Normative Influences in Science and their Impact on (Objective) Empirical Research

Kreienkamp, Jannis; Agostini, Max; Kunz, Marvin C.; Meyerhuber, Malte; de Matos Fernandes, Carlos A.

Published in:
Empirical Research and Normative Theory

DOI:
10.1515/9783110613797-007

Jannis Kreienkamp, Maximilian Agostini, Marvin Kunz, Malte Ingo Meyerhuber, and Carlos A. de Matos Fernandes

Normative Influences in Science and Their Impact on (Objective) Empirical Research[1]

Abstract: Empirical research has the ultimate goal to inform us about the "objectively true" state of the world. This ambition especially holds for the natural sciences, but also extends to the social sciences. In the context of recent developments and theoretical discussions, the authors aim to contribute to the discussion of objectivity in empirical research from a junior researcher's perspective, debating the influence of normative assumptions on empirical research. They analyse normative influences within the six conceptual steps of the empirical research process: (1) idea generation, (2) research funding, (3) research planning, (4) data collection, (5) data analysis, and (6) scientific output. The authors end with a summary of current directions that may help move to a more reflective, nuanced, and transparent scientific process.

1 Introduction – A Reflexivity Perspective

This chapter on normative influences in empirical research was a result of normative influences on us. All five authors enjoyed an education in the social sciences, varying in degree between psychology and sociology. This means that the examples we will draw upon are examples closely related to these disciplines.

As junior researchers, we are both in the most and least favourable position to write about normative influences; the least favourable because our limited research experience provides us only bounded insight into the normative research culture; and the most favourable because our limited experience leaves us mostly untouched by many of the normative influences of the research culture. We do not offer a complete picture, nor do we pretend to comprehensively understand the snapshot we are presenting. Yet, we attempt to present a new angle, showing how the bigger picture of gaining scientific insights is perceived by a new generation that will continue the venerable tradition of empirical research.

During our university education we were the recipients of lines of thoughts that were influenced by a field in uproar: social science research did not repli-

1 We stipulate that writing this chapter was a collective enterprise and all authors contributed equally to it.

https://doi.org/10.1515/9783110613797-007

cate (Open Science Collaboration 2015), prominent researchers were convicted of fraud (Carey 2011), and the underlying statistical framework of "standard inference" was being challenged (Simmons, Nelson, and Simonsohn 2011). As a result, our education was heavily influenced by a changing field, focussed on teaching us how to avoid the mistakes of the past.

In 2011, Diederik Stapel was accused and later convicted of scientific misconduct for fabricating data (see, e.g., Bhattacharjee 2013). This case stirred up the scientific community, especially at the universities where Stapel was previously employed as a researcher. One of these universities was the University of Groningen, where we were educated to become (empirical) researchers. The realisation of the susceptibility of the scientific community to misconduct influenced academic staff and the teaching methods at our university. In this climate of raised awareness, it was communicated that we, as a new generation of social scientists, are required to understand and safeguard against the exploitation and misuse of empirical research. This facet of our scientific education has influenced our thinking and is central to our discussion of normative influences in, and on, empirical research.

2 Background

According to Richard Dawkins (as quoted by Singh 2004, p. 497), science is the "disinterested search for the objective truth about the material world", offering us insights into the truth of the world we live in and the entities we are. The method of choice for this "disinterested search" – in many fields – is empirical research,[2] generally due to its methodological rigour and adherence to certain standards of scientific conduct. The way for the march of empirical research has only been paved in the 18th-century Enlightenment, when philosophers, like John Locke and David Hume, formed theories that aimed to pave a road towards a world knowable through empiricism. While this view has become somewhat diluted in recent years, many still assume scientific findings to approach objectivity. Such research has not always been in the hands of empirical approaches and may be less justified than is often assumed. Consider, e.g., the work of Kant (1964), who tried to find *a priori* insights to the way our reasoning and our perception of the world function; relying on his pure thinking rather than on methodical empirical observations. Thus, while gaining more influence

2 With "empirical research" we, here, mean the positivistic research tradition that relies on the analysis of (directly and indirectly) observable data.

in scientific inquiries, in this chapter we argue that empirical research cannot stand autonomously for itself; rather it is interwoven with and must acknowledge the political, cultural, and subjective context (i.e., normative influences), especially when interpreting the social world.

Normative influences in, and on, empirical research are a topical debate (see, e.g., the checklist discussed by Munafò and Smith 2018), due to their essentiality for researchers, for those who rely on the empirical findings, and for those who are studied. Two key aspects can be identified in the debate:

– Ought: Should there be normative influences in, and on, empirical research?
– Is: What normative influences exist within empirical research?

In general, the social sciences have been especially ambivalent regarding the first question, exemplifying both sides of the debate. E.g., parts of psychology have striven to move towards an objective, norm-free science. Ambitions in philosophy (*Vienna Circle*), theory (*behaviourism*), and methodology (*randomised controlled experiments*) worked towards general laws and a nomothetic natural science, aiming to make science free from the influence of politics, culture, and the subjective individual that uncovers positivistic truths (Popper 1973, Porter 1995). However, several voices have claimed that value-free science may be impossible (Gergen 1973); e.g., anthropological theories, as well as qualitative research methods have explicitly built upon the subjectivity of study objects and researchers. Whilst the first question (ought) is important, it may be less relevant to the practitioners and users in the field, as most of them are likely to adopt a more pragmatic and instrumentalist position.

In contrast to the first question, the second question of *what kind of normative influences exist* is thus of greater importance to both researchers and practitioners. Accurate scientific knowledge is relevant to the generating field, as well as individuals who experience direct or indirect consequences from it. As an example, the field of empirically informed ethics attempts to improve ethical theories by incorporating empirical results into the theory building process. These theories may then find implementation in medical or elderly care facilities, such as when a new finding is implemented to – potentially – improve elderly care. However, which empirical results are trustworthy? And which should be implemented? While empirical results can assist in advancing knowledge, it is precarious to rely on each piece of empirical data as if it represented the truth (cf. meta-analysis which accounts for multiple pieces of empirical data; see, e.g., Postmes and Spears 1998, Zimbardo 1969). This holds especially for scientific debates in which both sides provide empirical evidence for opposing claims.

The aim of this chapter is to inform philosophers, but also empiricists, about the implications of normative influences on empirical scientific processes. Our goal is not to discourage using empirical results altogether, but to place them into perspective and inspect them within their specific context.

3 The Six Conceptual Stages of Research

In general, we will investigate six *interwoven* stages, which we deem prominent in conducting empirical research and that appear relevant for scholars in general. Specifically, we will discuss: (1) idea generation, (2) research funding, (3) research planning, (4) data collection, (5) data analysis, and (6) scientific output. In the following sections, we will explore how normative assumptions can influence seemingly objective empirical research (for a graphical illustration see: https://www.doi.org/10.17605/OSF.IO/GYDB8).

3.1 Idea Generation

The first step in the research process is to decide on a phenomenon of systematic analysis. In this step, normative influences are often very direct and overt. One direct normative influence on research idea generation is, e.g., the political diversity (or lack thereof) within a scientific field. Most fields have a certain implicit outlook on society and recreate their political perspective through the people they attract (and keep within the field). E.g., when social psychologists recently investigated their field's political diversity, they realised that their field predominantly consisted of liberal voices, which may hinder the advancement of certain social psychology research domains (Crawford et al. 2015, Duarte et al. 2015). Liberal values may be ingrained into research questions and methods and may result in an one-sided approach to (politically) controversial topics. Moreover, the authors claim that conservative voices trying to get into the field experience a hostile climate and – sometimes – outright discrimination. Similarly, philosophy seems to be having issues with getting (and keeping) minorities and women in the field (Lombrozo 2013). The often-implicit political attitude and self-selection can, therefore, be problematic for the general validity of research, as it may strongly influence the kinds of questions scientists (dare to) ask and the results they find.

Apart from direct influences, such as liberal biases, more indirect influences are observable as well. E.g., a tendency in the social sciences is to avoid broad theorising to explain the observed phenomena (Kruglanski 2001); with the development of comprehensive theoretical frameworks being undervalued. The current trend away from broader theories has been referred to as the *Post-Normal*

Science of Precaution (Ravetz 2004).[3] Some argue that this shift leads towards circular research, a scattered field, and has increased the distance to the societal dialogue (Kruglanski 2001, Sarewitz 2016). The fear of theorising is especially apparent on a smaller scale: the model scale. As an example from the psychological discipline, developmental psychologists may focus on the development of the self, neuropsychologists focus on the neurological processes of the self, and social psychologists on the interplay of situational cues and changes of the self. An integrative underpinning of what the self actually consists of, however, is missing. Instead, every domain tends to undertake (mainly) its own theorising; van Zomeren (2016, p. 13) summarises this by arguing that

> without theoretically integrative efforts, science is blind; and a blind science is nothing more than a very large storage container of empirical trivia. It is and does what a very large storage container is and does: it is very large, and it contains and stores things. And within, it is divided into so many sections and subsections that one may spend a lifetime counting or easily lose count.

