Eva Barlösius, Uljana Feest, Torsten Wilholt (Leibniz Universität Hannover): From Raw Data to Primary Data: Research Practices, Tacit Assumptions and Replicability

Abstract: One wide-spread response to the replication crisis has been to call for more transparency in the scientific process. Attempts to institute such transparency range from methodological reforms to calls for pre-registered studies and changes in review-and publication policies (e.g., Munafo et al. 2017). Methodological reforms are mostly geared towards preventing p-hacking, i.e., the questionable practice of inflating the statistical significance of a result by manipulating the data in various ways. One suggestion put forth to counteract p-hacking is to require scientists to share their data in public repositories. While we believe these suggestions to be valuable and important, we challenge two assumptions at the heart of calls for methodological reform, i.e., (1) the assumption that the scientific process can be exhaustively described and governed by means of explicit rules and (2) the assumption that there can be such a thing as “neutral” data.

With regard to the first point, we highlight the existence of knowing-how and material/social practices that form a central aspect of scientific research, also sometimes referred to as “tacit knowledge” (e.g., Polanyi 1958, Kuhn 1961; Feest 2016). Following Polanyi, we conceptualize such tacit knowledge as having a social and a personal component: It is shared and passed on within the relevant communities, but certain individuals are recognized as being particularly well-versed at the skill(s) in question. In turn, this means that those individuals are trusted to “do their job well.” Sociologically speaking, it is clear that this kind of trust has been shaken at its core in recent times. However, we argue that such a reliance on epistemic trust cannot be eliminated from the scientific process altogether (cf. Wilholt 2013).

Touching on our second point, we will focus on a particular scientific skill, namely that of transforming raw data into “primary data” (Schönbrodt et al. 2015), a process that Leahey (2008) has referred to as “data editing,” and which she singles out as relying on tacit knowledge. This process is one in which researchers add, drop, or change features of the immediate outputs of scientific measurements in order to create data sets that can be shared and subjected to statistical
analysis. Data editing, we argue, is local and context-dependent in several respects, in that it (a) presupposes tacit knowledge with respect to the specifics of the instruments used, and (b) typically is done before the background of the purpose/question for which the data were generated. Taken together, these two aspects of context-dependence call into doubt the very notion that a shared data base of neutral primary data is a realistic aim.

We will situate our analysis within existing sociological analyses about data sharing practices (e.g., Barlösius et al 2018) and the philosophy and sociology of tacit/practical knowledge as well as social epistemology and the emerging field of the philosophy of data. In addition, some of our key claims will be illustrated by quotes from qualitative interviews with scientists. Our conclusion should not be taken as expressing a generalized skepticism about the possibility of overcoming the replication crisis. We do argue, however, that there will not be a purely mechanical fix for the crisis.

References
Kuhn, Thomas. 1962. The Structure of Scientific Revolutions. Chicago: The University of Chicago Press
Munafò, Marcus, Brian A. Nosek, Dorothy Bishop, Katherine Button, Christopher Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer Ware, and John Ioannidis. 2017. A manifesto for reproducible science. NATURE HUMAN BEHAVIOUR 1, 0021, DOI: 10.1038/s41562-016-0021
Tristan Berger (Université Paris 1 Panthéon-Sorbonne et ISJPS): Regulatory science and the replication crisis: A colossus with feet of clay

Abstract: In view of the dangers they are likely to entail, certain products may not be placed on the market without a scientific assessment of the health and environmental risks. For example, genetically modified organisms (GMOs), human and veterinary medicinal products, chemical substances or food additives must be the subject of toxicological, ecotoxicological or pharmacological studies. The public authorities then use these studies as a basis for accepting or refusing to place the product on the market. This “hybrid activity that combines elements of scientific evidence and reasoning with large doses of social and political judgement” is what Sheila Jasanoff calls regulatory science.

