). There is a tendency to privilege secrecy over open collaboration, especially in crowded domains such as biomedicine; and to rely on automated systems that may speed up data access and analysis, such as various forms of machine learning algorithms, and yet also tends to black-box the underlying analytic processes (Leonelli, 2017). This competition is particularly problematic for researchers from historically disadvantaged groups, for the reasons noted earlier.

These tendencies are greatly augmented by the multiple ways in which private and public research depend on each other, ranging from funding to digital infrastructures. The principles of reproducibility and competition are in strong tension with each other. It is very hard to check quality claims of industry given lack of transparency. We were told that they do yearly audits and conduct very stringent training of researchers, and worked hard on "open science within the company", but all this stays in-house (we were not given access to any training materials or programmes, despite explaining that those materials would be treated confidentially). Scientific institutes working at the boundary between public and private research are in a particularly advantageous position to tackle these issues by improving standards for scrutiny across the board, and ensuring oversight of the social implications of innovation developed from publicly funded research.

BOX 6: Research Institutes at the Translational Cutting Edge

There are several research institutes in Flanders that sit at the intersection of the public and private sectors.

- VITO: The mission of VITO is to accelerate the transition to sustainable energy and to facilitate the creation of a circular economy.
- Flanders Make: Their mission is to conduct high-tech research in support of companies in the manufacturing industry to develop and optimise products and production processes.
- VIB: VIB aims to support the translation of biological and biomedical research in Flanders into products for patients, farmers and consumers while, at the same time, creating new economic activity and new jobs.
- ILVO: The mission of ILVO is to conduct research in agriculture, fisheries, and food, with a particular focus on sustainability.
- VLIZ: The Flanders Marine Institute promotes marine knowledge creation and excellence through research about the ocean, seas, coast and tidal estuaries.
- Imec: Imec's mission is to enable nano- and digital technology innovation.

Although these institutes have clearly differentiated roles and span different domains, they also face common challenges arising from the need to bridge the public and private sectors. Supporting these institutes is crucial to Flanders' science ability to cross the so-called translational gap and ensure that there is a sustainable and supported pathway to take insights garnered within basic research to the market. At the same time, a principal challenge is whether the research institutes can ensure the responsible commercialization of their research outputs. In our consultations with stakeholders in Flanders, we have found more attention paid to the effectiveness of translational research than on its accountability to wider society. Some of the institutes have clear mandates to confront urgent social challenges such as waste control (VITO), overfishing (VLIZ) and more broadly food security (ILVO). There is however no institutional body in charge of overseeing whether transnational research carried out under those banners is socially responsible. This is a missed opportunity given the advantageous position of institutes such as the above to ensure the responsible translation of public findings to innovation.

PART 5. Recommendations

Based on the extensive discussions with stakeholders and the literature reviewed in this report, we formulate some recommendations that should form the basis for further discussion. Our recommendations arise from our analysis of the complexity of the reproducibility issue, and the tensions between competing values and imperatives that reproducibility involves.

We contextualize our recommendations within the overarching current reward structure that is governing academia. Academia worldwide is governed by a strong emphasis on a set of ostensibly "objective" metrics to measure performance, both at the individual level and at the level of institutions. We consider this emphasis to be misplaced and counter-productive. Most existing metrics, from counting publications or citations to other bibliometric indicators, emphasize short-term achievements and are therefore often antithetical to long-term strategic research efforts. Moreover, any metric will not just *measure* performance, but it will also *shape* the behavior of researchers and institutions. For example, if the number of publications (rather than their quality) is an important metric, then researchers will be incentivized to publish more papers at the possible expense of quality.

In addition to those in-principle problems, the emphasis on metrics creates a further challenge for reproducibility because there are no firmly established "metrics" for reproducibility. Moreover, even in institutions where metrics are complemented by qualitative evaluation – as for example during the U.K.'s REF program – the emphasis is on rewarding novelty over replication and quality. To the extent that recognition is provided, for example by journals that award "badges" for open data, preregistration, and so on, those fairly superficial indicators can foster "open washing"; that is, symbolic gestures to satisfy the demand for reproducibility without deep commitment to reproducibility. At present, therefore, widely-used performance criteria do not reward reproducibility. We noted earlier how this constitutes a persistent challenge in the Flemish institutional context. The problem is compounded by the fact that other activities that contribute to reproducibility and quality control remain largely invisible. For example, reviewing for journals or grant panels is largely unrecognized, although recent attempts to record reviewing activity via platforms such as ORCID or Publons (now subsumed in Web of Science) is seeking to raise the visibility of such essential activities.