Notably, this was already highlighted by McGuire (1973), when he argued that only using societally relevant problems, such as a sequential cause-effect model without underlying theory, disregards the complex nature of human behaviour. The apparent lower model scale would, therefore, benefit from broad interdisciplinary theorising that is able to span the different subdomains, achieving successful integration. In turn, an additional effort integrating the Post-Normal Science of Precaution might also help showcase and address the normative inconsistencies of small scale models, methods, and sub-disciplines.

3.2 Funding

While pursuing a promising research idea, researchers discover they are dependent on grants from public, commercial, or private funding sources to realise their research. The funding of a promising research idea has become an increasing concern for both experienced and aspiring researchers. In 2006, biochemist Roger Kornberg received the Nobel Prize for his research on the copy process of information in DNA. Following this award, he declared before the United States Senate that the current funding practices hinder important research projects and discourage scientists (Edwards and Roy 2017). Kornberg stated that his

3 Note that Ravetz sees this shift as positive, opening doors for new interdisciplinary collaborations.

fundamental research would have been impossible if he was raised a decade later, within the current funding environment. The following paragraphs will illustrate several normative influences on the funding process.

The allocation of funds is usually based on the quality of scientific proposals through peer review. This system may, however, be an ineffective method of financial redistribution. Some scientists claim that due to restrictions in review panels' size and available time, most panels are unable to accurately rank the quality of research proposals (Fang, Bowen, and Casadevall 2016). Additionally, 85 percent of reviewers in the field of medicine report an absence of training in reviewing grant proposals. For allocating funds in empirical research, a typical study selection is commonly limited to two or three reviewers reading a study proposal in-depth, a number too low to provide an acceptable level of accuracy (Kaplan, Lacetera, and Kaplan 2008). Funding and review panels thus often appear to lack the effective structure to adequately judge the quality of research proposals – introducing their personal normative agendas as a heuristic to decide between the large numbers of studies.

The problems with funding extend beyond individual and collective normative influences of review panels. Edwards and Roy (2017) criticise academia's contemporary incentive and reward structure for fostering unhealthy competition and unethical behaviour. They argue that a focus on quantitative performance metrics (such as the *h*-index) results in these indices being a target rather than a measure. According to *Goodhart's Law* (see, e. g., Elton 2004), a measure that becomes a target is no longer an adequate measure. For researchers, many factors such as hiring, promotion, funding, and awards depend on quantitative measures, which is why researchers are pressured to emphasise quantity of publications over quality. One consequence of this increasing pressure may be the temptation of questionable research practices (QRPs). The usage of QRPs gives researchers a competitive advantage over their colleagues, wherewith they more often publish successfully, and attract new funding. This results in a feedback loop, further promoting the application of QRPs and promising researchers utilising these practices an advantage. Moreover, according to Edwards and Roy, incentive structures in academia often lead researchers and other stakeholders to neglect the existence of QRPs. As long as the funding and incentive structure promotes and rewards the usage of QRPs, engaging in these practices will improve a researcher's chance to receive funding.

Cushman and colleagues (2015) illustrate further problematic funding structures. In their analysis, they reveal that although proposal quality, proposer demographics, requested amount of funding, and number of submitted proposals per researcher remained unaltered, the likelihood of receiving funding has substantially decreased for the last decades (see also Alberts et al. 2014). This is consistent with the observation that funding budgets remained stagnant or decreased, while the number of researchers has rapidly grown (Kimble et al. 2015). According to Cushman and colleagues, the likelihood of acquiring funding per application might decrease to or below six percent if the current trend continues. The six percent threshold illustrates a tipping point at which the time required to write proposals will take more working hours than the grant's financial payoff will allow a researcher to continue working. This scenario turns the process of applying for scientific funding into an endeavour that leaves scientists wondering whether they can afford to pursue their own interests and forces them to satisfy the expectations of external funding agencies. This could result in researchers serving the interest of private corporations for the sake of future funding (and thus financial stability; see, e.g., Sismondo 2008), or a focus on irrelevant in vogue topics. It should be noted that proponents of the currently existing funding structure state that the competition between researchers will subsequently provide better results and better research, similar to competition on the free market. A comparison between countries, funding structures, and scientific output has, however, not found any compelling evidence for this claim (Auranen and Nieminen 2010).

3.3 Research Planning

After receiving funding for one's research, implementing the research idea requires a detailed plan (while receiving funding also often requires a specific research plan, this section is related to all aspects of research that need planning, above and beyond just arriving at a plan that allows applying for funding). Normative assumptions also influence this process of planning and conceptualising empirical studies, sometimes leading to systematic influences in scientific results. Planning is, e.g., influenced by paradigmatic norms, habits, types of research, and the choice of research participants.

In particular, research planning is often a series of (conscious) decisions by researchers; these decisions, however, never occur in a normative vacuum. E.g., previous research is highly influential in research planning. If previous work, however, was limited to certain aspects, it may lead a researcher to neglect some relevant lines of enquiry for several reasons: First, analysing data with new (un-paradigmatic) questions is inherently precarious, which is why scientists may over-focus on a limited aspect of their research paradigm (Kuhn 1962). Second, due to the difficulty of publishing non-significant quantitative research findings (*file drawer problem*; Rosenthal 1979), scientists may be unaware of certain relations and may thus not investigate them. Third, a strong focus on linear cause-and-effect relations in social science research may lead to the neglect of many of the complexities within a studied system (McGuire 1973), further restricting the type of research project that is deemed to be of value. Lastly, many concepts can be defined very differently. These definitions are, however, at the core of the research process and not only influence a study's outcomes but also create a normative frame of key concepts. E.g., "the difference between a terrorist and a freedom fighter is a matter of perspective: It all depends on the observer and the verdict of history" (Linkola 2009, p. 160). If one were to study terrorists' attitudes, it is hard to imagine that stating, "You as a terrorist …" and "You as a freedom fighter …" led to the same responses.

Moreover, the subjects included in a study may also have an impact on outcomes. Research in the social sciences is often not as representative and generalisable as might be expected. E.g., between 2003 and 2007, 96 percent of participants in six psychology journals came from industrialised Western countries, which, however, only comprises 12 percent of the global population (Henrich, Heine, and Norenzayan 2010). Hence, study samples in the social sciences are often WEIRD (Western, Educated, Industrialized, Rich, and Democratic). Furthermore, the majority of these participants are undergraduate psychology students (Arnett 2008, Gallander Wintre, North, and Sugar 2001, Peterson 2001). While it is convenient to sample students from one's own university, it is also questionable whether claims regarding individuals can be abstracted from such a specific group. In some cases, e.g., when the underlying theory assumes a mechanism to be fundamental to all human beings, generalisation may be possible (Stroebe, Gadenne, and Nijstad 2018); however, in others not (Anderson and Stamoulis 2007, Gendron et al. 2014). As most psychological theories were constructed utilising such samples, the question of cross-cultural validity (or even within culture generalisability) often remains unanswered. Moreover, the theories and theorists themselves are often also WEIRD when theories and theorising rely on a very limited set of assumptions (for an exception, see Guo et al. 2013, van Zomeren 2016). In summary, research planning – and thus also the results – of empirical research are seldom as objective and "straightforward" as they seem to be. Once a researcher has planned a study, one has to collect data, a process which once more can be subject to normative influences.

3.4 Data Collection

Many claim that empirical data collection in the social sciences has moved the field away from subjective influences towards a more objective representation of reality or at least of the data. Randomised controlled trials (RCTs) have long been the prime example of this apparent cut with individual, political, or cultural influences in empirical sciences. They have often been called the gold standard of research in social sciences that allow for experimental manipulation (for a review, see Cartwright 2007, Meldrum 2000). The reason why RCTs are considered as such, was the introduction of several key components to safeguard against a range of biases: Randomisation, (placebo-)control, and masking are three key elements to the RCT method, each of which is a solution to a form of systematic bias researchers faced in the past (Kaptchuk 2001).

The element of a control group was introduced as evidence for the effectiveness of an intervention (Dehue 2005). A control group is the benchmark to which the treatment group is compared. If one can show that the intervention was the only thing that differed between two groups, this might be considered strong evidence for the intervention. To ensure that the two groups are as similar as possible and only differ by the factor of intervention the second element, randomisation, was introduced. By randomly assigning participants to either the treatment or the control group, one can assume that, given a reasonable sample size and random sampling, chance will allocate individual differences roughly equally, minimising the mean differences between the groups (for a historical review, see Dehue 1997, Hacking 1988). In short, this method would unfetter the design from conscious or unconscious allocation-decisions of the experimenter (Kaptchuk 2001). Lastly, masking describes the idea that neither the participants nor the researchers are aware of the participant's experimental group affiliation. This element controls for deliberate influences by the experimenter as well as for several well-known unconscious psychological effects. An example would be the *experimenter's bias*, the idea that the researcher's expectations unconsciously change his or her behaviour towards the study object. While RCTs produce useful results and develop the sciences in which they are applied, the question remains whether they can stick up to the reputation of eliminating normative influences.

RCTs, although (seemingly) promising as an extensive effort to move towards a more objective methodology, also introduce their own problems. Many have pointed out that even the simplest RCT experiment is never free of (normatively influenced) choices made by the scientists in the research process. Many of these decisions relate to the previous stages of research where someone had to define, operationalise, and measure a concept. Zwaan (2013), in his blog post "50 Questions About Messy Rooms", argues that in many cases already setting up control and experimental conditions is not as clear and objective as it is often claimed, and highlights that many aspects of study setups are subjective and influenced by personal (normative) biases and expectations.

