Meta-Psychology, 2026, vol 10, MP.2023.4097
https://doi.org/10.15626/MP.2023.4097
Article type: Commentary

Open data: Not Applicable
Open materials: Not Applicable
Open and reproducible analysis: Not Applicable
Open reviews and editorial process: Yes
Preregistration: No

Edited by: Rickard Carlsson
Reviewed by: Nicholas JL Brown, Kai Horstmann
Analysis reproduced by: Not Applicable
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/NRP6H

An analysis of functional relationships between systemic conditions and unethical behavior in German academia

Nicole Bössel-Debbert[1], Annette Kluge[2], Daniel Leising[3], Dorothee Mischkowski[4], Le Vy Phan[5], Franz Richter[3], Manfred Schmitt[6], and Jutta Stahl[7]
[1]University Medicine Greifswald
[2]Ruhr University Bochum
[3]Technische Universität Dresden
[4]MPI for Research on Collective Goods, Bonn; Leiden University
[5]Bielefeld University
[6]University of Kaiserslautern-Landau
[7]University of Cologne

This paper is an updated English version of a report filed by a commission that the German Psychological Society (DGPs) appointed in 2022. The commission's task was a) to identify factors in the academic system that enable and/or promote unethical behavior, and b) to propose corrective measures. Based on expert interviews, a literature review, feedback from the community, and discussions within the commission, the following problematic issues were identified: (P1) negligent or fraudulent research practices, (P2) abuse of power, (P3) inadequate supervision of Early Career Researchers, (P4) poor quality of teaching, (P5) counterproductive incentives, (P6) overburdening of professors with tasks, (P7) fixed-term and short-term employment, (P8) unnecessarily strong power imbalance, (P9) an ineffective peer-review system, (P10) questionable assessment practices in hiring professors, (P11) lack of clarity of, and low commitment to, ethical standards, and (P12) weak control and sanctioning mechanisms. The commission offers concrete recommendations for changes to the academic system that should make the occurrence of unethical behavior less likely. Even though some of these recommendations might be rather specific for the field of psychology and for the German academic system, most of them may be valid for science in general.

Keywords: Science, Reform, Meta-Science, Misconduct, System

Background

In January 2022, the board of the German Psychological Society (DGPs) appointed a commission tasked with (a) identifying the structural conditions that enable, or even promote, unethical behavior in academia, and (b) proposing concrete measures for improvement. In June 2023, the full report was published online, alongside a commentary by the DGPs leadership board (https://www.dgps.de/fileadmin/user_upload/PDF/Berichte/Bericht_AMWF20230626.pdf).The present paper is an abridged and updated version of that report.

The commission used a broad definition of unethical behavior as conduct whose short- and long-term consequences contradict widely accepted ethical imperatives, specifically the notion that scientific work should serve the common good and that harm to individuals should be avoided. This definition thus covers scientific misconduct (e.g., data forgery, plagiarism, guest authorship) as well as bullying, sexual harassment, corruption, and exploitation. Notably, it also covers bad (e.g., negligent, redundant) research and bad supervision of Early Career Researchers (ECRs).

The reason for installing the commission in the first place was the widely shared impression that unethical behavior in science is not rare, sometimes severe, and most often remains undetected and unsanctioned. This impression was fed both by a continuing stream of media reports about several quite drastic cases, and by findings of systematic research on the topic. Nielsen et al. (2010) discussed that, regardless of the workplace type (academic or non-academic), accurately estimating the prevalence of unethical workplace behavior, such as bullying, harassment, discrimination, and sexual violence, is non-trivial and complicated by conceptual (e.g., definitional, terminological) and methodological issues (e.g., measurement, recruitment of non-random and non-representative samples, affective bias). In their meta-analysis, these authors estimated that, on a global

average (based on 86 samples; 130,973 respondents), "at least 1 out of 10, and maybe as many as about 1 out of 5, workers are exposed to bullying in their workplaces" (p. 967). Thus, their prevalence estimate ranged from about 10 to 20 percent. A study in the US (Basile et al., 2020) estimated that about 5.6 percent of women and 2.5 percent of men have experienced some form of sexual violence by a workplace-related perpetrator in their lifetimes. A more recent meta-analysis by Dhanani et al. (2021) reported significantly higher prevalence estimates for "workplace mistreatment" (34 percent experienced, 44 percent witnessed). Notably, these authors did not find much evidence for sampling bias (i.e., systematic over- or under-reporting due to how samples were drawn).

Focusing more specifically on academic work environments, Lasser et al. (2021) estimated that 10 to 13 percent of PhD students (PhD network MPI) have been bullied by a superior. In a recent survey specifically addressing researchers working in psychology (follow-up: Stahl et al., 2024; initial study: Elson et al., 2021), 8 to 10 percent of the members of the German Psychological Society reported that they had experienced abuse of power, sexual or psychological types of harassment/bullying, or discrimination in academia.

At present, the question of whether academia, or German academia, is a particularly fertile breeding ground for unethical behavior cannot be conclusively answered, due to a lack of comparative studies. The present report is based on the premise that the prevalence of unethical behavior in (German) academia is substantial, and that there may be parameters of the current academic system that permit or even exacerbate and reward such behavior. The commission's task was to identify those parameters and to make suggestions for structural reforms.

Proces and Methods

Between its appointment and the delivery of its first report, the commission held 21 extended meetings. In its work, it relied on a qualitative approach, by

1. Collecting, reading, analyzing and discussing relevant literature;
2. Evaluating particularly problematic case examples as well as best practices;
3. Discussing personal experiences that commission members had had while serving in various academic functions;
4. Consulting experts on specific topics (i.e., power imbalances, employment relationships and fixed-term contracts, control and sanctioning mechanisms, personnel selection;
5. Soliciting feedback from the community.

The Acknowledgements section at the end of this manuscript lists those individuals who supported the commission's work by sharing their views and experiences. The analyses and recommendations that the commission developed constitute the main focus of this paper and are presented below.