Recommendation 1*. A commitment to a systematic reassessment of the role of research assessment and how it relates to reproducibility.*

The *Agreement on Reforming research Assessment*[1] that was launched by the European Commission in January 2022, and that has now been signed by more

[1] https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/reforming-research-assessment-agreement-now-final-2022-07-20_en

than 350 stakeholder institutions, provides a first step in that direction with its vision that "the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer-review is central, supported by responsible use of quantitative indicators. Among the imperatives for such a reassessment, is the careful evaluation of the characteristics and needs of different areas of research, and attention to avoid unintended discrimination among disciplines by instituting domain-specific reproducibility criteria as an overarching model of 'best practice'. The reassessment also needs to take the highly distributed nature of knowledge production into account, by giving due credit to activities sometimes classified as 'purely technical', 'support' or 'service' (such as data management and quality assessment), which are however highly skilled and crucial to research progress.

Recommendation 2. *A commitment to expand understandings of reproducibility to include disciplines and approaches that cannot, given their focus and methods, implement direct or computational reproducibility.*

While we signalled a wide variety of understandings of reproducibility, each of which is tailored to the specific characteristics of research practice across domains, there is a tendency in policy debates around reproducibility to focus only on what fits highly controlled and standardised experiments. We recommend that the understanding of reproducibility is extended to include the forms of reproducibility used in less standardised, more explorative and qualitative forms of research, whose emphasis on reproducible expertise has much to teach to those advocating for strict standardisation of research environments as the key solution to the reproducibility crisis.[2]

Recommendation 3*. A commitment to examine the implications of reproducibility initiatives on the full diversity of stakeholders, with particular emphasis on junior researchers, minoritized groups, and women.*

Enhancing reproducibility requires institutional reform and support, and it cannot be achieved by training researchers in new methods alone. On the contrary, the availability of training resources (workshops and so on), while well-intentioned, could be counter-productive unless those initiatives are accompanied by institutional infrastructure and recognition of the expertise and efforts involved in enhancing reproducibility. Having clear recognition structures for the expertise of data stewards, technicians, and early career researchers deeply involved in quality checks and data curation efforts, for instance, would immediately provide visibility to those efforts, those helping to address the tensions we mentioned

[2] See also Onghena (2020) for recommendations specifically tailored to the enactment of reproducibility within the humanities.

between different contributors to research. Similarly, placing more emphasis on responsible collaboration with non-academic stakeholders, including interested citizens as well as non-academic partners, will enhance the responsiveness and openness of both public and private research. Finally, it is crucial to flank training initiatives within research institutions with appropriately remunerated, permanent positions aiming to support existing researchers in complying with Open Science and reproducibility standards, which are extremely labour-intensive and cannot be simply added to the existing heavy workloads of researchers.

Recommendation 4*. Explore mechanisms to support building a research culture of open discussion that relies on mutual accountability to enhance error detection and correction without creating paranoia and mutual distrust.*

An almost inevitable consequence of increased transparency is an increased likelihood of error discovery. No one is immune against committing errors during analysis or experimentation, and the detection of errors is an essential aspect of the self-correcting nature of science. However, the detection of errors presently incurs a reputational cost for the researcher involved, and hence researchers who practice transparent research are taking a risk that their colleagues who are less transparent can avoid. A key path towards reproducibility consists of facilitating positive dialogue over mistakes and the significance of lessons learnt, without taking anything away from researchers' autonomy and the uniqueness of each research situation. An example of a constructive initiative in this respect is the *Journal of Trial and Error*, started by a group of PhD students at the University of Leiden and rapidly recognised internationally as a venue to publish negative results and experiments that did not work out, as well as generating a healthy debate around what constitutes a "significant failure" and when/how do mistakes deserve to be celebrated and recognised on a par with putative research successes (https://journal.trialanderror.org/). Similarly, many Open Access journals are now focusing less on assessing what may count as "new ideas" within a given domain (which is often a matter of dispute among referees and can be biased towards specific types of contributions and contributors), and more on the reliability and quality of the knowledge being produced. The emphasis placed by the FWO on negative results and integrity is also laudable in this respect, and should be strengthened in the future as a way to support a culture of sharing and learning from failures as much as from what happens to work out.

Recommendation 5*. Set up science communication strategies that focus on the engagement of multiple publics into research processes, and that ensure an understanding of science as a sophisticated yet fallible endeavour, whose outputs are credible precisely because they are highly scrutinised and constantly subject to further verification.*

As we stressed in this report, it is crucial to the future of science and its role in society to emphasise engagement of publics in research process, in ways that are informed, constructive and not instrumentalised in ways that are damaging to the rigor and reliability of research. To this end, there needs to be investment in a science communication strategy that includes both a strong commitment to (and venues for) the co-production of scientific knowledge, for instance through citizen science initiatives and public engagement programmes, and effectively provides information crucial to public life (e.g. actionable advice in case of global emergencies such as the COVID-19 pandemic).