Another type of scientific standard, at least in many social sciences, seems to be a preference to favour quantitative over qualitative data – something which also holds for the way that RCT experiments are typically set up. Quantitative research methods that rely on numerical results and statistics seem to be valued more than qualitative research methods that rely on more personal investigations of words, pictures, and objects. Such a development is especially apparent when considering the methodological courses offered at research-oriented universities and when considering what kind of research is published in the highest ranking and most cited journals. Both university education and top-tier journals focus largely on quantitative research methods and lack the inclusion of qualitative research (see, e. g., Shuval et al. 2011). Again, the argument many educators and philosophers of science offer is that quantitative results are clear, solid, and most importantly less affected by normative influences (yet, see also Tetlock 1994). Alan Bryman (1984) famously observed that the debate between quantitative and qualitative researchers is less methodological and more metaphysical, asking the important question of epistemological positions. Bryman writes that qualitative methods are often seen as phenomenological and constructivist by proponents of quantitative research, while quantitative research methods are described as positivistic from the perspective of researchers utilising qualitative research methods. Consequently, a social science that tries to lose its relativistic connotation, looking for hard facts, came to systematically favour a quantitative approach in order to show its positivistic effort to uncover the reality as it is without any normative connotation. In sum, the philosophers of science and research practitioners in the social sciences have gone to great lengths to build an argument for a purely empirical and somewhat "objective" and "quantitative" research process. Therefore, this process is in itself an important systemic influence on data collection that implicitly and explicitly restricts researchers to a limited valued frame of operation.

The problems with qualitative research that these researchers point out hold, however, also for quantitative research: In the case of developmental research, according to Peterson (2016), researchers cannot rely on strictly following standard procedures, as babies and children can easily violate standard protocols (e. g., throwing temper tantrums, being too tired, or being too excited). Excluding every infant that breaks protocol may result in the need for very large initial samples, only to reach a very small (and highly specific, thus even more biased) final sample – something that often is logistically impractical or impossible. Peterson claims that in order to gain meaningful insights (and statistical significance), certain factors are required: Flexibility with protocols, early analyses of the data, consideration of failure as a way to detect boundary conditions, and the analysis of unexpected statistically significant results. For many researchers, these factors may raise (or even constitute) red flags for "bad" research. However, most people would also agree that such developmental research has yielded some beneficial insights into the development of humans. Therefore, the process of collecting data may be (severely) influenced by a field's ideals and standards, and while questioning each one of these should be done in moderation, reflecting on them may nonetheless be beneficial.

3.5 Data Analysis

The next research stage of interest, logically following after data collection, is the analysis of the gathered data. One may assume that at least this process should be free from normative influences, as statistical methods are mathematical equations and therefore "objective" in their nature. However, also in the process of data analysis, normative influences and personal judgements play a large role. For instance, Kahneman (2011), in his popular-scientific book *Thinking, Fast and Slow*, exemplifies that scholars are taken in by human biases as well – also influencing how they use statistics to answer specific questions. In his research on heuristics and biases, he argues that lay people, as well as experts in statistics (e. g., statistics professors), maintained biases or heuristics that lead to incorrect statistical inferences. Kahneman implicitly makes an argument that scholars, in general, can (and should) improve their "how-to" knowledge on data analysis. It appears that certain normative influences on data analysis are influenced by either a lack of statistical knowledge or a rather unintentional reliance on internalised heuristics and biases (and thus also one's "go-to" methods of analysis). We will address two aspects in more detail: Normative influences in (1) finding statistical results, and in (2) describing statistical results.

Firstly, the process of data analysis can be influenced by statistical interpretation of the findings. A normative challenge arises when scholars interpret the statistical output of the data, for instance, to distinguish between the magnitude of the statistical certainty (probability value or p-value) and the practical impact of the finding (Cohen 1992, Kirk 1996). Plainly, it appears that during data analysis a certain grey-zone of statistical interpretation arises wherein scholars have to navigate. E. g., a non-significant statistic can be observed (i. e., a high p-value), even when the effect size (i. e., the comparative practical size of the effect; see Cohen 1992) can be substantial. Furthermore, small non-significant effect sizes (e. g., due to small samples) can be meaningful – and in practice even save lives. Or, the other way around, an effect can be statistically significant but irrelevant in reality. The leeway of scholars occurs within the bounds of practical or statistical relevance and tends to favour statistical over practical significance (e. g., Roberts 2015). Furthermore, one can remove unfortunate outliers, transform the data, or interpret statistical assumptions loosely to "improve" the likelihood of finding statistically significant results – all without having violated formal guidelines. The aforementioned points relate to the issue raised by Roberts (2015), who discusses the importance of a statistically significant result, and why scholars tend to pursue a statistically significant result. The success of researchers seems to depend on a *p*-value below 0.05.[4] With non-significant research findings ($p > 0.05$) scientific journals are less likely to give coverage to one's results, even if they are highly relevant in practice. All in all, there is a seemingly normative pressure on researchers to pursue findings in their studies with "good" (i.e., significant) statistical results (see Cohen 1992, Kirk 1996).

Secondly, another strand of normative influences on data analysis considers the notion of "hypothesising after the results are known" (i.e., HARKing; see Kerr 1998, p. 196). One speaks of HARKing when a researcher hypothesises based on data (i.e., post hoc), but presents the findings as if they were predicted before the data was collected (Edwards and Roy 2017). This is problematic, because Type I errors (finding an effect that does not exist) may be translated into theory and the process of scientific research may be misrepresented (Rubin 2017). The scope of HARKing has been investigated in several qualitative and quantitative studies. For instance, qualitative investigations of social scientists report severe transgressions of scientific dishonesty, such as questionable lab protocols and scheduled HARKing meetings (Peterson 2016). Quantitative studies, looking at a broader group of scientists, indicate similar results (John, Loewenstein, and Prelec 2012). Explicitly, 74 percent of researchers stated to not always report all the measures they used, and 71 percent continued data collection until statistical significance was reached (John, Loewenstein, and Prelec 2012). Thus, what is considered as a good statistical result can influence the a priori described aim of the empirical study. This shifts the supposed right way of deductive research towards a more inductive way of hypothesising because the interpretation of statistics by researchers can be value-laden.

HARKing has often been described as a practice of malicious intent – researchers using their data to cheat. Issues of at least equal importance are the data analysis decisions that are made unintentionally or theory-guided but are still data contingent and neglect the implications of potential analyses testing the same question. A recent article by Gelman and Loken (2013) argues that the same issue that underlies HARKing – namely doing analyses until one finds a statistically significant result – can still be problematic even if one

4 A *p*-value below 0.05 suggests that if we drew infinitely random samples from the population, 95 percent of the samples would reject the null hypothesis (frequentist understanding), or as a theoretical statement, the finding has a probability of 0.95 of correctly rejecting the null hypothesis (Kass 2011).

only does a single analysis. Especially when the analysis depends heavily on the data structure, even a single analysis is problematic because there are many possible analyses that could have been done with the same data and still answer the same research question.[5] This includes two main aspects. (1) One influence is the arbitrary decision of how a hypothesis is tested. Gelman and Loken offer a series of examples to illustrate this point. One compelling example especially highlights this "garden of forking paths" (Gelman and Loken 2013, p. 1) – the assertion that a specific analysis testing a research hypothesis could be contrary to a series of other analyses testing the same hypothesis. Especially broad or vague hypotheses, such as "political orientation has an influence on voting turnout". There are multiple ways in which we could test such a hypothesis – all of which would be based on theoretical grounds. If we were to find a political orientation effect among men (because, e. g., men are more ideological) we would see the hypothesis confirmed. If we were to find an effect only in state elections but not in county elections (because ideological issues often arise on a state level), we would also see our hypothesis confirmed. Choosing any specific analysis is in parts an often-unacknowledged arbitrary choice. (2) A second form of unintentional influence on the analysis performed is the arbitrary decision made during the analysis process (Steegen et al. 2016). E. g., a researcher might decide to define a person as poor if they earn less than 60 percent of the median income. This can be a reasonable definition and analysis choice, yet, defining someone as poor if they cannot afford basic necessities might be just as valid of a definition of poverty. Deciding between the two options often does not solely depend on theory (both decision rules have theoretical and empirical backing) but it is an arbitrary decision by the researcher that might have led to different results if they had chosen a different method.

Consequently, normative influences (e. g., subjective expectations and biases) remain prevalent in finding and describing statistical results in the process of data analysis (see, e. g., Gelman and Loken 2013). The topic of statistical analysis relates closely to the subsequent section of scientific output in general.

5 A recent study exemplifies this point. Scholars received the same dataset and the same question: "[are] soccer players with dark skin tone [...] more likely than those with light skin tone to receive red cards from referees" (Silberzahn et al. 2018, p. 338). They found that analytic choices within each scientific field resulted in different outcomes.

3.6 Output

Every process of scientific inquiry aims to be shared in one way or another with the scientific community or the society at large. Scientific communication ranges from journal publications to conference proceedings, or output aimed at the general public. Scientific communication, or scientific output as we shall label it, is a necessary part of every scientific process. The creation of scientific output overarches many of the processes we have discussed thus far.