Basic Concepts

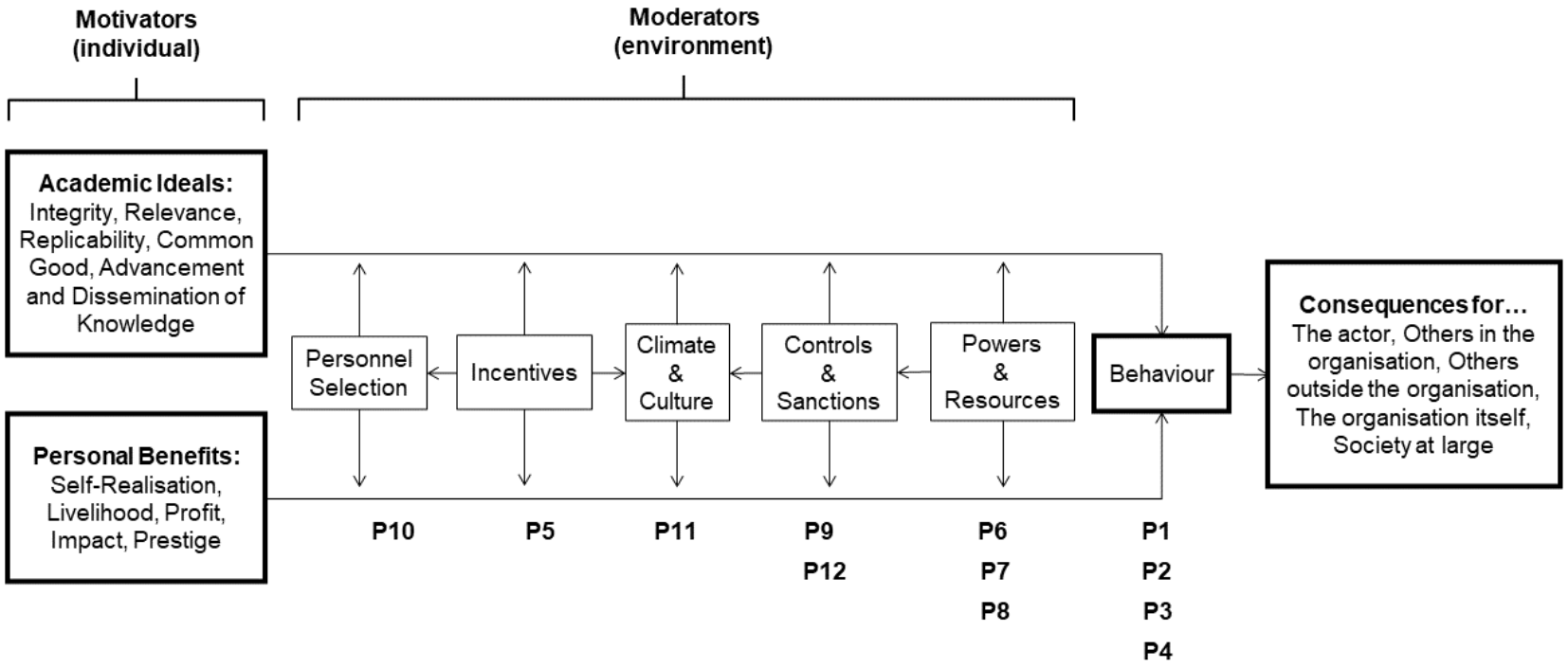
Figure 1 depicts a heuristic model of behavior regulation, incorporating variables that the commission identified as being particularly relevant in the given context. According to the model, the behavior of individuals in academic institutions may be fueled primarily by two different kinds of "motivators":

- *Academic ideals* such as contributing to the growth and transmission of knowledge for the common good. People may exhibit certain types of behavior (e.g., error monitoring and correction) in the service of such ideals. Notably, the very existence of public research institutions is justified primarily through the extent to which the behavior of their members serves the achievement of these goals.
- *Prospective personal benefits* such as financial profit, prestige, or making a livelihood. People may exhibit certain behaviors (e.g., covering up mistakes) in the service of these motives.

Presumably, the behavior of most individuals working in academic institutions is influenced by motivators from both domains, although the relative strength of the respective influences may differ between individuals. Moreover, the influences of the two kinds of motivators may be well aligned with each other (e.g., when researchers' desire for recognition by others in the scientific community motivates them to collect particularly strong data), but they may also come into conflict (e.g., when researchers use questionable scientific practices to increase their chances of getting their research published and thus getting tenure).

Figure 1

Model of behavioral regulation in academia (with two types of motivators on the individual side and five moderators on the work environment side), plus references to twelve problem areas identified in this report (P1 to P12; see text for details); ECRs = Early Career Researchers.



However, human behavior depends not only on the motives of individuals, but also on the conditions under which they operate. Key environmental factors that the commission considered in its analyses are also contained in Figure 1 ("Moderators"):

1. Personnel Selection. How people (with specific behavioral inclinations and skills) are recruited to, and retained by, the organization;
2. Incentives. Which behavior is rewarded (directly or indirectly);
3. Climate and Culture. Which norms and values guide the behavior of the organization's members and are viewed as being common by them.
4. Control and Sanctions. Whether unethical behavior is likely to be detected and, if necessary, sanctioned;
5. Powers and Resources. How much liberty and discretion members of the organization have to behave as they please.

These contextual conditions of the current academic system may move the behavior of the people working in it away from its original purposes (i.e., to serve the aforementioned academic ideals). The more these original purposes are sacrificed, however, the more the legitimacy of this kind of work is at risk of getting lost.

Consequently, the behavior resulting from these interactions between motivators and moderators should be considered problematic to the extent that it leads to ethically unacceptable consequences. This extent may be determined by comparing the long-term consequences that the respective behavior will likely have for different parties (see Figure 1). For example, negligent or fraudulent researcher behavior should be considered problematic because it consumes resources primarily to the personal benefit of the researcher (career-wise), without yielding any return on investment (e.g., a relevant increase in knowledge) to those who contributed these resources (e.g., taxpayers).

Twelve particularly important "problem areas" (abbreviated P1 to P12) were identified by the commission. They are listed in Figure 1 and discussed in more detail in the second part of the present paper. Note that our analysis does not distinguish between individual academic institutions, but rather aims to describe typical institutional conditions. The arrows connecting some moderators with one another in Figure 1 are intended to symbolize that these variables may affect each other (further connections and directions are possible).

Multiple Nestedness

The behavior of individuals in academic institutions is nested within several layers of rules, regulations, and guidelines that have been established at the various levels of the academic system. Also, various actors may influence the behavior of individuals in the system, by enforcing these rules, regulations, and guidelines. Figure 2 demonstrates this for the German academic environment.