The economic, health and environmental stakes involved in decisions in this area are sometimes colossal. The question of the reproducibility of studies is therefore particularly acute. All the more so as cases of fraud have been detected, creating, according to the French Agency for the Safety of Health Products (Agence française de sécurité sanitaire des produits de santé), “a certain doubt as to the reliability of the data submitted to the competent national authorities”. However, it is extremely difficult to reproduce these studies, for two reasons: on the one hand, it is extremely expensive and, on the other hand, the raw data from these studies are not systematically accessible, as the rules vary from one legal system to another.

While the implementing Commission Regulation (EU) No 503/2013 of 3 April 2013 on GMOs, as well as Regulation (EU) 2019/6 of the European Parliament and of the Council of 11 December 2018 on veterinary medicinal products, and Regulation (EC) No 1331/2008 of the European Parliament and of the Council of 16 December 2008 on additives, food enzymes and flavourings provide for public authorities to have access to raw data from studies, this is not the case for Directive 2001/83/EC on medicinal products or Regulation (EC) No 1907/2006 on chemicals, which did not provide for systematic access to raw data from evaluations. The latter may then be protected under business secrecy, preventing the reproduction of the studies transmitted to the authorities and consequently questioning their reliability. It thus appears that, far from being limited to the academic sphere, the reproducibility crisis is also affecting regulatory science, more particularly the fields in which the raw data of studies are likely to be protected under business secrecy. This crisis seems to pave the way for a gradual increase in the level of transparency, as evidenced by the Seralini affair and subsequent regulatory developments in the field of GMOs.

Florian Charvolin (CNRS, Centre Max Weber, Lyon): The issue of transparency of biodiversity data in citizen science. To blur or not to blur information?

Abstract: In the biodiversity sciences, the question of replicability consists in the capacity given to amateurs and professionals to go back to the sites where species have once been spotted. It is also to be able to use the data made available on the internet for new researches. Most of citizen sciences are organized as an operation chain going from the naming of the species in outings, the reporting
of the sight on internet sites, and the database processing of it. I will tackle the example of Faune France a national database of 70000 contributors and millions of sightings of Fauna, to introduce the controversy among its staff members on the necessity of transparency, conceived as open access to raw data with all the metadata.

Faune France administrators are wary on the issue, and different takes can be noticed on the necessity of blurring the data, to give an information "degradée" (downgraded or degraded in English) to the public. In some cases the argument is to protect the spotted species from any inconvenience of being visited by many people, for example if they are nesting. In other cases there is a reluctance from amateur organizations to give the information that they so harshly collected to private research bureaux ("bureaux d'étude" in French) that will take advantage of them to sell reports to public administrations. In that case there may be a competition between private research bureaux and the member's associations of Faune France which sometimes also sell reports to public administrations. It is also argued that raw data contain errors by the very means that they are collected by amateurs, and they shouldn't be given without previous quality control by the organization banking the data. Last but not least some of these associations, claim that everybody should follow the Aarhus Convention and set as a principle to give the open access to all the data collected.

The communication will unfold the different arguments present in the controversy, and will stress the entwined nature of their economic, social and epistemological nature.

Benjamin Derbez (Université de Bretagne Occidentale / Labers) & Meoïn Hagège (Université Paris-Est Créteil / CEpiA Inserm U955): Biological Universality and Replicability: The Case of Elderly Subjects in Clinical Oncology Research

Abstract: In the field of biomedical research, there are two meanings for the word replicability. In the first sense it refers to the reproducibility of the results of a clinical trial by means of its repetition under identical conditions. In a second sense, replicability can refer to the applicability of knowledge produced in an experimental setting to the real life conditions of clinical practice; in this case, a clinical patient does not always resemble the one tested in a clinical trial. While the question of replication is generally overcome by the statistical calculation of confidence intervals, the question of translation is otherwise a daunting one for clinicians who may be confronted with the gap between published clinical trial data and the therapeutic responses observed in their patients. This situation calls into question the presupposition of the "biological universality" of bodies (Lock and N'Guyen, 2010). This is particularly the case in oncology when the patients treated are considered "older". In this paper, we propose to focus on the issue of producing standards of cancer treatment for patients undergoing geriatric care. We will first show how the evolution of biomedical research practices and standards has led to an under-representation of older subjects in cancer clinical trials. This under-representation is problematic for clinicians because patients over 65 years of age represent more than half of the population treated for cancer in France. Based on
interviews conducted with oncologists and geriatricians as part of the GerOncoSAge sociological study on practices of inclusion of older patients in clinical trials, we will highlight the translation difficulties that these professionals face in their medical activity. Finally, we will rely on the results of these same interviews and question the relevance of the hierarchy of levels of evidence accepted in the paradigm of evidence-based medicine. The statistical principle underlying the production of evidence, also known as "all other things being equal", cannot be applied heuristically to the clinic with older subjects, as they are a particularly heterogeneous population, often with several pathologies and lower tolerance for cancer treatments. This paper discusses the issue of whether the production of medical knowledge for geriatric patients requires to revisit the reliance on randomized clinical trials as references in establishing causality.