Conclusions: How Does the Pursuit of Reproducibility Help Us Address the Science Crisis?

We have argued that the pursuit of reproducibility as an overarching epistemic value is not a magic formula for developing or assessing scientific research. In and of itself, improving reproducibility does not necessarily 'fix' concerns around research quality, since it does not address systemic issues with rewards and incentives for research, and does not provide a universal solution to methodological issues, since reproducibility means different things to different fields and research approaches.

However, examining the 'crisis of reproducibility' provides an excellent opportunity to explore systemic concerns in the current system of scientific knowledge production, and particularly to address the lack of support and resources for researchers to explicitly and regularly discuss:

1. methodological commitments within and across disciplines, and beyond academia.
2. how learning from mistakes and problems happens in everyday practice - and is documented in ways that are intelligible and actionable by others (including by privately funded research).
3. the strategies used to choose which research components need to be preserved in the long term, and how.
4. how Open Science can be implemented to support and improve ongoing research practices, and be geared towards community participation and societal benefit.
5. how credit systems in science need to recognise contributions by early career researchers, technicians and support staff, and emphasise long-term outcomes over short-term gains.

Among the means available to achieve such goals, there is the newly formed Belgian Reproducibility Network, which – if well-used – could serve as a catalyst for discussions among relevant stakeholders and therefore for progress on all of these fronts. It is also crucial for research institutions to bring in support structures and transform their systems of workload allocation to match the training and new demands placed on researchers through reproducibility and Open Science policies. New developments in Artificial Intelligence (AI) could help foster more effective ways of checking the soundness of research, at least in some domains. There have been exciting recent developments in how AI can assist with scientific tasks such as assisting with evidence synthesis for policy makers (Porciello, Ivanina, Islam, Einarson, & Hirsh, 2020).

The over-arching goal of the Flemish research community should be to "make reproducibility talk" unnecessary. Scientists no longer debate continental drift, the link between HIV and AIDS, or the existence and cause of climate change.

Consensus has been achieved on those issues and so, perhaps somewhat ironically, scientists no longer debate them because continuous scrutiny of well-confirmed ideas would be wasting valuable time. Once we have achieved a sufficient standard of reproducibility, it, too, should no longer attract much attention as a topic of discussion.

References

Allison, D. B. et al (2016). A tragedy of errors. *Nature*, *530*, 27-30.

Baba, A., Cook, D. M., McGarity, T. O., & Bero, L. A. (2005). Legislating “sound science”: The role of the tobacco industry. *American Journal of Public Health*, *95,* S20-S27. doi: 10.2105/AJPH.2004.050963.

Bailes, M., Bates, S., Bhalerao, V., Bhat, N., Burgay, M., Burke-Spolaor, S., . . . van Straten, W. (2011). Transformation of a star into a planet in a millisecond pulsar binary. *Science*, *333,* 1717-1720.

Bartling S, Friesike S (eds., 2014), *Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Springer.

Beaulieu, A. and Leonelli, S. (2021) *Data and Society: A Critical Introduction*. London, UK: SAGE. https://uk.sagepub.com/en-gb/eur/data-and-society/book 269709

Bezuidenhout L, Leonelli S, Kelly A and Rappert B (2017) Beyond the Digital Divide: Towards a Situated Approach to Open Data. *Science and Public Policy* 44(4): 464-475

Boulton G et al (2012) *Science as an Open Enterprise.* 02/12. London: The Royal Society Science Policy Centre.

Bringmann, L. F., Elmer, T., & Eronen, M. I. (2022). Back to basics: The importance of conceptual clarification in psychological science. *Current Directions in Psychological Science*. DOI: 10.1177/09637214221096485.

Brulle, R. J. (2013). Institutionalizing delay: foundation funding and the creation of US climate change counter-movement organizations. *Climatic Change*, *122,* 681-694. doi: 10.1007/s10584-013-1018-7

Brulle, R. J. (2018). The climate lobby: a sectoral analysis of lobbying spending on climate change in the USA, 2000 to 2016. *Climatic Change*. doi: 10.1007/s10584-018-2241-z

Brulle, R. J., Hall, G., Loy, L., & Schell-Smith, K. (2021). Obstructing action: foundation funding and US climate change counter-movement organizations. *Climatic Change*, *166,* 17. doi: 10.1007/s10584-021-03117-w

Brysse, K., Oreskes, N., O’Reilly, J., & Oppenheimer, M. (2013). Climate change prediction: Erring on the side of least drama? *Global Environmental Change*, *23,* 327-337. doi: 10.1016/j.gloenvcha.2012.10.008