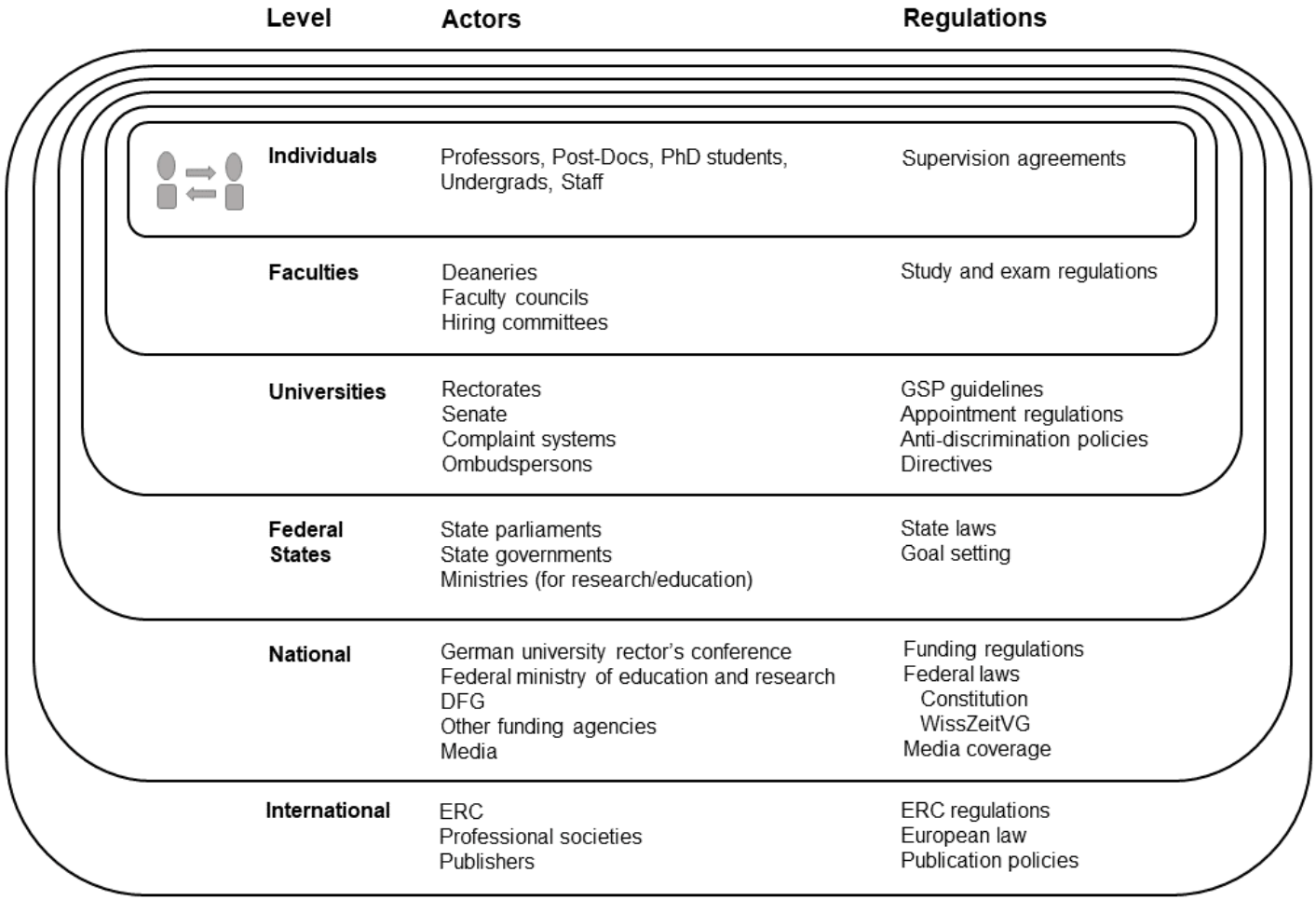
The relevant actors primarily include three groups of people: (1) people who have been given primary responsibility for tasks in the context of research, teaching, and training of young researchers (usually professors), (2) people who are currently preparing for such a position (they are referred to as Early Career Researchers (ECRs) in the present paper), and (3) students. They all interact with one another on a daily basis against the background of the institutional environment and its implicit and explicit reward structure.

The behavior of these different groups of actors is situated within the regulations of the faculties and universities. These, in turn, belong to the jurisdiction of the responsible ministries of the federal states, and at the same time they are also part of the higher-level (national and international) science system.

We consider this multiple nestedness relevant for the present analyses in at least two regards: First, for all interventions meant to alleviate a problem of the academic system, it will be necessary to identify the relevant level at which they should take place. Second, changes taking place at one level may conflict with regulations or expectations located at other levels. For example, if a hiring committee decided to use criteria other than the usual ones (such as number of authorships, citations, impact factors, and acquired grant money), it might work against the interests of the faculty and the university, which – at present – are themselves evaluated largely on the basis of these criteria (www.topuniversities.com/university-rankings-articles/world-university-rankings/world-university-ranking-methodologies-compared).

Figure 2

Multiple nestedness of relevant actors and institutions in the German science system. DFG = Deutsche Forschungsgemeinschaft (German Research Foundation), ERC = European Research Council, GSP = Good Scientific Practice, WissZeitVG = Wissenschaftszeitvertragsgesetz (German Act on Temporary Scientific Contracts).



Detailed Analysis of Existing Problems and Suggested Solutions

The overall goal of the commission's work was to derive concrete recommendations on how to improve the academic system. The remainder of this paper contains detailed analyses of 12 "problem areas", along with the commission's recommendations for improvement.

Our analyses begin with some particularly problematic kinds of behavior and their consequences (P1 to P4). Their existence is basically the cause and the starting point for all of the following elaborations. We then analyze various characteristics of the academic work environment (P5 to P12) that may contribute to the emergence and/or persistence of this problematic behavior and its consequences (see Figure 1).

Notably, the twelve problem areas that the commission identified are interrelated in complex ways. Often, a problematic outcome is brought about by several problematic system features, and sometimes it is even possible to delineate entire "chains" of effects. For example, the overburdening of professors with tasks (P6) may lead to reduced diligence in peer review (P9), and this low diligence (P9) may then make it more likely for other researchers to get away with questionable research practices or outright fraud (P1).

Moreover, the same feature of the academic system may be the cause of several problematic outcomes. As a consequence, a proposed solution targeting one cause may help alleviate several problems at once. For example, the large power differentials between senior and junior researchers (P8) not only enable various forms of power abuse (P2), but also impair the functioning of control mechanisms (P12) (i.e., most junior researchers do not dare to file complaints). In the following, we hint at these complexities occasionally, but acknowledge that a comprehensive analysis of this multilateral interplay would exceed the scope of the present paper.

The current paper reflects the commission's informed opinion. However, not all of the recommendations were considered equally helpful by all members of the commission. The consequences of some of the proposed changes are difficult to foresee in a system as complex as science. In some cases, only a combination of several measures may lead to the desired improvement, while implementing individual measures in an isolated fashion may even lead to a deterioration.

Some of the proposed measures (e.g., supervisor training) may incur increases in workload, whereas others (e.g., limits on authorships or advisorships) may actually help reduce workload. Again, we occasionally hint at this, but abstain from attempting a comprehensive analysis.

Furthermore, the different types of measures (e.g., changes in faculty policies vs. changes to the law) will require different timelines to be implemented, which is why their (likely) effects must be monitored both individually and in combination with each other, at different points in time. All of this introduces a significant level of uncertainty. The commission proposes that systematic research accompanying the suggested measures and their effects should be conducted from the start.

Given the staggering complexity of the overall task, it is impossible to tie all of the recommendations presented here to empirical evidence. Many of these measures have never been implemented, but this may not be used as an argument against ever trying them.

Due to how the current process was initiated (see above), the present paper does focus on the situation within psychology, and within the German academic system. Based on the many discussions we had, however, we are convinced that most of the issues we address are not specific to a given field or a given national system, but rather universal. We will come back to this aspect later.

Problem Area 1. Negligent or Fraudulent Research Practices (P1)

Most scientific research is supposed to generate robust, relevant knowledge for the benefit of society. However, many published studies lack methodological quality and published results often cannot be replicated (Errington et al., 2021; Open Science Collaboration, 2015).