Hugh Desmond (KU Leuven, University of Antwerp): Incentivizing Replication is Insufficient to Safeguard Default Trust

Abstract: When thinking about the social structures of science, philosophers of science and meta-scientists have for some time now predominantly adopted an ‘economic approach’, where scientists are modeled as credit-maximizing agents responding to incentives such as promotion, funding, or publication criteria (Kitcher 1990; Strevens 2006; Higginson and Munafò 2016; Smaldino and McElreath 2016; Heesen 2018; Holman and Bruner 2017; O’Connor 2019).

Yet in applied ethics, sociology of science, and to a large extent actual science policy making, an ‘ethical approach’ informs research on social structures of science: scientists are predominantly understood to be agents concerned with ideals such as honesty, respect, or reliability, and are capable of acting contrary to incentive structures (Carvalho 2017; Desmond 2019; ESF-ALLEA 2017; Forsberg et al. 2018; Godecharle et al. 2013).

Philosophers of science and meta-scientists do not openly dismiss the ethical approach. Yet, it does often not seem so clear what precisely, if anything, the ethical approach brings to the explanatory table that cannot be covered by the economic approach. For instance, concern with honesty could be explained as minimizing the expected penalties (negative credit) following a strategy of dishonesty; concern for reliability could be explained as maximizing replicable studies, which can be modeled as having a higher pay-off than non-replicable studies (as in Heesen 2018).

What is it, if anything, that prevents one from taking a cynical stance on the ideals of individual scientists, i.e., that they are mere window-dressing, ineffectual against the brutal reality of credit-maximization? Indeed, the view that scientists should be credit-maximizers, in the interests of scientific progress, seems viable (Kitcher 1990).

This paper will suggest that the explanatory limits of the economic approach are reached concerning the phenomenon of default trust between scientists. This will be defined in detail later in the talk, but the core idea is that scientists tend to believe that their colleagues are telling the truth – or are at least attempting to do so. This default trust in each others’ assertions underlies
many core scientific behaviors – I consider peer review and collaboration as illustrative examples – and thus may be considered integral to scientific research.

Given widespread problems with reproducibility and replicability (Baker 2016) – or at least, the perception of such problems (Fanelli 2018) – such default trust can be said to be under pressure. I then consider one of the most important proposed policy reforms to the credit-based incentive structure of science: incentivizing replication research. Replication research basically acts to disincentivize low-credence assertions. Can it safeguard default trust? Using an expanded version of Heesen's model of when replicable (or trustworthy) assertions maximize credit, I will then show that, no matter how much replication research is incentivized, default trust cannot be justified in a culture of credit-maximization.

The upshot is that, on the descriptive side, the economic model cannot account for an important explanandum concerning scientific practice (i.e., default trust). On the normative side, it means that in a culture of extreme credit-maximization, default trust between scientists would ultimately be eroded, and replaced with a ‘default lack of trust’ (or default distrust), and this would be detrimental to science.


Andreas Diekmann (Institute of Sociology, University of Leipzig and Department of Humanities, ETH Zurich): Irreproducible results. Fraud, error, and the importance of replication

Abstract: Once again the Lancet science scandal with Covid-19 publications has shown that the scientific publishing system is vulnerable to fraud. But not only falsified data can destroy trust in the validity of science. Systematic errors and distortions undermine trust far more. How is it possible to detect fraudulent and erroneous publications and what can we do about it? The protection of whistle blowing, statistical methods of fraud detection, reforms of the publication system, and, above all, the promotion of replications are necessary measures to strengthen the credibility of scientific results.