Cataldo, J. K., Bero, L. A., & Malone, R. E. (2010). “A delicate diplomatic situation”: tobacco industry efforts to gain control of the Framingham study. *Journal of Clinical Epidemiology*, *63,* 841-853. doi: 10.1016/j.jclinepi.2010.01.021

Chang, A. C., & Li, P. (2015). Is economics research replicable? Sixty published papers from thirteen journals say “usually not”. *Finance and Economics Discussion Series*,*2015* (83), 1-26. doi: 10.17016/feds.2015.083

Dennis, S., Garrett, P., Yim, H., Hamm, J., Osth, A. F., Sreekumar, V., & Stone, B. (2019). Privacy versus open science. *Behavior Research Methods*, *51,* 1839-1848. doi: 10.3758/s13428-019-01259-5

Burgelman, J et al (2019) Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data* 2, 43.

Derksen and Morawski (2022) Kinds of replication.

Dunlap, R. E., & McCright, A. M. (2011). Organized climate change denial. In J. S. Dryzek, R. B. Norgaard, & D. Schlosberg (Eds.), *The Oxford handbook of climate change and society* (pp. 144-160). Oxford, UK: Oxford University Press.

Ebeling, M., 2016. *Healthcare and Big Data: Digital Spectres and Phantom Objects*. Palgrave

Edwars MA, Roy S (2017) Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. *Environmental Engineering Science* 34, 1.

Elliott, K. (2020) A Taxonomy of Transparency in Science. *Canadian Journal of Philosophy* 1-14.

Elliott, Kevin and David B. Resnik. 2019. “Making open science work for science and society.” *Environmental Health Perspectives* 127:075002.

Eubanks, V., 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Illustrated edition, New York: St. Martin’s Press.

European Commission (2018) *Open Science: Altmetrics and Rewards*. Final Report for the Mutual Learning Exercise Open Science: Altmetrics and Rewards of the European Commission. URL: https://rio.jrc.ec.europa.eu/en/policy-support-facility/mle-open-science-altmetrics-and-rewards

European Commission (2016) *Open Innovation, Open Science, Open to the World. A vision for Europe*, Luxembourg, Publications Office of the European Union.

European Commission (2018) *Open Science: Altmetrics and Rewards*. Final Report for the Mutual Learning Exercise Open Science: Altmetrics and Rewards of the European Commission. URL: https://rio.jrc.ec.europa.eu/en/policy-support-facility/mle-open-science-altmetrics-and-rewards

Fecher, B. and Friesike, S. (2014) Open Science: One Term, Five Schools of Thought. In S. Bartling, S. Friesike (eds.), *Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Springer, 17-47.

Feest, U. (2019) Why Replication Is Overrated. *Philosophy of Science* 86:5, 895-905.

Fernandez Pinto, M. (2020) Open Science for Private Interests? How the Logic of Open Science Contributes to the Commercialisation of Research. *Frontiers in Research Metrics and Analysis* 5: 588331.

Fidler, F. and Wilcox, J. (2018) Reproducibility of Scientific Results. In Zalta E (ed) *The Stanford Encyclopedia of Philosophy*.

Fox Tree, J. F., Lleras, A., Thomas, A., & Watson, D. (2022). *The inequitable burden of open science.* Retrieved from https://featuredcontent.psychonomic.org/the-inequitable-burden-of-open-science/

Freudenburg, W. R., & Muselli, V. (2010). Global warming estimates, media expectations, and the asymmetry of scientific challenge. *Global Environmental Change*, *20,* 483-491. doi: 10.1016/j.gloenvcha.2010.04.003

Garvey, A., & Paavola, J. (2021). Community action on natural flood management and the governance of a catchment-based approach in the UK. *Environmental Policy and Governance*, *32,* 3-16. doi: 10.1002/eet.1955

Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kington, R. (2011). Race, ethnicity, and NIH research awards. *Science*, *333,* 1015-1019. doi: 10.1126/science.1196783

Güttinger S (2020) The Limits of Replicability. *European Journal for Philosophy of Science* 10:10.

Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences*, *113,* 4296-4301. doi: 10.1073/pnas.1516047113

Hornsey, M. J. (2020). Why facts are not enough: Understanding and managing the motivated rejection of science. *Current Directions in Psychological Science*. doi: 10.1177/0963721420969364

Joslyn, S., & LeClerc, J. (2016). Climate projections and uncertainty communication. *Topics in Cognitive Science*, *8*. doi: 10.1111/tops.12177

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2,* 196-217. doi: 10.1207/s15327957pspr0203_4

Kunert, R. (2016). Internal conceptual replications do not increase independent replication success. *Psychonomic Bulletin & Review*, *23,* 1631-1638. doi: 10.3758/s13423-016-1030-9

Lash, T. L. (2021). Getting over TOP. *Epidemiology*, *33* (1), 1-6. doi: 10.1097/ede.0000000000001424

Leonelli S (2023) *The Philosophy of Open Science*. Cambridge University Press.