Probably the main reason for this lack of replication lies in the everyday (mis)behavior of researchers, especially their use of so-called Questionable Research Practices (QRPs) (e.g., passing off random findings as effects that were predicted by theory). The line between QRPs and outright fraud is not clear-cut, with the former being commonly perceived as less severe and thus more permissible. While there is evidence for a widespread use of QRPs (Fanelli, 2009; John et al., 2012), the amount of fraud is more difficult to determine, for obvious reasons. It should be noted, however, that even the use of QRPs may suffice to discredit a piece of scientific work entirely, and should thus be generally regarded as unethical (Miller et al., 2025).

In addition to non-replicability, which is a problem shared by other scientific disciplines (e.g., Errington et al., 2021), psychology has a number of more specific weaknesses such as a lack of precision (formalization) in many of its theories and a confusing landscape of constructs, terminology, and measures (Anvari et al., 2025; Block, 1995; Flake & Fried, 2020). This lack of binding standards may easily be exploited for one's own benefit (e.g., by choosing an analysis pipeline that maximizes

publishability rather than informativeness). Thus, research is less efficient than it should be, and resources are wasted to a significant extent. In addition, such deficiencies may harm other researchers when they have to base their own research on unsound findings published by their colleagues. Ultimately, the public's trust in psychology and empirical science as a whole may decline, as may the identification of the next generation of potential researchers with their own discipline (Brachem et al., 2022).

Causes / Contributing Factors

Negligent and fraudulent research practices may be promoted by a variety of factors:

1. Academia is largely lacking a constructive error-management culture (Frank et al., 2024). To learn from mistakes (e.g., in experimental programs or data processing), open communication about these mistakes is necessary. However, fear of negative consequences (e.g., financial penalties, loss of face), may get in the way of such openness and lead to concealment of mistakes or even deliberate deception about mistakes that have happened.

2. Another highly relevant factor is the current incentive structure in (German) academia: Both the careers of individual researchers and the evaluations of academic institutions are strongly tied to primarily quantitative metrics such as the number of authorships and citations, impact factors, and grant money. However, the validity of these metrics as measures of research productivity (especially in the sense of generating reliable knowledge gains) is highly questionable. They are also easily and commonly manipulated (Borkenau, 2012; Chapman et al., 2019; Fong & Wilhite, 2017; John et al., 2012; Pruschak & Hopp, 2022).

3. The cost of conducting and publishing studies of dubious scientific value is borne not by those who conduct and publish them, but by other researchers and by the public.

4. The risk of detection and sanction is low, since traditional peer review seems unable to enforce good scientific practice (see P9) and the effort necessary for identifying the use of QRPs would be very high in the present reviewing system. The oversight mechanisms currently in place seem to be rather ineffective (see P12).

5. The overburdening of professorships with tasks (see P6) may promote the use of QRPs: Due to the multitude of tasks that they have to engage with every day, many researchers simply lack the time to take proper steps to promote research quality (e.g., preliminary studies to estimate effect sizes, sufficiently specific pre-registrations, replications, thorough theoretical deliberations, or even just reading others' work).

6. The occurrence of QRPs, as well as their non-detection, may also be promoted by the existing power differentials in research teams (see P8). Scientific misconduct is often not reported because potential whistleblowers fear retaliation (e.g., in the form of non-renewal of their fixed-term contracts, see P7). In the survey by Elson et al. (2021), some supervisors were described as explicitly asking employees to manipulate data (see also Stahl et al., 2024).

7. The practice of giving mostly fixed-term contracts with relatively short durations to ECRs (see P7) creates a certain pressure continuously to produce new research successes in order to secure one's continued employment. This may lead some ECRs to lower their ethical standards and adopt QRPs to save their careers.

8. Personal integrity does not yet play an important enough role as a selection criterion in academic hiring procedures. In fact, these procedures often rely heavily on metrics that are easily manipulated (see P5, P10). With little effective oversight in place (see P9, P12), there may thus be a risk of selecting personnel with particularly low ethical standards.

9. Even when scientific misconduct is absent, vague theory and terminology, as well as arbitrary measuring practices, will reduce the refutability of one's hypotheses, which in turn will improve the chances of having one's research published.

Recommendations

(1.1.) Engage in more active culture assessment and development. This may include awarding certificates like "great place to work" or "or great place to do research" (note that these are separate issues). In this, it will be important to enable independent evaluations (e.g., by letting assessors interview organization members of their own choosing, anonymously) and to take measures to avoid positive evaluations that are rooted in the assessors' own interests (e.g., of being hired again). The relevant criteria and process must be transparent, continuously monitored and, if necessary, further developed.

(1.2.) Provision of a mature infrastructure for the implementation of Good Scientific Practices (such as Pre-Registration and Open Data) by universities and other research institutions, as well as third party-agents (e.g., ZPID, Open Science Framework).

(1.3.) Explicit recognition and consideration of the additional workload necessary to comply with higher research quality standards (e.g., larger samples, replication trials) in the context of grant applications.

(1.4.) Promotion of training in Good Scientific Practice for individual researchers (e.g., as a routine part of ECR training, and as compulsory professional development for professors).

(1.5.) Good Scientific Practice and Open Science should be key topics in teaching, not only as theoretical concepts, but also in application (e.g., as part of student research projects) (Scheffel et al., 2023).

(1.6.) The teaching of psychology must be significantly strengthened in terms of theory formation, theory specification (formalization), and theory testing. The commission recommends fostering this development through cooperation with other fields (mathematics, philosophy, physics, computer science) whose members possess the relevant expertise.

(1.7.) The problem of the confusing landscape of terms, constructs and measures in psychology can only be solved through coordinated efforts. It is necessary to (a) document the multiple existing redundancies and ambiguities of meaning (e.g., Cain et al., 2008; Eva et al., 2025; Hodson et al., 2018, and then (b) reduce them by gradually establishing more explicit consensus on term, construct, and measurement use (see Leising et al., 2024 for suggestions on how such a process may be organized). Due to the high degree of complexity of this task (Leising et al., 2022b, 2022a) and the political dimension involved, it may be necessary for journals and/or professional societies to make this a priority point on their agenda, and to explicitly mandate and massively incentivize such a development (e.g., by creating high public visibility).

(1.8.) Develop and gradually implement elements of a more constructive error-management culture. Here, one may borrow from existing approaches in other fields such as medicine (e.g., web-based error reporting systems; Hoffmann et al., 2008). Generally speaking, the making of errors needs to be destigmatized, and error detection needs to be systematized and rewarded (Elson, 2024).

(1.9.) Establish a DGPs division focusing on Theoretical Psychology. This new group should not work separately from those addressing more specific content-related questions, but rather establish sustainable ties with them and attempt to provide helpful contributions to their work. A solid foundation in philosophy of science, logic, and mathematics is necessary for this, which requires the participation of appropriately qualified colleagues (including, ideally, colleagues from other fields, see 1.6). As an alternative to founding an entire new section, explicitly strengthening this subject within the already existing Methodology division may be considered.