David Pontille and Didier Torny (Centre de sociologie de l’innovation, I3, UMR 9217, CNRS): When replication is prohibited

Abstract: The replication crisis label presupposes the desirability of reproducing a phenomenon or result as an ideal of science regulation. In line with a Mertonian philosophy of science (Merton, 1973), advocates of reproducibility consider replication to be an essential criterion for the validation of scientific statements. However, the implementation of such an objective is variable according to scientific fields, with a reproducibility gradient between the pure reproduction of a computer code in certain data-intensive sciences (Kitzes et al., 2017) and the possibility of following the same interpretative path from the same materials in humanities (Leonelli, 2018).

Moreover, both studies on experimental practices (Collins, 1992) and on fraud (Broad & Wade, 1982) have long since shown that the core of scientific practices consists in producing newness and, as a result, replication activity is very marginal, if not non-existent, and certainly not valued in the careers of researchers. On the contrary, it is in the field of scientific misconduct that quasi-industrial practices of reproduction have emerged where there used to be lots of craftsmanship (Biagoli, 2012). On this basis, many companies and independent scientists have developed tools for detecting plagiarism in texts, for detecting the reprocessing and partial reproduction of images, and, through statistical tools, for manufacturing by non-random repetition and distribution of digital data. But it is with the discovery of a veritable epidemic of photocopied images and the subsequent naming of the paper mill, that is "a shady company that produces scientific papers at demand", that the question of industrial-like scientific reproduction (Bik, 2020) is being raised. Hundreds of articles containing identical figures, but describing different phenomena, were thus catalogued, in the manner of copycat student copies from traditional university essay mills.

Our presentation will discuss a variety of tools and ways to determine and qualify (improper) reproduction. We will thus be able to shed original light on the debate on reproducibility by showing how misconduct studies redefine it, valuing variation, discrepancy and difference rather than similarities as the default evidence of scientific production.

https://scienceintegritydigest.com/2020/02/21/the-tadpole-paper-mill/


**Julian Reiss (Johannes Kepler University Linz): On the Use of Prediction Markets in the Replication Crisis**

Abstract: Prediction markets go back at least to bets mediaeval pundits made on who would be the papal successor. A recent application is the replication market, in which anyone can sign up to bid on the chance of studies in the social and behavioural sciences (from 3,000 studies deemed eligible by the Center for Open Science) to replicate, understood as 'getting a statistically significant result in the same direction as the original claim' and 'using data that was not used in the original study' (www.replicationmarkets.com).

The goal of this paper is to discuss the capacity of prediction markets to generate reliable information about the likelihood of future events in general, to assess the application to the replication crisis, and, finally, to address the question whether prediction markets can help to solve the replication crisis. In particular, it will be examined whether this new institution can contribute to realigning researchers’ incentives both to produce replicable results and to invest resources in replication.

**Agnès Robin (Université de Montpellier, UMR 5815): The Part of Law in the Absorption of the ‘Replicability Crisis’**

Abstract: Our contribution proposes to study the question the ‘replicability crisis’ under legal aspects. It means to analyze the regulation linked to replicability of research results. Because law is not properly or directly concerned by the question of results’ replicability, except the research
integrity soft law (European code of conduct). Until recently, law was not so concerned by results’ replicability. Mainly through industrial property which commands to protect only operating and industrially useful innovations. In a way the industrial property assure the replicability of scientific results which is evaluated through the eyes of the skilled person in the art (inventive activity). The undirect regulatory function of the industrial property is nevertheless not available on all scientific results but only on those which can pretend to enter the market. As sociologists have already written (Pontille, 2016), the author’s Law can’t properly take in charge the question of the scientific responsibility because of the link, too thin, between the scientific “authorship” conception and the author’s Law conception and despite the efforts of scientific editors consortium (e.g. ICMJE) to force researchers to be transparent and accurate about functions of the participants on a research project. Nevertheless an evolution can be observed through the French and European “open science” policy, forcing as possible the dissemination and the archiving of research data which allows peers to access the data and verify results. The “open science” policy which follows the objective to make data reusable contributes to the science data movement and to the production of combined data (“second hand data” or secondary data). The scientific integrity question, and its underlying question (scientific responsibility), comes up in a particular way considering that the ‘big data analyze’ is operating a split between the genesis’ context of data and data whose intelligibility can be lost (Bachimont, 2016). The question is all the more acute because of the emergence of spontaneous actors (PubPeer) who nowadays increase the reputational risk of researchers but contributes to provide the proof of frauds to public authorities.