Such output may be influenced by the norms or expectations within the researchers' academic communities and institutions (university or otherwise). The communal judgement of academic output (e. g., through peer-review), often, has a general tendency to support a certain status quo. Take, e. g., the research by Nobel-prize winner George Akerlof, who studied a special problem of asymmetric information for markets. This later led to a paradigm shift in economics. Before this occurred, though, the publication of his research was turned down by several referees of well-known journals: *The Academic Economic Review* and *The Review of English Studies* both rejected it for "triviality", a reviewer of the *Journal of Political Economy* stated: "If this paper would be correct, no goods could be traded at all" (Akerlof 2003, par. 14). There are, thus, systemic restrictions in the publishing process that (normatively) value certain paradigm-corroborating papers over others (at times at the expense of important research findings).

Apart from finding acceptance for one's findings by journals, researchers may also experience normative pressures from their universities and funding agencies. Academic institutions may only grant rewards for researchers' efforts when they succeed in publishing a certain number of scientific articles in certain high-tier journals (Edwards and Roy 2017, Roberts 2015). These kinds of productivity norms can often be manipulated. E. g., researchers' output can be quantified using a measure that calculates the personal productiveness and impact of publications: the *h*-index (Hirsch 2005).[6] While the *h*-index takes into account the quantity of published studies, it can be easily manipulated by self-citation (Gálvez 2017). Edwards and Roy (2017) argue that a focus on quantity over quality, when linked to quantitative performance measures, such as the *h*-index, causes these measures to be a target rather than a measure. For researchers, many factors – such as hiring, promotion, funding, and awards – depend on these quantitative measures, which is why they are seemingly pressured to em-

6 The *h*-index quantifies the number of an author's publications and their citations (Hirsch 2005). With one published paper being cited once, the *h*-index will be 1, with two published papers, both of them being cited twice, the *h*-index is 2, and so on.

phasise quantity of publications over their quality. Hence, there is pressure on researchers who do empirical studies, through certain expectations of the institutions they rely on (see also subsection 3.2 on *Goodhart's Law*).

Moreover, the aforementioned form of pressure strongly relates to the possible personal (i. e., subjective) goals of a researcher in comparison to the collective goal of building a foundation for future research. For instance, a researcher can become widely known after publishing counterintuitive or "flashy" findings. Although pursuing counterintuitive findings may not be obscured by normative influences, it could diminish the objectivity of a researcher when the two collide. Roberts describes this as "clickbait worthy research" (Roberts 2015, p. 2). His argument centres on the tendency to neglect certain long-lasting problems, but rather pursuing something "flashy" that might result in popularity for the scholar publishing the findings. However, "researchers [should] seek to answer some of the most fundamental questions that humans can ask about nature" (National Academies 2009, p. 1). Pursuing clickbait-worthy findings conflicts with a core view that the National Academies and others postulate and can make way for value-laden judgements which relate back to the stage of idea generation. In the end, pursuing personal (subjective) goals, influenced by the institutional context, may influence what kind of output is generated, instead of asking oneself what a study may contribute to the literature.

Output does not only relate to the scientific system as a whole, it also has very real-life implications on the presented output itself (e. g., papers, presentations, and so forth); the underlying pressure for it often being the drive to publish one's findings. When publishing, however, many hurdles have to be taken. One of these hurdles may be the need to present a clear and understandable narrative – something that, in theory, should not pose a problem, as one should just write down how the research was conducted. Research should be presented in a clear fashion with a clear introduction, followed by a fitting method, and ended with significant findings that correspond to the hypotheses in the introduction. This is because a clear and understandable narrative often helps a different person (e. g., reviewers, editors, or readers) to understand one's findings. However, this can become dangerous if the suggestion of a clear narrative becomes a need or even a "tyranny" (Roberts 2015, p. 2). The argument that underlies the clean narrative is that it might be beneficial to rotate the order and writing or changing a previous narrative to be more consistent with one's findings (i. e., HARKing; see Kerr 1998). People may thus be prone to not honestly report on all the hurdles and inspirations on the way, but rather try to present the most coherent story – something that often does not reflect the actual research process. Thus, there are subjective expectations in sciences concerning the narrative of empirical studies which can, partially, deter the objectivity of findings, for instance as we discussed with the example of HARKing (see subsection 3.5).

4 Moving Forward – The Next Steps

When reflecting on how these research stages deal with normative influences, one is quickly faced with several questions: Is there no objectivity in empirical research? And should we, if that were the case, stop doing empirical research altogether? The latter question would be answered with a determined "no" by most scientists (ourselves included). We hope that theorists and practitioners can use the six steps of the research process to reflect upon empirical results and thereby understand and use the inferences from research, while being aware of the normative influences. We would also like the reader to further consider the following points.

Firstly, we would like to propose an active and public discussion about the normative influences of the academic system and its incentive structures. Secondly, we suggest utilising multiple research (method) perspectives in the research process. Thirdly, we want to propose some concrete ways in which individual researchers can develop a nuanced understanding of their scientific output's normative influences. To communicate these points, we will use the remainder of this chapter to suggest concrete measures both on a structural and individual level.

4.1 Change the System – Addressing the Incentive Structure

We recognise the enormous effort necessary to change an established system. However, recently several strings of thought have emerged to propose systemic alternatives. For instance, some have called for a renaissance of a "democratic educator" tradition (Rustin 2016, p. 147), others focussed on the values (e. g., cultural and intellectual purpose) universities should strive for (Collini 2012, Miedema 2012, Thomas 2010), and again others argue for more transparent journal procedures (e. g., pre-registration, openness of peer-reviews; see Gonzales and Cunningham 2015, Polka et al. 2018). In the following paragraphs, we would like to highlight some of these developments and discuss how the larger scientific incentive system and the research process could profit from them.

Many thinkers mention a reconstructing of governmental funding for the academic incentive system. For instance, one funding system that is not entirely novel but might help offer a more independent, low-pressure research environment, could be long-term institutional grants (like the *Harnack Principle* of the Max Planck Society in Germany). Some have argued that larger, independent, and unconditional grants for relevant topics might give researchers the leeway to spend some of their resources on free and bold theorising (which links back to the idea generation stage). It might introduce a different notion of scientific freedom than the current situation and also offer a space for the large-scale theorising the social sciences dearly need to develop, even though a fair and unbiased distribution of funds may be a challenge in itself. Recognising this need, some research councils have already implemented changes that lead to longer grant durations (e. g., Burgio 2017) or larger researcher freedom due to a broadened scope of evaluation panels (e. g., Hornyak 2017), emphasising original research and merit thereof over pure quantity of publication. Inspired by such approaches, we propose that rethinking governmental funding might have the potential to positively change normative assumptions in the process of empirical research and the six stages therein.

With a change in (governmental) funding, we also have to ask the question who should profit from these investments. We believe that research funded by society, through governmental channels, should be reported back to and impact society. An interesting development in this area is the publication process in open science formats (e. g., Butler 2016, Waldrop 2008),[7] allowing interested individuals to read scientific output and to investigate data themselves (for more information, see the initiative "Science Without Publication Paywalls" by cOAlition S; see Schiltz 2018). This allows for transparency of the data, enables new collaborations, and makes efficient use of the data (see also Else 2018). A complementary approach is to use the open access channels to include other disciplines and their expertise. E. g., in a recent paper, a collaboration of 84 people from multiple disciplines worked on an openly accessible paper, providing expertise from multiple angles (Lakens et al. 2018). With today's technology, these endeavours do not prove to be impossible anymore, as many software and internet services allow for a contemporaneous collaboration, no matter where on the world researchers are based. From this, we argue that changing the standards of how scientific results are accessed and communicated can benefit society as a whole and might in the process create an awareness of the limitations of any individual scientific discipline.

An additional issue closely linked to this is the journal-based review process. We will not take an absolute stance in the discussion on what kind of review

7 There are different formats of open science, some scholars advocate for open data formats; for a review, see Arzberger et al. (2004), Janssen, Charalabidis, and Zuiderwijk (2012).

process may be the best (see, e.g., the discussion between Fiske 2016 and Gelman 2016), but the process of reviewing certain scientific output can possibly be made less value-laden.[8] For instance, citation pushing (e.g., reviewers suggesting their own work when commenting on a paper, even though it is not relevant in that context) may be deterred if one knows who is reviewing. One could also imagine that if the reviewing and editing process would be "rewarded", a reviewer could gain credit through the possibility of adding her name to a section of the paper. This would hopefully encourage the reviewer to provide useful feedback as she gets partly responsible for the research project. Another possibility would be to "pay" reviewers from the journal income. Any payment that reviewers receive should, however, be strictly limited to funding for their own future research projects. One could argue that paying a reviewer would bias the reviewer to support a possible manuscript. One possible solution to this would be to make the outcome of the review independent of the payment process. Reviewers who do not live up to the expectations may, however, not be invited for reviews in the future. Nonetheless, crediting the reviewers may be an interesting development that could help change the academic reward system, but then again, ironically, it may also fuel other biases. Although we realise that solely generating research without any value-laden judgement is untenable, a step forward could be openness in the process.