(1.10.) Employ the Research Quality Evaluation (RESQUE) collector App (Gärtner et al., 2025) to showcase the good science merits of a researcher's / department's / university's publications, independently of the peer review conducted by scientific journals.

(1.11.) Establish the explication of all authors' actual contributions as a standard practice (Allen et al., 2014; https://credit.niso.org/). Note that adopting the CRediT system implies that the *order* of authors does not contain any additional information as compared to the CRediT roles themselves. One may thus move to alphabetical ordering as already common in some disciplines. Note further that the CRediT system does acknowledge types of contributions (i.e., supervision, funding acquisition, project administration) that were previously seen as *not* justifying authorship. Under the RESQUE framework, the size of the contribution that each author made in their respective area(s) is now supposed to be directly quantified, via a statement (Leising et al., 2025). Implementing these changes would be the responsibility of scientific journals.

Problem Area 2. Abuse of Power (P2)

Surveys at universities and research institutes (Arcudi et al., 2019; Beadle et al., 2020; Elson et al., 2021; Olsthoorn et al., 2020; Peukert et al., 2020; Regler et al., 2019; Scherpenberg et al., 2021; Schraudner et al., 2019; Stahl et al., 2024) suggest that abuse of power is not an exception in academia. Most abuses of power are committed by academics already firmly established in the system (usually professors), while the victims tend to be ECRs, students, and administrative staff (Elson et al., 2021; Stahl et al., 2024). This is unsurprising given the large power differential between the former and the latter groups (see P8). In addition, certain groups (e.g., racialized persons; first-generation academics) seem to be at an even greater risk of becoming victims (e.g., Niemann et al., 2020). However, we do acknowledge that ways exist for persons lower in the hierarchy to abuse the power that they do hold against persons higher up in the hierarchy (e.g., by refusing to cooperate, by badmouthing leaders).

Abuse of power can take many different forms in varying degrees of severity: (a) Misappropriating another person's intellectual property, including plagia-

rism as well as "guest" and "honorary" authorships, (b) Extorting extra work (e.g., delegating one's own teaching responsibilities to ECRs), (c) Demanding problematic research practices and shifting responsibility for them onto the dependent person in case of detection, (d) Bullying (e.g., threats, humiliation), (e) Physical abuse, and (f) Sexual harassment (Stahl et al., 2024). In addition to those directly affected by such behavior, others may be harmed more indirectly (e.g., when the detection of the abuse of power by one person harms the reputation of colleagues working at the same department).

Causes / Contributing Factors

Abuse of power in the academic context often results from an interplay of several factors:

1. Some personality dispositions (especially a readiness to maximize one's own personal benefit at the expense of others, overstepping boundaries and violating rules and even laws in doing so; Moshagen et al., 2018; Paulhus and Williams, 2002) make abuses of power more likely. Research on business leaders clearly shows the relevance of these traits (Palmer et al., 2020); their importance in higher education contexts, on the other hand, is only recently being considered more systematically (Cohen & Baruch, 2022; Forster & Lund, 2018). Selection procedures for academic leadership positions do not yet cover this domain (see P10).

2. Some structural characteristics of the science system (especially in combination with the abovementioned personal characteristics) may promote the abuse of power (Cohen & Baruch, 2022; Lasser et al., 2021). A central factor in this regard is the unnecessary centralization of too much power in the hands of professors (see P8). This makes it very risky for potential complainants to voice any concerns because they must fear retaliation (e.g., non-renewal of fixed-term contracts, see P7) by the person whose behavior they complain about (Horbach et al., 2020).

3. The existing mechanisms for detecting and sanctioning abuses of power in academia are weak and largely ineffective (see P12). The responsible bodies within academic institutions tend to be understaffed, have hardly any power to impose sanctions, and are often not even known to potential complainants. Surveys also show that these "internal" complaint channels are not trusted much, and that their use tends to produce results that are rarely experienced as satisfactory by complainants (Elson et al., 2021; Höbel et al., 2022). The vast majority of misconduct cases thus remains unreported and unsanctioned. The knowledge of this low effectiveness of the existing control mechanisms is in itself likely to promote abuses of power, as perpetrators can feel relatively safe from consequences.

4. Another important factor that likely fosters abuses of power in academia is the general lack of awareness of, and commitment to, certain ethical standards (see P11). In the course of the past few decades, it seems to have been a widely accepted tenet that professors are (and should be) allowed to demand a lot from their employees in almost every respect, since the path to a professorship "is rocky" ("it's always been that way", "we all had to go through it"). The line between legitimate uses and illegitimate abuses of power may thus be blurred, which makes detection and sanctioning of the latter less likely.

5. The common use of fixed-term contracts for ECRs (see P7) creates a situation of high job insecurity and – desired – personnel turnover (see also Rahal et al., 2023). Given the fact that institutional grievance procedures usually take months or even years, a typical outcome is that such procedures have not yet concluded when the complainant's contract expires, and are then simply abandoned with no palpable effects on the – permanently employed – accused.

6. Many forms of power abuse in academia come at little cost to the perpetrator, while at the same time being highly likely to "pay off". This concerns, for example, the practice of coercing undeserved authorships from people in less powerful positions (e.g., as reviewer, editor, or advisor). The likelihood of ever being "caught" with this type of behavior is very low. At the same time, long publications lists are a major asset in trying to attain a permanent position, obtain research funding, or just get a raise (see P5).

Recommendations

The commission did not identify any measures for improvement that would be *uniquely* suited for fighting power abuse in academia. However, many of the measures that are proposed in the other 11 chapters of this report would serve that goal, too. We attempt to point out when that is the case.

Problem Area 3. Inadequate Supervision of ECRs (P3)

Some ECRs experience a severe lack of competent and responsible supervision by their advisors (Mackie & Bates, 2018). Deficiencies often named in this regard comprise (a) being pushed into bad scientific practice, (b) chronic inaccessibility of the supervisor, (c) lack of access to needed resources, (d) supervisors' lack of interest in the ECRs' research topics and/or personal and professional development, and (e) too little freedom for ECRs to make and be accountable for their own decisions (which is at odds with the idea that ECRs are supposed to learn how to work independently as researchers).

Poor supervision almost by definition tends to damage the ECRs concerned, both in terms of professional development and/or well-being. In the long run, however, it also has the potential to harm the scientific process, as inadequate supervision may lead to inefficient and low-quality work. In addition, the reputation of professors, faculties, and universities may be damaged when poor supervision becomes publicly known. Furthermore, negative supervision experiences may drive talented ECRs out of the scientific community (Christian et al., 2021; Kis et al., 2022), so that their qualifications will no longer be available to science.

Causes / Contributing Factors

1. The overburdening of professors with tasks (see P6) plays an important role: German professors often have so many responsibilities and are busy with so many things at once that they can only provide high-quality supervision for a very small number of ECRs, if any.

2. The current incentive system (see P5) rewards the completion of as many doctorates as possible without ever raising the question of supervision quality.