Leonelli, S (2022) Open Science and Epistemic Pluralism: Friends or Foes? *Philosophy of Science.* https://doi.org/10.1017/psa.2022.45

Leonelli, S. (2021) Data Science in Times of Pan(dem)ic. *Harvard Data Science Review* 3(1) https://doi.org/10.1162/99608f92.fbb1bdd6

Leonelli S, Lovell B, Fleming L, Wheeler B and Williams H. (2021) From FAIR data to fair data use: Methodological data fairness in health-related social media research. *Big Data and Society* 8 (1) DOI: 10.1177/20539517211010310

Leonelli S (2018) Re-Thinking Reproducibility as a Criterion for Research Quality. *Research in the History of Economic Thought and Methodology*. Volume 36B, 129-146.

Leonelli, S. (2017) *Biomedical Knowledge Production in the Age of Big Data.* Report for the Swiss Science and Innovation Council, published online November 2017: http://www.swir.ch/images/stories/pdf/en/Exploratory_study_2_2017_Big_Data_SSIC_EN.pdf

Leonelli, S., Davey, R., Arnauld, E., Parry, G. and Bastow, R. (2017) Data Management and Best Practice in Plant Science. *Nature Plants* 3, 17086. doi: 10.1038/nplants.2017.86

Levin N, Leonelli S et al (2016) How Do Scientists Understand Openness? Exploring the Relationship between Open Science Policies and Research Practice. *Bulletin for Science & Technology Studies* 36: 128-41.

Levy, K. E., & Johns, D. M. (2016). When open data is a Trojan Horse: The weaponization of transparency in science and governance. *Big Data & Society*, *3* . doi: 10.1177/2053951715621568

Lewandowsky, S. (2021). Climate change, disinformation, and how to combat it. *Annual Review of Public Health*, *42,* 1-21. doi: 10.1146/annurev-publhealth-090419-102409

Lewandowsky, S., & Bishop, D. (2016). Research integrity: Don't let transparency damage science. *Nature*, *529,* 459-461. doi: 10.1038/529459a

Lewandowsky, S., Mann, M. E., Brown, N. J. L., & Friedman, H. (2016). Science and the public: Debate, denial, and skepticism. *Journal of Social and Political Psychology*, *4,* 537-553. doi: 10.5964/jspp.v4i2.604

Lewandowsky, S., & Oberauer, K. (2016). Motivated rejection of science. *Current Directions in Psychological Science*, *25,* 217-222. doi: 10.1177/0963721416654436

Lewandowsky, S., & Oberauer, K. (2020). Low replicability can support robust and efficient science. *Nature Communications*, *11,* 358. doi: 10.1038/s41467-019-14203-0

Lewandowsky, S., Oreskes, N., Risbey, J. S., Newell, B. R., & Smithson, M. (2015). Seepage: Climate change denial and its effect on the scientific community. *Global Environmental Change*, *33,* 1-13. doi: 10.1016/j.gloenvcha.2015.02.013

Love, A. (2020) Developmental Biology. *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/archives/sum2022/entries/biology-developmental/

Lowes, S., & Montero, E. (2021). The Legacy of Colonial Medicine in Central Africa. *American Economic Review*, *111,* 1284-1314. doi: 10.1257/aer.20180284

Ma, M. A., Gutiérrez, D. E., Frausto, J. M., & Al-Delaimy, W. K. (2021). Minority Representation in Clinical Trials in the United States: Trends Over the Past 25 Years. *Mayo Clinic Proceedings*, *96,* 264-266. doi: 10.1016/j.mayocp.2020.10.027

Mak, W. W. S., Law, R. W., Alvidrez, J., & Pérez-Stable, E. J. (2007). Gender and ethnic diversity in NIMH-funded clinical trials: Review of a decade of published research. *Administration and Policy in Mental Health and Mental Health Services Research*, *34,* 497-503. doi: 10.1007/s10488-007-0133-z

McDonald, H. (2014). *Boston College tapes fiasco harms search for truth in Troubles.* Retrieved from https://www.theguardian.com/politics/2014/may/06/boston-college-tapes-fiasco-northern-ireland-troubles

Mirowski, Philip. (2018). The future of (open) science." *Social Studies of Science* 48(2):171-203.