Others have argued, more philosophically, that scholarship and learning are intrinsic values in themselves and what universities should again strive for (Collini 2012), or that universities have a cultural and intellectual purpose of free speculations and inter-generational transmission (Thomas 2010). Still others have started building a (mostly) new vision of academia – in symbiosis with the current system and the broader society (Miedema 2012). These new initiatives do their best to propose more democratic and inclusive environments, which, in our minds, should also aid a more open and reflective research climate.

All the previous points result in the question of what academic output is and what it ought to be. With a change of the incentive system, (more) people may realise that doing academic work and getting academic achievements do not have to be zero-sum, and that, e.g., very often publication- and citation-numbers are not ideal in determining the value of a project to society. One should be aware that the excellence of different people might become evident in different measurements. Limiting academic output solely to the production and publishing of papers fails to address science in its complexity. Educator, theoriser, practitioner, reviewer, communicator, and connector are academic roles that are not

8 For an interesting podcast episode on incivility in reviewing, see Inbar and Inzlicht (2018).

sufficiently rewarded by the current academic system. It could, thus, be beneficial (and fair) to change the incentive system so that not a single measurement or a single dimension of measurement is applied, but that a multitude of measures provides a better picture.

4.2 Robust Research

There are also very practical possibilities for researchers to disentangle normative influences from empirical research: e. g., by changing how scholars conduct research (e. g., Munafò and Smith 2018), as well as how they handle and present data (e. g., Weissgerber et al. 2015). We will thus elaborate on how individuals can reflect on influences on the research process.

Collaborating with multiple disciplines is inherent to the recently encouraged triangulation approach (Munafò and Smith 2018). In this approach, any research question is investigated by a team consisting of multiple disciplines, applying multiple methods, and using theories from different angles. The use of mixed methods may help to reveal much more than a single method could (e. g., Pool et al. 2010), and may yield insights that would be ignored when employing only one method.[9] This basically means that the process of replication is conducted from different fields (for an overview on replication, see Open Science Collaboration 2015). Particularly, triangulation calls for a stronger, possibly bolder, and more integrative theoretical and statistical approach to empirical research. An interesting addition to this would be a combination with a change to the incentive system. Munafò and Smith (2018) suggest crediting every person involved in the triangulation process, specifically defining which person contributed in which way. This would acknowledge every person involved in the project, giving credit to people that may sometimes be left behind in the current system.

In terms of data handling, researchers should try to avoid normative default options and use the statistical methods and standards most suitable. E. g., the notion of setting the Type I error rate (i. e., the chance of inferring an effect that is not there; also known as the *alpha level*) to 5 percent has become almost a doctrine in the scientific domains. In light of the recent replication crisis, some have suggested to set the "new" alpha to 0.5 percent, requiring "stronger" findings in order to achieve statistical significance (Benjamin et al. 2018). On the opposite side, other voices have called for abandoning alpha levels all together and

9 An interesting approach can be to combine the theoretical design with the statistical design, e. g., in *dynamic systems modelling*; see Kunnen (2012).

looking at the size of the actual effects instead (following the seminal paper by Cohen 1992, Sullivan and Feinn 2012). Another interesting approach argues to "give p's a chance" and to let researchers reflect on their results (Albers 2017, par. 1, see also Lakens et al. 2018). This would give a researcher the freedom to set the alpha level in accordance with one's study. If a Type I error has to be avoided at all costs (e.g., if a medical treatment has strong negative side effects) the researcher can set their Type I error more rigidly (e.g., alpha to 0.1 percent). However, if this is not the case, a more lenient Type I error rate can be utilised. Sincerely reporting on these choices, in combination with an indication of statistical power, would provide a broader picture of the investigated reality. The move away from strong normative default options to more individual decision making can also be found when applying general research methods. E.g., Dehue (2002) argues that we should refrain from using a randomised control trial in every situation, because qualitative and mixed methods research can be superior in some situations. All these options have in common that they encourage the researcher to use the methods most suitable for the research question at hand (while justifying their use) and refrain from normative defaults assumed to be "good".

Moreover, a scholar can try to present the data more thoroughly and comprehensively with less normative decisions (e.g., Weissgerber et al. 2015). E.g., a recent study shows that bar charts and simple slope graphs (a graph of predicted effects at very high or low values of the predictor) remain prevalent (47 percent of total observed) as visualisation methods in psychology (Stulp 2017). They provide descriptive visual information but might overstate the results (e.g., because it does not show the large variation in the data) or misrepresent the data and relationships to the human eye. The choice for visualising the data with these kinds of plots can be problematic (e.g., Anscombe 1973). The graphs used are often too simplistic to portray the vast complexity of the underlying data (see, e.g., the *Datasaurus Dozen* visualised by Locke and D'Agostino McGowan 2017). Weissgerber and colleagues (2015) argue that presenting the data in a more complete manner can show, for instance, the distribution of data rather than just presenting a simple slope (as Anscombe visualised by presenting four same slopes with different data points in each graph). Scholars can still convey their message but presenting the data more comprehensibly can leave some of the subjective interpretation of the data to the reader, and thus divert some of the interpretation to the reader without prematurely excluding alternative explanations. It is, however, not a new suggestion, as illustrated by a quote by Edward Tufte (2001, p. 105), who demanded early on: "*above all else show the data*".

4.3 Reflexivity

We finally want to advocate for a method that could sit at the core of our previous points. At the beginning of our chapter, we introduced ourselves and explained our own view on normative influences and the normative influences in, and on, our education. By doing so, we tried to give our readers an understanding of our own background, making our arguments not an objective representation of truth but a subjective interpretation of the current field. We believe that such a reflexivity approach can also benefit research in such a way that it would give researchers and their audience agency to understand why specific decisions in the research process were made and their potential consequences.

Reflexivity has been a key term in qualitative social science research in the last decades, especially in the field of sociology, but the concept has been applied in many different ways (Holland 1999). The concept is, however, old. Mead (1934, p. 134) defined reflexivity as "the turning back of the experience of the individual upon [her- or himself]". In recent years, reflexivity was argued to be one of the most appealing buzzwords of sociology and epistemology (Tsekeris 2010). Reflexivity has generally been understood in the social sciences and in recent years as "awareness of the influence the researcher has on the people or topic being studied, while simultaneously recognising how the research experience is affecting the researcher" (Probst 2015, p. 37). It has been argued to be essential in qualitative research (Ahmed, Hundt, and Blackburn 2011, Blaxter, Hughes, and Tight 2001, D'Cruz, Gillingham, and Melendez 2007, Gilgun 2008, Koch and Harrington 1998, Lazard and McAvoy 2017), as a means to, and end to, overcome the criticism that qualitative research was anecdotal and subject to researcher's bias (Patniak 2013). Reflexivity in qualitative research aims to monitor the effects of a researcher's involvement in the research, thus improving both the research's accuracy as well as the findings' credibility and context by clarifying the researcher's biases, values, and beliefs (Cutcliffe 2003). There are many ways in which this may be done and there is no standardised procedure for doing so. Attia and Edge (2017) recently argued that reflexivity should be considered as consisting of both a prospective and a retrospective component. The former is concerned with the impact of the researcher on the study, while the latter is concerned with the impact of the study on the researcher.

As we have illustrated throughout this chapter, quantitative research can also, against its best efforts and intentions, be subjected to all kinds of biases, and the introduction of reflexivity throughout empirical research may, therefore, be highly beneficial. E. g., Dehue (2002, p. 86) argues that "the designing of surveys and tests demands the taking of decisions as to which categories to use, and how to further specify them in survey questions and test items. After a research project, the original decisions are removed from the construction like redundant moulds”.

Moreover, this would allow others to somewhat control for researchers’ values and beliefs when pooling together results from many different studies, such as utilising meta-analyses. However, this has barely been done, with a few exceptions (Ryan and Golden 2006, Walker, Read, and Priest 2013). Nonetheless, journals seem to require such processes more and more (see, e. g., Nature 2018). Therefore, embracing, or at least actively acknowledging, decisions potentially influenced by the normative framework instead of removing them, could provide a broader and more holistic view on the research process. Such an embracement could provide important information and help not only the practitioners of science but also the users of science (e. g., philosophers). With our reflexivity example at the beginning of this chapter, we elucidate one way to implement reflexivity, nevertheless, this is by all means not the only way to do so.

5 Conclusion

In this chapter we aimed to raise an argument that in each step of the research process, in (1) idea generation, (2) research funding, (3) research planning, (4) data collection, (5) data analysis, and (6) scientific output, normative influences play a role. Furthermore, we aimed to answer the “is” question postulated at the beginning of this piece: What normative influences exist within empirical research? These stages are, moreover, not independent of each other. Systemic influences often intertwine these domains. Finally, we propose that addressing the incentive system, robust research, and reflecting on normative influences could provide a more holistic picture of the overall interwoven research process.

Acknowledgements

We express our sincerest appreciation to Martijn van Zomeren, Trudy Dehue, and Alexander Max Bauer, for intellectual support and advice from the start of this process.