3. The power imbalance between professors and ECRs (see P8) can make ECRs reluctant to voice criticism of inadequate supervision, or to demand more and better supervision (e.g., more frequent and detailed feedback).

4. There is a relative lack of clarity as to what constitutes "good" supervision in the first place. As there are no well-established standards for this, ECRs often do not have a clear idea of what quantity and quality of supervision they may legitimately expect.

5. Personnel management and supervision skills do not play a relevant role in training on the path to becoming a professor, or in the assessment of applications in the context of hiring professors (P10).

6. It is often unclear whom ECRs may turn to if they are dissatisfied with their supervision, and what measures may be taken to alleviate such a situation. Sanctions for bad supervision do not exist (P12).

Recommendations

(3.1.) Let universities and other research institutions, the German Psychological Society, and institutions funding research (e.g., German Research Foundation, Federal Ministry of Research, Technology, and Space) declare a plausible maximum number of ECRs (e.g., five) who may collaborate scientifically with the same supervisor at the same time.

(3.2.) With an increasing number of ECRs, the time and energy that a supervisor may devote to each individual ECR is likely to decrease. There is no automatism here, but it is just not plausible that the number of ECRs supervised by the same person may be raised indefinitely without any loss in supervision quality. Therefore, a large number of PhD students working with the same supervisor should only be admitted under special conditions (e.g., co-supervision), and the quality of supervision should then be regularly and carefully checked (e.g., by interviewing the ECRs concerned, see 3.7.).

(3.3.) Devise and disseminate explicit guidelines regarding the cooperation between ECRs and professors.

(3.4.) Increase the use of supervision agreements to increase the predictability and controllability of scientific collaboration for all parties involved. In such agreements, essential parameters of the planned collaboration should be put in writing (e.g., preliminary research topic, modalities of changing topic and/or supervisor, access to resources, rights to use data, authorships, frequency and format of progress reports, feedback). Standard ways of dealing with complaints and conflicts, including responsibilities for handling them, should also be clarified. The commission has published a template for supervision agreements (https://www.dgps.de/fileadmin/user_upload/PDF/Empfehlungen/Handreichung_BV_zur_Weitergabe__Fassung_Mai_06_2024_.pdf).

(3.5.) Regular feedback to professors on the quality of their scientific cooperation with and supervision of ECRs, by the latter (e.g., in the context of semi-annual meetings, or based on the results of - possibly anonymous - surveys).

(3.6.) As long as the current practice of merging the roles of boss and scientific collaborator persists (the "supervisor" is both at once), it will be necessary to establish explicit, legally binding regulations of the procedure for changing supervisors, including the conditions under which each party may opt out of an existing collaboration, and provisions for enabling a successful completion of the candidate's PhD regardless.

(3.7.) Use of Thesis Advisory Committees (TACs) to establish independent oversight, documentation, and, if necessary, protection of the ECRs' research projects. Here, a group of other researchers monitors the progress of a doctoral project by way of regular, separate meetings with both parties (supervisor and PhD student). Notably, supervisors shall not be members of a TAC overseeing the progress of ECRs they work with. TACs may include members from other universities and/or disciplines – this seems even desirable as it may help promote the quality and interdisciplinarity of the project concerned. The establishment of a TAC should be a routine part of supervision agreements.

(3.8.) For academic positions with supervision responsibilities, the commission recommends systematically introducing validated tools for personnel development, such as 360° feedback. Given the potential of such assessment to be abused (e.g., to retaliate against a supervisor for perceived injustices), this should be tried primarily in the context of personnel development (rather than personnel selection).

Problem Area 4. Poor Quality of Teaching (P4)

Besides research, teaching is one of the central tasks of universities, as defined in German federal higher-education acts. However, complaints about the poor quality of academic teaching in Germany are not uncommon and include (a) lack of commitment by teaching staff, (b) outdated content, (c) content of questionable or obviously low relevance, (d) presentation formats that are not conducive to learning, (e) redundancies, as well as (f) unsuitable examination formats.

Those who are affected the most by these deficits are students, who not only have to deal with unnecessary frustrations and obstacles in the course of their studies, but also with the feeling of not being valued in their role and position. Also, learning success may fall short of what would be desirable and achievable. Another group of persons affected are those lecturers who are more committed to teaching than would be absolutely necessary, as their commitment is (apart from student feedback) rarely rewarded. In fact, time invested in teaching will not be available for research activities anymore, and professors who do invest much of their time this way are commonly regarded as being second-rate or "deadwood".

Causes / Contributing Factors

1. At present, high-quality teaching is not explicitly incentivized much within the German academic environment. For tenured professors, negative evaluations by students do not have any serious consequences. It is mostly up to the teaching staff to decide whether or what conclusions they want to draw from such feedback, or whether even to obtain such feedback in the first place. The easiest option is not to draw any conclusions, to disregard or not obtain feedback, and to continue as before.

2. Teaching skills play a subordinate role in academic hiring procedures at German universities (see P10). This can be seen as a direct reflection of the current incentive structure in which publications, citations, and third-party funding are given much greater weight in evaluations of academic institutions.

3. It is unclear how teaching quality should best be assessed. At present, teaching quality is not measured in terms of achieving long-term teaching goals (i.e., increasing the competencies of students) – which is the primary goal of teaching – but in terms of grades and pass rates, and in terms of evaluations by students. These criteria are prone to error: Typically, the content of classes and exams as well as grading criteria are determined by the same persons who also teach the classes. This creates a situation in which teachers may achieve high pass rates, good average exam grades, good teaching evaluations by their students, and a low number of complaints simply by lowering their standards (known as the "good grades in return for bad teaching" principle). In fact, this may be regarded as an instance of power abuse – this time for the benefit of students (at least superficially). The problem may be exacerbated by the fact that the mere number of graduations is commonly used as a metric of institutional performance for universities.

4. A final important factor is, once again, the overburdening of professors with tasks (see P6). Considering the number of tasks with which professors are regularly expected to engage, a rational response will be for them to reduce their effort in those areas where it will have the least unpleasant consequences for themselves (e.g., teaching).

Recommendations

(4.1.) Create explicit, effective rewards (including financial ones) for teachers who invest time and effort into the quality of their teaching.

(4.2.) Give greater weight to teaching quality in hiring procedures and in evaluations of faculties and universities.

(4.3.) Establish continuous training of teaching skills for teachers in academia. Participation should be encouraged or even made mandatory (e.g., once a year).

(4.4.) Academic institutions should proactively and individually approach their teachers, offering voluntary external evaluation and expert advice on their teaching.

(4.5.) Professors teaching the same subject should cooperate more closely with one another: They may streamline and harmonize the content of their courses, and work on creating shared teaching materials. Such exchanges would likely promote the quality of teaching overall (because each topic could be covered by the teacher with the greatest expertise, respectively), and reduce the workload for individual teachers. They may (and should) also result in the creation of Open Educational Resources (i.e., freely available teaching materials). The DGPs may take on a steering role here.