National Academies of Sciences, Engineering, and Medicine 2019. Reproducibility and Replicability in Science. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303

Nguyen, C. Thi (2021) Transparency is Surveillance. *Philosophy and Phenomenological Research* 00: 1-31

Noble, S.U., 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. Illustrated edition, New York: NYU Press.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*, 2600-2606. doi: 10.1073/pnas.1708274114

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, *26,* 1596-1618. doi: 10.3758/s13423-019-01645-2

Onghena, P. (2020) *Repliceerbaarheid in de KVAB empirische menswetenschappen*. KVAB Standpunt 64, 2020.

Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science*, 349(6251), aac4716. doi:10.1126/science.aac4716.

Open Science Policy Platform (2020) *Progress on Open Science: Towards a Shared Research Knowledge System. Final Report of the Open Science Policy Platform*: European Commission, Directorate-General for Research and Innovation. URL: https://ec.europa.eu/research/openscience/pdf/ec_rtd_ospp-final-report.pdf#view=fit&pagemode=none doi: 10.2777/00139 Romero F (2019) Philosophy of Science and the Replicability Crisis. *Philosophical Compass* 14: e12633.

Oreskes, N. (2019) *Why Trust Science?* Princeton, NJ: Princeton University Press.

Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt*. London, UK: Bloomsbury Publishing.

Porciello, J., Ivanina, M., Islam, M., Einarson, S., & Hirsh, H. (2020). Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. *Nature Machine Intelligence*, *2,* 559-565. doi: 10.1038/s42256-020-00235-5

Reardon, S. (2011). Decrying CIA Vaccination Sham, Health Workers Brace for Backlash, 333, 395-395. doi: 10.1126/science.333.6041.395

Rubin, M. (2017). When does HARKing hurt? identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. Review of General Psychology, 21, 308-320. doi: 10.1037/gpr0000128

Rubin, M., & Donkin, C. (2022). Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. Philosophical Psychology, 1-29. doi: 10.1080/09515089.2022.2113771

Schwartz, J. (2018). "transparency" as mask? the EPA's proposed rule on scientific data. New England Journal of Medicine. doi: 10.1056/NEJMp1807751.

Science, Digital et al. (2022): *The State of Open Data 2022*. Digital Science. Report. https://doi.org/10.6084/m9.figshare.21276984.v5

Radder (1996) *In and About the World*. New York: SUNY Press.

Strømland, E (2019) Preregistration and reproducibility. Journal of Economic Psychology, 75, Part A, 102143, https://doi.org/10.1016/j.joep.2019.01.006

Taylor L., 2017. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society,* 4(2), pp.205395171773633.

Tempini, N., & Leonelli, S. (2018). Concealment and discovery: The role of information security in biomedical data re-use. *Social Studies of Science*, *48,* 663-690. doi: 10.1177/0306312718804875

Thaler, K. M. (2021). Reflexivity and Temporality in Researching Violent Settings: Problems with the Replicability and Transparency Regime. *Geopolitics*, *26* (1), 18-44. doi: 10.1080/14650045.2019.1643721

United Nations (2019) Towards Global Open Science: Core Enabler of the UN 2030 Agenda. https://research.un.org/ld.php?content_id=50970566

Weber, M. (2022) *Philosophy of Developmental Biology*. Cambridge: Cambridge University Press.

Whatmore, S. J., & Landström, C. (2011). Flood apprentices: an exercise in making things public. *Economy and Society*, *40,* 582-610. doi: 10.1080/03085147.2011.602540

Worrall, J. (2014). Prediction and accommodation revisited. *Studies in History and Philosophy of Science Part A*, *45,* 54-61. doi: 10.1016/j.shpsa.2013.10.001

Yates, I., Byrne, J., Donahue, S., McCarty, L., & Mathews, A. (2020). Representation in clinical trials: A review on reaching underrepresented populations in research. *Clinical Researcher, 34* (7).

Zuboff, S., 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 1st ed. New York: Public Affairs.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2017). Making replication mainstream. *Behavioral and Brain Sciences*, *41,* E120. doi: 10.1017/S0140525X17001972

Annex 1 CV of the two Thinkers:

Stephan Lewandowsky, University of Bristol, Chair in Cognitive Psychology, with special interest in the way misinformation in society can persist, in how myths and misinformation can spread, and in the variables that determine whether or not people accept scientific evidence, for example surrounding vaccinations or climate science. Remarkable contributions:

– Lecture "Being open but not naked" at the 6th World Conference on Research Integrity in Amsterdam (2017),: https://www.youtube.com/watch?v=3A8QUSSPPhk
– Paper together with Klaus Oberauer "Low replicability can support robust and efficient science" in Nature Communications, January 2020 https://www.nature.com/articles/s41467-019-14203-0
– A strong presentation at the ALLEA workshop Ethical aspects to open science in Brussels in early 2018

Sabina Leonelli, University of Exeter, Turing Institute, active in Global Young Academy, Open Science Group, philosopher of science, Co-Director of the Exeter Centre for the Study of the Life Sciences (Egenis), leading the Data Studies research strand; Turing Fellow at the Alan Turing Institute in London. Editor-in-Chief of the international journal History and Philosophy of the Life Sciences, together with Professor Giovanni Boniolo, and Associate Editor for the Harvard Data Science Review. Vice President of the European Philosophy of Science Association, steering committee member of the Society for the Philosophy of Science in Practice, Plan S ambassador.