Bibliography

Ahmed, Dunya Ahmed Abdulla; Hundt, Gillian Lewando, and Blackburn, Clare (2011): "Issues of Gender, Reflexivity and Positionality in the Field of Disability. Researching Visual Impairment in an Arab Society". In: *Qualitative Social Work* 10 (4), pp. 467 – 484.
Akerlof, George (2003): "Writing the 'The Market for "Lemons"'. A Personal and Interpretive Essay". In: *The Nobel Prize*. https://www.nobelprize.org/prizes/economics/2001/akerlof/article/, retrieved on November 23, 2018.
Albers, Casper (2017): "All We Are Saying is Give p's a Chance". In: *Mindwise Psychometrics*. http://mindwise-groningen.nl/all-we-are-saying-is-give-ps-a-chance/, retrieved on November 24, 2018.
Alberts, Bruce; Kirschner, Marc; Tilghman, Shirley, and Varmus, Harold (2014): "Rescuing US Biomedical Research from Its Systemic Flaws". In: *Proceedings of the National Academy of Sciences* 111 (16), pp. 5773 – 5777.
Anderson, Leigh and Stamoulis, Kostas (2007): "Applying Behavioural Economics to International Development Policy". In: Mavrotas, George and Shorrocks, Anthony (eds.): *Advancing Development*. London: Palgrave Macmillan, pp. 664 – 685.
Anscombe, Francis (1973): "Graphs in Statistical Analysis". In: *The American Statistician* 27 (1), p. 17.
Arnett, Jeffrey (2008): "The Neglected 95 %. Why American Psychology Needs to Become Less American". In: *American Psychologist* 63 (7), pp. 602 – 614.
Arzberger, Peter; Schroeder, Peter; Beaulieu, Anne; Bowker, Geof; Casey, Kathleen; Laaksonen, Leif; Moorman, David; Uhlir, Paul, and Wouters, Paul (2004): "An International Framework to Promote Access to Data". In: *Science* 303 (5665), pp. 1777 – 1778.
Attia, Mariam and Edge, Julian (2017): "Be(com)ing a Reflexive Researcher. A Developmental Approach to Research Methodology". In: *Open Review of Educational Research* 4 (1), pp. 33 – 45.
Auranen, Otto and Nieminen, Mika (2010): "University Research Funding and Publication Performance. An International Comparison". In: *Research Policy* 39 (6), pp. 822 – 834.
Benjamin, Daniel; Berger, James; Johannesson, Magnus; Nosek, Brian; Wagenmakers, Eric-Jan; Berk, Richard; Bollen, Kenneth; Brembs, Björn; Brown, Lawrence; Camerer, Colin; Cesarini, David; Chambers, Christopher; Clyde, Merlise; Cook, Thomas; de Boeck, Paul; Dienes, Zoltan; Dreber, Anna; Easwaran, Kenny; Efferson, Charles; Fehr, Ernst; Fidler, Fiona; Field, Andy; Forster, Malcolm; George, Edward; Gonzalez, Richard; Goodman, Steven; Green, Edwin; Green, Donald; Greenwald, Anthony; Hadfield, Jarrod; Hedges, Larry; Held, Leonhard; Ho, Teck-Hua; Hoijtink, Herbert; Hruschka, Daniel; Imai, Kosuke; Imbens, Guido; Ioannidis, John; Jeon, Minjeong; Jones, James Holland; Kirchler, Michael; Laibson, David; List, John; Little, Roderick; Lupia, Arthur; Machery, Edouard; Maxwell, Scott; McCarthy, Michael; Moore, Don; Morgan, Stephen; Munafò, Marcus; Nakagawa, Shinichi; Nyhan, Brendan; Parker, Timothy; Pericchi, Luis; Perugini, Marco; Rouder, Jeff; Rousseau, Judith; Savalei, Victoria; Schönbrodt, Felix; Sellke, Thomas; Sinclair, Betsy; Tingley, Dustin; van Zandt, Trisha; Vazire, Simine; Watts, Duncan; Winship, Christopher; Wolpert, Robert; Xie, Yu; Young, Cristobal; Zinman, Jonathan, and

Johnson, Valen (2018): “Redefine Statistical Significance”. In: *Nature Human Behaviour* 2 (1), pp. 6–10.
Bhattacharjee, Yudhijit (2013): “The Mind of a Con Man”. In: *The New York Times*. http://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html, retrieved on November 25, 2018.
Blaxter, Loraine; Hughes, Christina, and Tight, Malcolm (2001): *How to Research*. Buckingham: Open University Press.
Bryman, Alan (1984): “The Debate About Quantitative and Qualitative Research. A Question of Method or Epistemology?”. In: *The British Journal of Sociology* 35 (1), pp. 75–92.
Burgio, Gaetan (2017): “The NHMRC Program Grant Overhaul. Will It Change the Medical Research Landscape in Australia?”. In: *The Conversation*. https://theconversation.com/the-nhmrc-program-grant-overhaul-will-it-change-the-medical-research-landscape-in-australia-78343/, retrieved on November 14, 2018.
Butler, Declan (2016): “Dutch Lead European Push to Flip Journals to Open Access”. In: *Nature* 529 (7584), p. 13.
Carey, Benedict (2011): “Fraud Case Seen as a Red Flag for Psychology Research”. In: *The New York Times*. http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html, retrieved on November 14, 2018.
Cartwright, Nancy (2007): “Are RCTs the Gold Standard?”. In: *BioSocieties* 2 (1), pp. 11–20.
Cohen, Jacob (1992): “A Power Primer”. In: *Psychological Bulletin* 112 (1), pp. 155–159.
Collini, Stefan (2012): *What Are Universities For?* London: Penguin.
Crawford, Jarret; Duarte, José; Haidt, Jonathan; Jussim, Lee; Stern, Charlotta, and Tetlock, Philip (2015): “It May Be Harder than We Thought, But Political Diversity Will (Still) Improve Social Psychological Science”. In: *Behavioral and Brain Sciences* 38, e164.
Cushman, Priscilla; Hoeksema, Todd; Kouveliotou, Chryssa; Lowenthal, James, and Peterson, Bradley (2015): “Impact of Declining Proposal Success Rates on Scientific Productivity”. In: *ArXiv E-Prints*, arXiv:1510.01647.
Cutcliffe, John (2003): “Reconsidering Reflexivity. Introducing the Case for Intellectual Entrepreneurship”. In: *Qualitative Health Research* 13 (1), pp. 136–148.
D’Cruz, Heather; Gillingham, Philip, and Melendez, Sebastian (2007): “Reflexivity, Its Meanings and Relevance for Social Work. A Critical Review of the Literature”. In: *British Journal of Social Work* 37 (1), pp. 73–90.
Dehue, Trudy (2002): “A Dutch Treat. Randomized Controlled Experimentation and the Case of Heroin-Maintenance in the Netherlands”. In: *History of the Human Sciences* 15 (2), pp. 75–98.
Dehue, Trudy (1997): “Deception, Efficiency, and Random Groups. Psychology and the Gradual Origination of the Random Group Design”. In: *Isis* 88 (4), pp. 653–673.
Dehue, Trudy (2005): “History of the Control Group”. In: Everitt, Brian and Howell, David (eds.): *Encyclopedia of Statistics in Behavioral Science*. Vol. 2. Chichester: Wiley & Sons, pp. 829–836.
Duarte, José; Crawford, Jarret; Stern, Charlotta; Haidt, Jonathan; Jussim, Lee, and Tetlock, Philip (2015): “Political Diversity Will Improve Social Psychological Science”. In: *Behavioral and Brain Sciences* 38, e130.

Edwards, Marc and Roy, Siddhartha (2017): “Academic Research in the 21st Century. Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition”. In: *Environmental Engineering Science* 34 (1), pp. 51 – 61.
Else, Holly (2018): “Radical Open-Access Plan Could Spell End to Journal Subscriptions”. In: *Nature* 561, pp. 17 – 18.
Elton, Lewis (2004): “Goodhart’s Law and Performance Indicators in Higher Education”. In: *Evaluation & Research in Education* 18 (1 – 2), pp. 120 – 128.
Fang, Ferric; Bowen, Anthony, and Casadevall, Arturo (2016): “NIH Peer Review Percentile Scores Are Poorly Predictive of Grant Productivity”. In: *eLife* 5, e13323.
Fiske, Susan (2016): “A Call to Change Science’s Culture of Shaming”. In: *Association for Psychological Science*. https://www.psychologicalscience.org/observer/a-call-to-change-sciences-culture-of-shaming/, retrieved on November 14, 2018.
Gallander Wintre, Maxine; North, Christopher, and Sugar, Lorne (2001): “Psychologists’ Response to Criticisms About Research Based on Undergraduate Participants. A Developmental Perspective”. In: *Canadian Psychology, Psychologie Canadienne* 42 (3), pp. 216 – 225.
Gálvez, Ramiro (2017): “Assessing Author Self-Citation as a Mechanism of Relevant Knowledge Diffusion”. In: *Scientometrics* 111 (3), pp. 1801 – 1812.
Gelman, Andrew (2016): “What Has Happened Down Here Is the Winds Have Changed”. In: *Statistical Modeling, Causal Inference, and Social Science*. http://andrewgelman.com/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/, retrieved on November 14, 2018.
Gelman, Andrew and Loken, Eric (2013): “The Garden of Forking Paths. Why Multiple Comparisons Can Be a Problem, Even When There Is No ‘Fishing Expedition’ or ‘p-Hacking’ and the Research Hypothesis Was Posited Ahead of Time”. In: *Department of Statistics, Columbia University in the City of New York*. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf, retrieved on November 14, 2018.
Gendron, Maria; Roberson, Debi; van der Vyver, Jacoba Marietta, and Barrett, Lisa Feldman (2014): “Perceptions of Emotion from Facial Expressions Are Not Culturally Universal. Evidence from a Remote Culture”. In: *Emotion* 14 (2), pp. 251 – 262.
Gergen, Kenneth (1973): “Social Psychology as History”. In: *Journal of Personality and Social Psychology* 26 (2), pp. 309 – 320.
Gilgun, Jane (2008): “Lived Experience, Reflexivity, and Research on Perpetrators of Interpersonal Violence”. In: *Qualitative Social Work* 7 (2), pp. 181 – 197.
Gonzales, Joseph and Cunningham, Corbin (2015): “The Promise of Pre-Registration in Psychological Research”. In: *American Psychological Association*. http://www.apa.org/science/about/psa/2015/08/pre-registration.aspx, retrieved on November 14, 2018.
Guo, Lan; Stone, Dan; Bryant, Stephanie; Wier, Benson; Nikitkov, Alex; Ren, Chunyan; Riccio, Edson Luiz; Shen, Milton; Trabelsi, Samir, and Zhang, Li Fang (2013): “Are Consumers’ Financial Needs and Values Common Across Cultures? Evidence from Six Countries”. In: *International Journal of Consumer Studies* 37 (6), pp. 675 – 688.
Hacking, Ian (1988): “Philosophers of Experiment”. In: *Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2 (1), pp. 147 – 156.