(4.6.) Criteria and methods for teaching evaluations (especially by students) should be critically scrutinized and optimized. Suggestions in this regard were made, for example, by Kreitzer and Sweet-Cushman (2021).

(4.7.) The question of appropriate forms of examination in psychology should also be critically reflected upon. Complaints by students about "presentation seminars" and multiple-choice examinations, for example, are common, but viable alternatives are hardly ever specified.

Problem Area 5. Counterproductive Incentives (P5)

Scientists regularly compete with each other for limited resources such as positions and project funding. In order to make decisions on the allocation of these resources as objective as possible, the use of so-called metrics of research productivity has become firmly established. However, common quantitative metrics such as publication and citation numbers, third-party funding totals, impact factors, and h-indices are increasingly seen as lacking validity, so much so that their use may actually be detrimental to the original goals of science (European Commission, Directorate-General for Research and Innovation, 2021; Leising et al., 2022b, 2022a)).

It is possible that stronger contributions to a domain's knowledge base have better chances of being published in prestigious ("high-impact") journals, and of becoming cited. However, the same outcome is often achieved by papers with relatively weak scientific content. This is partly due to the fact that quality assurance through peer review does not yet function well enough (see P9). Among the negative consequences of this situation are (a) a dubious scientific evidence base, and (b) biased academic hiring and promotion procedures to the disadvantage of researchers who put more emphasis on the quality, rather than the quantity, of their work.

Many of the commonly used metrics are easy to manipulate and are in fact commonly manipulated (Chapman et al., 2019; Falagas & Alexiou, 2008). For example, surveys show that the acquisition of so-called guest or honorary authorships by way of mutual favors or coercion is a common form of scientific misconduct (Fong & Wilhite, 2017; Gasparyan et al., 2013; Pruschak & Hopp, 2022). The benefits of this practice for individual researchers and the institutions they work for are obvious under the current incentive structure, while the cost and the risk of discovery and sanctions (see P12) is close to zero for everyone involved (i.e., those who grant guest authorships and those who accept them). The consequence is that authorships say very little about people's actual scientific contributions to a research paper – it may range from several years of intensive work to exactly zero. The mere number of a person's authorships should thus be viewed as a discredited metric, yet it continues to play a key role in research evaluations.

There also exist some counterproductive incentives in the area of third-party funding. At some academic institutions in Germany, success in acquiring third-party funding is actually the only relevant criterion by which the performance of professors is measured and possibly rewarded (e.g., with a tax-funded bonus to a researcher's income). This is because a certain percentage of these sums is typically added on top as an "overhead" to be used directly by the institution. German academic institutions may actually need that money, as they continue to be described as being short of budget (e.g., Vallbracht, 2024). As a consequence, it is rational for research institutions to spend as much money as possible as quickly as possible on their research, and to reward those researchers who help them achieve that goal (e.g., with bonuses to their salaries). Researchers whose research is the costliest are the most likely to be rewarded that way.

Notably, all of this happens irrespective of any potential returns on investment for the public that funds the whole endeavor, and it also happens within a system in which actual quality control is more or less unheard of. The current incentive structure is primarily oriented towards quantity and attaches too little importance to actual scientific progress in terms of content

(i.e., giving reliable answers to relevant scientific questions). This is also the case, for example, when professors, faculties, and universities are rewarded (e.g., in university rankings) for the sheer number of completed doctorates (Hachmeister, 2019), but decisions on the respective quality criteria are made solely by those same professors, faculties, and universities. This approach obviously entails a strong incentive to increase completion rates by lowering quality standards.

Causes / Contributing Factors

1. The current dysfunctional metrics were probably once introduced with the best of intentions, namely with the aim of establishing greater objectivity in assessing scientific productivity. However, their validity relies too much on the honesty of individual researchers, and on the functioning of the peer-review system. Apparently, it was underestimated how easily these metrics may be manipulated, and how willing many would be to do just that (Chapman et al., 2019; Fong & Wilhite, 2017; John et al., 2012; Pruschak & Hopp, 2022).

2. It is not uncommon for academics working in the system to justify their own acquiescence by saying that they are more or less powerless to change the current form of metrization, even though they actually share the growing criticism of it. Professors routinely refer to faculty or university administrations as the responsible agents, and these in turn refer to ranking agencies and to political leadership. This inclination to deny responsibility is a key factor that keeps the problem going.

Recommendations

(5.1.) Raise public – and especially political – awareness of the lack of validity of common metrics and the possibility of developing and using more valid metrics. This may be achieved by (e.g.) talking to the media, by writing blog posts or open letters, or via official petitions to parliaments.

(5.2.) Academic institutions and advocacy groups should publicly distance themselves from the use of invalid productivity metrics in science evaluation. One way of doing this would be to join the Coalition for the Advancement of Research Assessment (https://coara.eu/) which the German Psychological Society has already done.

(5.3.) Specify and start using more valid metrics, especially ones capturing the methodological rigor of studies (e.g., for representative data, sufficient statistical power, and reproducibility checks that were successfully carried out: Gärtner et al., 2025; Leising et al., 2022b, 2022a; Schönbrodt et al., 2025). Scores attained by individual articles may be aggregated for individual researchers (e.g., as part of promotion or hiring proceedings) or for entire institutions (e.g., as part of rankings).

(5.4.) Use more valid metrics in specifying strategic objectives (e.g., between professors, faculties, and university leadership). For example, a faculty may be financially incentivized for increasing its number of replication attempts (by X percent in a given time-span) – irrespective of the outcome of those attempts.

(5.5.) Abandon the pure *number* of completed doctorates as an evaluation criterion in all of the previously mentioned contexts.

(5.6.) Use more valid metrics as criteria in hiring and promotion procedures (see 10.3.).

(5.7.) Voluntary public commitment by researchers to a (relatively low) maximum number of authorships per average year (Frith, 2020). This would be a highly plausible way to prioritize quality over quantity, as it shifts the focus to one's most important projects, instead of one's being only superficially involved in as many projects as possible. This measure would also help reduce the overall number of manuscripts to be peer-reviewed, which should enable more thorough peer reviews in turn (see P9). As with some of the other recommendations we make, this one will be most effective if adopted by many researchers at once.

Problem Area 6. Overburdening of Professors with Tasks (P6)

Current expectations regarding German professors' performance are considered by many to be far too high. A study by Weihs et al. (2018) yielded an estimate of 56 hours average working time per week [95 percent CI: 35, 80]. In another study by Jackenroll and Julmi (2018), a substantial percentage of German professors reported feeling exhausted (15 % none to mild; 55 % mild to intermediate; 27 % intermediate to strong; 3 % strong to very strong).

As a general rule, university professors in Germany are expected to (a) teach eight to nine classes (typically including lectures and seminars) per year. This requires keeping teaching materials up to date (in both analog and digital format) based on scientific innovation and student feedback, as well as preparing and grading class papers and exams. All of this is accompanied by a lot of administrative work (e.g., archiving; handling of legal complaints). Most professors also supervise Bachelor's and Master's theses, which involves regular meetings with students, reviewing drafts, and reading and grading the final versions of theses.