In 2021-2022, Fellow at the Wissenschaftskolleg zu Berlin, conducting a project on "Excellence and Diversity in Global Scientific Practice".
https://www.godan.info/news/making-crop-data-sharing-responsible-and-reliable
https://www.yumpu.com/en/document/read/8933819/sabina-leonelli-cv-esrc-genomics-network

Annex 2 The steering group of the Thinkers program

The Thinkers:
Sabina Leonelli, University of Exeter
Stephan Lewandowsky, University of Bristol
Coordinators:
Joos Vandewalle KTW, KU Leuven
Patrick Onghena, KMW, KU Leuven
Staff members:
Inez Dua, staff member KVAB, Logistic and administrative support
Bert Seghers, staff member KVAB, Young Academy support

Members:
Freddy Dumortier, permanent secretary KVAB, UHasselt
Jan De Houwer, UGent
Sadia Vancauwenbergh, UHasselt
Geert Molenberghs, KU Leuven, UHasselt, KAGB
Mieke Van Houtte, UGent, KMW
Willy Verstraete, FWO, KNW, UGent
Geert Leroux-Roels, UZ Gent, KAGB
Inge Van Nieuwerburgh, UGent
Roosmarijn Vandenbroucke, Young Academy and Belgian Reproducibility Network
Joris Vandendriessche, Young Academy and Belgian Reproducibility Network

Cover text

The report explores three aspects of reproducibility with the goal of providing recommendations to stakeholders about how to ensure reproducibility. Replication perse is only one aspect of reproducibility, and to thoroughly understand reproducibility requires casting a wider net that includes components such as transparency, research practices, and the role of theory in science. Second, the relationship between these components and reproducibility can be complex and counterintuitive. Third for science to yield robust, reproducible and credible knowledge, we must transform research institutions, evaluation and practices so that, ultimately, it is no longer necessary to talk about it. We will have achieved reproducibility when we no longer debate it.

Based on their deep insights in the nature and the operational elements of the research process, and extensive conversations with the stakeholders in Flanders, the thinkers have obtained a good perspective on the status of reproducibility of research in Flanders. They appreciate the efforts and initiatives to enhance it, and give recommendations for the research assessment, a broader understanding of reproducibility, the full diversity of stakeholders, building a research culture of open discussion, implement science communication strategies that focus on the engagement of multiple publics into research processes.