Henrich, Joseph; Heine, Steven, and Norenzayan, Ara (2010): “The Weirdest People in the World?”. In: *Behavioral and Brain Sciences* 33 (2–3), pp. 61–83.
Hirsch, Jorge (2005): “An Index to Quantify an Individual’s Scientific Research Output”. In: *Proceedings of the National Academy of Sciences* 102 (46), pp. 16569–16572.
Holland, Ray (1999): “Reflexivity”. In: *Human Relations* 52 (4), pp. 463–484.
Hornyak, Tim (2017): “Japan Shakes up Research Funding System”. In: *Nature Index*. https://www.natureindex.com/news-blog/japan-shakes-up-research-funding-system/, retrieved on November 14, 2018.
Inbar, Yoel and Inzlicht, Michael (2018): “You’re Not Wrong Walter, You’re Just an A$$hole”. In: *Two Psychologists Four Beers*. http://michaelinzlicht.com/getting-better/2018/5/24/two-psychologists-four-beers-in-search-of-the-campus-free-speech-crisis-pe9hd/, retrieved on November 14, 2018.
Janssen, Marijn; Charalabidis, Yannis, and Zuiderwijk, Anneke (2012): “Benefits, Adoption Barriers and Myths of Open Data and Open Government”. In: *Information Systems Management* 29 (4), pp. 258–268.
John, Leslie; Loewenstein, George, and Prelec, Drazen (2012): “Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling”. In: *Psychological Science* 23 (5), pp. 524–532.
Kahneman, Daniel (2011): *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
Kant, Immanuel (1964): *Groundwork of the Metaphysic of Morals*. London: Harper Torchbooks.
Kaplan, David; Lacetera, Nicola, and Kaplan, Celia (2008): “Sample Size and Precision in NIH Peer Review”. In: *PLoS ONE* 3 (7), e2761.
Kaptchuk, Ted (2001): “The Double-Blind, Randomized, Placebo-Controlled Trial. Gold Standard or Golden Calf?”. In: *Journal of Clinical Epidemiology* 54 (6), pp. 541–549.
Kass, Robert (2011): “Statistical Inference. The Big Picture”. In: *Statistical Science* 26 (1), pp. 1–9.
Kerr, Norbert (1998): “HARKing. Hypothesizing After the Results Are Known”. In: *Personality and Social Psychology Review* 2 (3), pp. 196–217.
Kimble, Judith; Bement, William; Chang, Qiang; Cox, Benjamin; Drinkwater, Norman; Gourse, Richard; Hoskins, Aaron; Huttenlocher, Anna; Kreeger, Pamela; Lambert, Paul; Mailick, Marsha; Miyamoto, Shigeki; Moss, Richard; O’Connor-Giles, Kate; Roopra, Avtar; Saha, Krishanu, and Seidel, Hannah (2015): “Strategies from UW-Madison for Rescuing Biomedical Research in the US”. In: *eLife* 4, e09305.
Kirk, Roger (1996): “Practical Significance. A Concept Whose Time Has Come”. In: *Educational and Psychological Measurement* 56 (5), pp. 746–759.
Koch, Tina and Harrington, Ann (1998): “Reconceptualizing Rigour. The Case for Reflexivity”. In: *Journal of Advanced Nursing* 28 (4), pp. 882–890.
Kruglanski, Arie (2001): “That ‘Vision Thing’. The State of Theory in Social and Personality Psychology at the Edge of the New Millennium”. In: *Journal of Personality and Social Psychology* 80 (6), pp. 871–875.
Kuhn, Thomas (1962): *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
Kunnen, Saskia (2012): *A Dynamic Systems Approach to Adolescent Development*. London: Psychology Press.
Lakens, Daniel; Adolfi, Federico; Albers, Casper; Anvari, Farid; Apps, Matthew; Argamon, Shlomo; Baguley, Thom; Becker, Raymond; Benning, Stephen; Bradford, Daniel;

Buchanan, Erin; Caldwell, Aaron; van Calster, Ben; Carlsson, Rickard; Chen, Sau-Chin; Chung, Bryan; Colling, Lincoln; Collins, Gary; Crook, Zander; Cross, Emily; Daniels, Sameera; Danielsson, Henrik; DeBruine, Lisa; Dunleavy, Daniel; Earp, Brian; Feist, Michele; Ferrell, Jason; Field, James; Fox, Nicholas; Friesen, Amanda; Gomes, Caio; Gonzalez-Marquez, Monica; Grange, James; Grieve, Andrew; Guggenberger, Robert; Grist, James; van Harmelen, Anne-Laura; Hasselman, Fred; Hochard, Kevin; Hoffarth, Mark; Holmes, Nicholas; Ingre, Michael; Isager, Peder; Isotalus, Hanna; Johansson, Christer; Juszczyk, Konrad; Kenny, David; Khalil, Ahmed; Konat, Barbara; Lao, Junpeng; Larsen, Erik Gahner; Lodder, Gerine; Lukavský, Jiří; Madan, Christopher; Manheim, David; Martin, Stephen; Martin, Andrea; Mayo, Deborah; McCarthy, Randy; McConway, Kevin; McFarland, Colin; Nio, Amanda; Nilsonne, Gustav; de Oliveira, Cilene Lino; de Xivry, Jean-Jacques Orban; Parsons, Sam; Pfuhl, Gerit; Quinn, Kimberly; Sakon, John; Saribay, Adil; Schneider, Iris; Selvaraju, Manojkumar; Sjoerds, Zsuzsika; Smith, Samuel; Smits, Tim; Spies, Jeffrey; Sreekumar, Vishnu; Steltenpohl, Crystal; Stenhouse, Neil; Świątkowski, Wojciech; Vadillo, Miguel; van Assen, Marcel; Williams, Matt; Williams, Samantha; Williams, Donald; Yarkoni, Tal; Ziano, Ignazio, and Zwaan, Rolf (2018): “Justify Your Alpha”. In: *Nature Human Behavior* 2 (3), pp. 168 – 171.

Lazard, Lisa and McAvoy, Jean (2017): “Doing Reflexivity in Psychological Research. What’s the Point? What’s the Practice?”. In: *Qualitative Research in Psychology* (Early Access), pp. 1 – 19.

Linkola, Pentti (2009): *Can Life Prevail? A Radical Approach to the Environmental Crisis*. London: Integral Tradition Publishing.

Locke, Steph and D’Agostino McGowan, Lucy (2017): “DatasauRus. Datasets from the Datasaurus Dozen”. In: *Datasaurus*. https://itsalocke.com/datasaurus/, retrieved on November 14, 2018.

Lombrozo, Tania (2013): “Name Five Women in Philosophy. Bet You Can’t”. In: *NPR*. https://www.npr.org/sections/13.7/2013/06/17/192523112/name-ten-women-in-philosophy-bet-you-can-t/, retrieved on November 14, 2018.

McGuire, William (1973): “The Yin and Yang of Progress in Social Psychology. Seven Koan”. In: *Journal of Personality and Social Psychology* 26 (3), pp. 446 – 456.

Mead, George (1934): *Mind, Self, and Society*. Chicago: The University of Chicago Press.

Meldrum, Marcia (2000): “A Brief History of the Randomized Controlled Trial. From Oranges and Lemons to Gold Standard”. In: *Hematology, Oncology Clinics of North America* 14 (4), pp. 745 – 760.

Miedema, Frank (2012): *Science 3.0. Real Science, Real Knowledge*. Amsterdam: Amsterdam University Press.