Felipe Romero (University of Groningen): The Many Faces of Scientific Self-Correction

Abstract: Authors have made progress studying causes of the replicability crisis and proposing solutions. However, one critical philosophical issue that remains unaddressed is how philosophical theories about scientific self-correction are affected by the crisis. The goal of this paper is two-fold. First, argue that the replicability crisis reveals that the statistical view of scientific self-correction is insufficient. The crisis shows that the human, social, and institutional aspects of science introduce errors that statistical inference alone cannot correct. Second, I present an analysis of scientific self-correction as a process spanning multiple levels on top of statistical self-correction. I show the payoffs of the proposed analysis in the context of psychology and the life sciences.

Silvia Tossut (Department of Philosophy and Educational Sciences – University of Turin; MSIO project; LLC research) : The Psychological Science Accelerator: An Epistemological Perspective

Abstract: §1. The Psychological Science Accelerator (PSA) is a recent attempt to overcome the replicability crisis in psychology. In this paper I provide an epistemological analysis of the PSA, to assess whether it can be a way out the crisis. The PSA is a distributed network of labs, construed on a Cristopher Chartier’s intuition and designed to enable and support crowdsourced research
projects. In order to face the crisis, Chartier stated the necessity of joining efforts and collective work, to accelerate the accumulation of field-specific knowledge. Interestingly, the PSA functions as a replication tool, but its focus is the acceleration of psychological science, through open, full and transparent communication.

§2. The PSA relies on the idea that communication is central, since openness and transparency are crucial virtues of scientific projects. The online platform, the institutional design, the democratic procedures and the network of scholars constitute the PSA that, as a whole, provides the infrastructure needed to accelerate rigorous psychological science.

I argue that the PSA needs an extra assumption explaining how the acceleration connects to the issues due to replicability. I suggest that the sense in which the PSA can be a way out the replicability crisis has little or nothing to do with acceleration.

§3. A first set of considerations concerns the possibility for transparent communication to obtain in huge groups. As to now, the PSA network includes 70 labs, which are supposed to realize direct, democratic communication on the online platform of the project. Still, it is hard to think that so many people can communicate in a way similar to the personal communication. Communication within a cooperating team seems to have different features, especially with respect to some epistemic relevant element, in particular mutual trust. Epistemology of testimony as well as ontological analyses of groups can be of help in singling out problems in this area.

§4. A powerful tool to examine the possible effects of the PSA’s communicative structure is Zollman’s network epistemology. According to the results of Zollman’s simulations, full communication among the members of a network is not the best communicative structure for scientific communities, since fast convergence on a results comes at the cost of the loss of reliability. If Zollman’s conclusion is correct, faster accumulation of knowledge cannot be achieved simply increasing direct communication among the members, because the final result on which the members of the network will converge won’t qualify as knowledge, due to its lack of reliability. Applied to the PSA, this means that the project is doomed to fail as an accelerator of psychological science: widespread, transparent communication does not guarantee that knowledge is finally produced.

§5. In the conclusive part of the paper, I argue that psychological science needs the PSA for it can coordinate efforts, set specific rules and standards to fill the public space with reliable evidence, and provide an injection of trust in the discipline after the discouraging results of previous replication projects. The PSA can be positive for the field, to the extent that it constitutes a collaborative environment, improving the standards for procedures and tests, and enhance the debate concerning the interpretation of massive data.