RECENT POSITION PAPERS

55. Addendum. Jean-Pierre Henriet. – *Mijlpalen in internationale wetenschappelijke samenwerking*, KVAB/Klassen Natuurwetenschappen, 2017.
56. Piet Swerts, Piet Chielens, Lucien Posman – *A Symphony of Trees. Wereldcreatie naar aanleiding van de herdenking van de Derde Slag bij Ieper, 1917*, KVAB/Klasse Kunsten, 2017.
57. Willy Van Overschée e.a. – *De mobiliteit van morgen: zijn we klaar voor een paradigmawissel*?, KVAB/Klasse Technische Wetenschappen, 2018.
58. Tinne De Laet e.a. - *"Learning Analytics" in het Vlaams hoger onderwijs*, KVAB/Klasse Technische Wetenschappen, 2018.
59. Dirk Van Dyck, Elisabeth Monard, Sylvia Wenmackers e.a. – *Onderzoeker-gedreven wetenschap. Analyse van de situatie in Vlaanderen*, KVAB/Klasse Natuurwetenschappen, 2018.
60. Liliane Schoofs – *Doctoraathouders geven het Vlaanderen van morgen vorm*, KVAB/Klasse Natuurwetenschappen, 2018.
61. Luc Bonte, Aimé Heene, Paul Verstraeten e.a. – *Verantwoordelijk omgaan met digitalisering. Een oproep naar overheden en bedrijfsleven, waar ook de burger toe kan/moet bijdragen*, KVAB/Klasse Technische Wetenschappen, 2018.
62. Jaak Billiet, Michaël Opgenhaffen, Bart Pattyn, Peter Van Aelst – *De strijd om de waarheid. Over nepnieuws en desinformatie in de digitale mediawereld*, KVAB/Klasse Menswetenschappen, 2018.
63. Christoffels Waelkens. – *De Vlaamse Wetenschapsagenda en interdisciplinariteit. Leren leven met interdisciplinaire problemen en oplossingen*, KVAB/Klasse Natuurwetenschappen, 2019.
64. Patrick Onghena – *Repliceerbaarheid in de empirische menswetenschappen*, KVAB/Klasse Menswetenschappen, 2020.
65. Mark Eyskens – *Als een virus de mensheid gijzelt. Oorzaken en gevolgen van de Coronacrisis*, KVAB/Klasse Menswetenschappen, 2020.
66. Jan Rabaey, Rinie van Est, Peter-Paul Verbeek, Joos Vandewalle - *Maatschappelijke waarden bij digitale innovatie: wie, wat en hoe?, KVAB - Denkersprogramma 2019*, KVAB/Klasse Technische Wetenschappen, 2020.
67. Oana Dima (auteur), Dirk Inzé, Hubert Bocken, Pere Puigdomènech, René Custers (eds)., *Genoombewerking voor veredeling van landbouwgewassen. Toepassingen van CRISPR-Cas9 en aanverwante technieken*, ALLEA-KVAB/Klasse Natuurwetenschappen, 2020.
68. Marie-Claire Foblets, *De multiculturele samenleving en de democratische rechtsstaat – Hoe vrijwaren we de sociale cohesie?*, KVAB/Klasse Menswetenschappen 2020
69. Joost Van Roost, Luc Van Nuffel, Pieter Vingerhoets e.a., *De rol van gas in de Belgische energietransitie – Aardgas en Waterstof,* KVAB/Klasse Technische Wetenschappen, 2020.
70. Richard Bardgett, Joke Van Wensem, *Bodem als natuurlijk kapitaal* – KVAB Denkersrapport 2020, KVAB/Klasse Technische Wetenschappen, 2021
71. Jos Smits e.a., *Multifunctionele eilanden in de Noordzee*, KVAB/Klasse Technische Wetenschappen, 2021.
72. Elisabeth Monard, red., *Kunst, Wetenschap en Technologie in Symbiose*, KVAB/Klasse Technische Wetenschappen, 2021.
73. Jan Wouters, Maaike De Ridder, *De problematiek van de rechtsstaat en democratische legitimiteit binnen de Europese Unie*, KVAB/Klasse Menswetenschappen, 2021.
74. Hilde Heynen, Bart Verschaffel, e.a., *Architectuurkwaliteit vandaag, Reflecties over architectuur in Vlaanderen*, KVAB/Klasse Technische wetenschappen en Klasse Kunsten, 2021.
76. Bea Cantillon, *Het armoedevraagstuk en de tragiek van de welvaartsstaat, Zeven termen voor een nieuw sociaal contract*, KVAB/Klasse Menswetenschappen, 2022.
78. Jo Tollebeek, Marc Boone en Karel van Nieuwenhuyse, *Een Canon van Vlaanderen, Motieven en bezwaren*, KVAB Klasse Menswetenschappen, 2022.
79. Luc Taerwe e.a., *Duurzaam Beheer van Infrastructuur, Niet alleen een kwestie van budgetten*, KVAB/Klasse Technische Wetenschappen, 2022.
80. Willem Salet, Marleen Spiekman, Staf Roels, Tom Coppens, Ivo Van Vaerenbergh, *Naar klimaatneutrale woongebouwen in 2050, KVAB Denkersprogramma 2022*, KVAB/Klasse Technische Wetenschappen, 2022.

The report explores three aspects of reproducibility with the goal of providing recommendations to stakeholders about how to ensure reproducibility. Replication perse is only one aspect of reproducibility, and to thoroughly understand reproducibility requires casting a wider net that includes components such as transparency, research practices, and the role of theory in science. Second, the relationship between these components and reproducibility can be complex and counterintuitive. Third for science to yield robust, reproducible and credible knowledge, we must transform research institutions, evaluation and practices so that, ultimately, it is no longer necessary to talk about it. We will have achieved reproducibility when we no longer debate it.

Based on their deep insights in the nature and the operational elements of the research process, and extensive conversations with the stakeholders in Flanders, the thinkers have obtained a good perspective on the status of reproducibility of research in Flanders. They appreciate the efforts and initiatives to enhance it, and give recommendations for the research assessment, a broader understanding of reproducibility, the full diversity of stakeholders, building a research culture of open discussion, implement science communication strategies that focus on the engagement of multiple publics into research processes.

The Academy's Standpunten series (Position papers) contributes to the scientific debate on current social and artistic topics. The authors, members and working groups of the Academy, write in their own names, independently and in full intellectual freedom. The quality of the published studies is guaranteed by the approval of one or several of the Academy's classes.

Vlaamse overheid

RICHTING MORGEN
MET WETENSCHAP, TECHNOLOGIE, CREATIVITEIT & INNOVATIE