Munafò, Marcus and Smith, George Davey (2018): “Repeating Experiments Is Not Enough”. In: *Nature* 553 (7689), pp. 399 – 401.

National Academies (2009): *On Being a Scientist. A Guide to Responsible Conduct in Research*. Washington: National Academies Press.

Nature (2018): “Nature Journals Tighten Rules on Non-Financial Conflicts”. In: *Nature* 554 (7690), p. 6.

Open Science Collaboration (2015): “Estimating the Reproducibility of Psychological Science”. In: *Science* 349 (6251), aac4716.

Patniak, Esha (2013): “Reflexivity. Situating the Researcher in Qualitative Research”. In: *Humanities and Social Science Studies* 2 (2), pp. 98 – 106.

Peterson, David (2016): “The Baby Factory. Difficult Research Objects, Disciplinary Standards, and the Production of Statistical Significance”. In: *Socius, Sociological Research for a Dynamic World* 1 (2), pp. 1–10.

Peterson, Robert (2001): “On the Use of College Students in Social Science Research. Insights from a Second-Order Meta-Analysis”. In: *Journal of Consumer Research* 28 (3), pp. 450–461.

Polka, Jessica; Kiley, Robert; Konforti, Boyana; Stern, Bodo, and Vale, Ronald (2018): “Publish Peer Reviews”. In: *Nature* 560 (7720), pp. 545–547.

Pool, Robert; Montgomery, Catherine; Morar, Neetha; Mweemba, Oliver; Ssali, Agnes; Gafos, Mitzy; Lees, Shelley; Stadler, Jonathan; Crook, Angela; Nunn, Andrew; Hayes, Richard, and McCormack, Sheena (2010): “A Mixed Methods and Triangulation Model for Increasing the Accuracy of Adherence and Sexual Behaviour Data. The Microbicides Development Programme”. In: *PLoS ONE* 5 (7), e11600.

Popper, Karl (1973): *Objective Knowledge. An Evolutionary Approach*. Oxford: Clarendon Press.

Porter, Theodore (1995): *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Postmes, Tom and Spears, Russell (1998): “Deindividuation and Antinormative Behavior. A Meta-Analysis”. In: *Psychological Bulletin* 123 (3), pp. 238–259.

Probst, Barbara (2015): “The Eye Regards Itself. Benefits and Challenges of Reflexivity in Qualitative Social Work Research”. In: *Social Work Research* 39 (1), pp. 37–48.

Ravetz, Jerry (2004): “The Post-Normal Science of Precaution”. In: *Futures* 36 (3), pp. 347–357.

Roberts, Brent (2015): “The Perverse Incentives That Stand as a Roadblock to Scientific Reform”. In: *The Inquisitive Mind* 25 (3). http://www.in-mind.org/article/the-perverse-incentives-that-stand-as-a-roadblock-to-scientific-reform/, retrieved on November 14, 2018.

Rosenthal, Robert (1979): “The File Drawer Problem and Tolerance for Null Results”. In: *Psychological Bulletin* 86 (3), pp. 638–641.

Rubin, Mark (2017): “When Does HARKing Hurt? Identifying When Different Types of Undisclosed Post Hoc Hypothesizing Harm Scientific Progress”. In: *Review of General Psychology* 21 (4), p. 308.

Rustin, Michael (2016): “The Neoliberal University and Its Alternatives”. In: *Soundings* 63 (3), pp. 147–176.

Ryan, Louise and Golden, Anne (2006): “‘Tick the Box Please’. A Reflexive Approach to Doing Quantitative Social Research”. In: *Sociology* 40 (6), pp. 1191–1200.

Sarewitz, Daniel (2016): “Saving Science”. In: *The New Atlantis* 49 (1), pp. 4–40.

Schiltz, Marc (2018): “Science Without Publication Paywalls. cOAlition S for the Realisation of Full and Immediate Open Access”. In: *PLoS Med* 15 (9), e1002663.

Shuval, Kerem; Harker, Karen; Roudsari, Bahman; Groce, Nora; Mills, Britain; Siddiqi, Zoveen, and Shachak, Aviv (2011): “Is Qualitative Research Second Class Science? A Quantitative Longitudinal Examination of Qualitative Research in Medical Journals”. In: *PLoS ONE* 6 (2), e16937.

Silberzahn, Raphael; Uhlmann, Eric; Martin, Daniel; Anselmi, Pasquale; Aust, Frederik; Awtrey, Eli; Bahník, Štěpán; Bai, Feng; Bannard, Colin; Bonnier, Evelina; Carlsson, Rickard; Cheung, Felix; Christensen, Garret; Clay, Russ; Craig, Maureen; Rosa, Anna

Dalla; Dam, Lammertjan; Evans, Mathew; Cervantes, Ismael Flores; Fong, Nathan; Gamez-Djokic, Monica; Glenz, Andreas; Gordon-McKeon, Shauna; Heaton, Tim; Eriksson, Karin Hederos; Heene, Moritz; Mohr, Alicia Hofelich; Högden, Fabia; Hui, Kent; Johannesson, Magnus; Kalodimos, Jonathan; Kaszubowski, Erikson; Kennedy, Deanna; Lei, Ryan; Lindsay, Thomas; Liverani, Silvia; Madan, Christopher; Molden, Daniel; Molleman, Eric; Morey, Richard; Mulder, Laetitia; Nijstad, Bernard; Pope, Bryson; Pope, Nolan; Prenoveau, Jason; Rink, Floor; Robusto, Egidio; Roderique, Hadiya; Sandberg, Anna; Schlüter, Elmar; Schönbrodt, Felix; Sherman, Martin Felix; Sommer, Amy; Sotak, Kristin; Spain, Seth; Spörlein, Christoph; Stafford, Tom; Stefanutti, Luca; Täuber, Susanne; Ullrich, Johannes; Vianello, Michelangelo; Wagenmakers, Eric-Jan; Witkowiak, Maciej; Yoon, Sangsuk, and Nosek, Brian (2018): "Many Analysts, One Data Set. Making Transparent How Variations in Analytic Choices Affect Results". In: *Advances in Methods and Practices in Psychological Science* 1 (3), pp. 337 – 356.

Simmons, Joseph; Nelson, Leif, and Simonsohn, Uri (2011): "False-Positive Psychology. Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant". In: *Psychological Science* 22 (11), pp. 1359 – 1366.

Singh, Simon (2004): *Big Bang. The Most Important Scientific Discovery of All Time and Why You Need to Know About It*. London and New York: Fourth Estate.

Sismondo, Sergio (2008): "How Pharmaceutical Industry Funding Affects Trial Outcomes. Causal Structures and Responses". In: *Social Science & Medicine* 66 (9), pp. 1909 – 1914.

Steegen, Sara; Tuerlinckx, Francis; Gelman, Andrew, and Vanpaemel, Wolf (2016): "Increasing Transparency Through a Multiverse Analysis". In: *Perspectives on Psychological Science* 11 (5), pp. 702 – 712.

Stroebe, Wolfgang; Gadenne, Volker, and Nijstad, Bernard (2018): "Do Our Laws Only Apply to Students? The Problem with External Validity". In: *Basic and Applied Psychology* 40, pp. 384 – 395.

Stulp, Gert (2017): "Visualising Differences in Visualisations in Social Psychology". Talk given at the University of Groningen.

Sullivan, Gail and Feinn, Richard (2012): "Using Effect Size or Why the p Value Is Not Enough". In: *Journal of Graduate Medical Education* 4 (3), pp. 279 – 282.

Tetlock, Philip (1994): "Political Psychology or Politicized Psychology. Is the Road to Scientific Hell Paved With Good Moral Intentions?". In: *Political Psychology* 15 (3), p. 509.

Thomas, Nancy (2010): "The Politics of Academic Freedom". In: *New Directions for Higher Education* 152, pp. 83 – 90.

Tsekeris, Charalambos (2010): "Reflections on Reflexivity. Sociological Issues and Perspectives". In: *Contemporary Issues* 3 (1), pp. 28 – 37.

Tufte, Edward (2001): *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.

Waldrop, Mitchell (2008): "Science 2.0". In: *Scientific American* 298 (5), pp. 68 – 73.

Walker, Susan; Read, Susan, and Priest, Helena (2013): "Use of Reflexivity in a Mixed-Methods Study". In: *Nurse Researcher* 20 (3), pp. 38 – 43.

Weissgerber, Tracey; Milic, Natasa; Winham, Stacey, and Garovic, Vesna (2015): "Beyond Bar and Line Graphs. Time for a New Data Presentation Paradigm". In: *PLoS Biology* 13 (4), e1002128.

Zimbardo, Philip (1969): “The Human Choice. Individuation, Reason, and Order Versus Deindividuation, Impulse, and Chaos”. In: *Nebraska Symposium on Motivation* 17, pp. 237 – 307.
van Zomeren, Martijn (2016): *From Self to Social Relationships. An Essentially Relational Perspective on Social Motivation*. Cambridge: Cambridge University Press.
Zwaan, Rolf (2013): “50 Questions About Messy Rooms and Clean Data”. In: *Rolf Zwaan*. https://rolfzwaan.blogspot.com/2013/08/50-questions-about-messy-rooms-and.html, retrieved on November 14, 2018.