The law requires German professors also to (b) contribute to their institutions' self-government. This implies participation in the meetings of various committees (e.g., faculty, senate, teaching committee, examination committee, hiring committee, ethics committee, open science committee, library committee, admissions committee, budget committee), as well as studying and commenting on the documents used in the work of these committees.

German professors are also (c) the responsible personnel and budget managers of their respective workgroup, including not only the bureaucratic duties that this entails (e.g., continuous monitoring of available funds, initiating and overseeing hiring procedures, reading and signing contracts, sometimes firing people), but also a significant responsibility for their employees' well-being and career progress (e.g., goal-setting, evaluation of goal achievement, giving feedback, help with networking, identifying and managing conflicts, writing reference letters).

Beyond meeting these "basic" requirements, professors are also expected to (d) continuously deliver research that is internationally competitive. This requires long-term, intensive content-related work at a high intellectual level and in constant coordination with changing teams of colleagues within and outside the respective institution. Typical research projects take several years from the beginning to completion: The literature must be studied, grant proposals must be written, submitted to funding bodies, and often revised. The same proposal must also be submitted to ethics committees.

The research that is going on in a professor's group must be supervised, discussed, and continuously optimized. Data must be collected, stored, analyzed and interpreted, and errors must be discovered and eventually corrected. Depending on the field one works in, data collection for a single project may take months or even years. Project funds (e.g., subject compensation, materials, wages) must be spent and correctly accounted for.

Furthermore, research results have to be presented at conferences and submitted to scientific journals in the form of manuscripts. The publishing process usually requires extensive (often multiple) rounds of manuscript reviews, at the same journal, or at different journals. Drafting and continuously optimizing a single, high-quality manuscript until it is ready for publication usually involves hundreds or even thousands of hours of work. The current development towards Open Science has increased this effort even more (e.g., by requiring pre-registration, open data, etc.).

In addition, professors regularly appear on the "other side" of the scientific process as well: as reviewers for manuscripts or grant proposals, and in the context of hiring procedures and award committees. Composing such reviews may easily consume entire work days. Yet another task is to make one's own scientific work visible (e.g., by having it registered in relevant databases, or via direct communication with the public using the available media channels).

Finally, professors are expected to be (e) good representatives of their respective institutions and of science as a whole. For example, they are to appear at social events such as alumni meetings or career days. They answer e-mails regarding all sorts of topics by students, colleagues, and journalists. And they volunteer to work as members of the governing bodies of science (e.g., for their professional societies).

The first to suffer from this overload of tasks are the professors themselves, many of whom are constantly working at the limits of their own capacity, often accompanied by the feeling that they have to cut back on quality in order to be able to meet all the demands at least to some extent. This contributes directly to the lack of quality in research, teaching, and supervision.

The overburdening of professors with tasks also contributes directly to the lack of thoroughness in peer reviews (see P9) and in the work of appointment committees (see P10). Oversight (see P12) will also be weak when professors have no time and energy left to pay attention to what their colleagues are doing, and when the responsible offices (e.g., for scientific integrity) are staffed with other professors who can only engage with this demanding type of work "on the side".

Causes / Contributing Factors

1. It has repeatedly been argued that academic institutions in Germany lack the basic funding that they would need to function properly (Institut für Demoskopie Allensbach, 2019; Vallbracht, 2024; Wissenschaftsrat, 2023). If this is true, then it is likely that too few people are expected to shoulder too much work.

2. The low level of formal flexibility in the distribution of tasks (e.g., the teaching load is typically the same for all professors) results in an inefficient allocation of resources.

Recommendations

The main activities of professors should be in the areas of research and teaching. Generally speaking, the commission recommends that professors be relieved of some of the more extraneous tasks they are currently burdened with, to enable them to concentrate more on these core responsibilities of theirs.

(6.1.) Increase the number of specialized positions that relieve professors of delegable tasks (examples: teaching coordinator, open science coordinator, exam coordinator, lab manager, data steward). The creation of such positions also prevents the coercion of ECRs to perform such tasks "on the side".

(6.2.) Allow for more flexibility with regard to the relative importance of the different tasks (research, teaching, administration, "third mission"), both between professorships and within professorships over time. Most notably, this includes making the teaching load more flexible, which may require changes to the respective state law.

(6.3.) Make it possible for professorships to be filled part-time (e.g., through job-sharing). This may require changes to state law. Halving the number of tasks assigned to a person this way may help reduce the felt workload and pressure overall, thus contributing to better well-being and better quality of the remaining work.

(6.4.) Make the overburdening of professorships better visible by publicly naming it as a problem, and through systematic research on the topic (e.g., by surveying professors regarding their workload).

(6.5.) Increase the basic funding of higher-education institutions in Germany, to enable adequate staffing (see also 6.1.). This may be achieved at no additional cost by repurposing some of the money that would otherwise be used for competitive grants.

Problem Area 7. Fixed-Term and Short-Term Employment (P7)

Recent studies show that 81 percent of academic staff in Germany are employed on a fixed-term basis (in 2013, the figure was 83 percent). According to a recent evaluation, about one third of these employment contracts have a duration of less than twelve months. In a survey of around 6,000 employees holding such a position, every second person stated that they considered the duration of their contract to be too short to achieve the desired qualification goal (Sommer et al., 2022).

The commission sees three main disadvantages of the current situation (see also Rahal et al., 2023): (a) The quality of academic work is likely to suffer if employees must permanently concern themselves with their own professional future (and write job applications, for example) in addition to their actual scientific tasks. Slim time perspectives may come into conflict with the basic requirement for diligence in research (e.g., the necessity to check for errors, attempt replications, and so on). (b) The constant dependence of ECRs on renewals of their fixed-term contracts makes them vulnerable to exploitation attempts by superiors (Scherpenberg et al., 2021). (c) Precarious and unstable employment conditions, even after attaining a doctorate, make it comparatively unattractive to remain in German academia (both when compared to academic employment in other countries, and to regular employment outside of academia). For example, the lack of compatibility between such employment conditions (which require high levels of regional mobility) and family planning is obvious (Kuhnt et al., 2022). This may lead women in particular to quit academia after completing their doctorate (Lemser, 2019). In a recent survey, 72 percent of professors agreed that career prospects for ECRs are “too uncertain” (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), 2024).

Causes / Contributing Factors

1. The lack of non-competitive basic funding limits the institutions’ ability to plan ahead and virtually forces them to retain as much budget flexibility as possible through temporary appointments.

2. In 2007, the German federal government introduced a new law (i.e., the German Act on Temporary Scientific Contracts; WissZVG; see Sommer et al., 2022) with the original intention of limiting fixed-term employment. To achieve that goal, this law prohibits further fixed-term employment for persons who have worked in academia for 12 years or more. However, the effect was pretty much the opposite of what had been originally intended: instead of being given permanent positions, ECRs are now regularly unable to find further employment once they have reached the 12-year limit.

3. Given that professors in Germany currently have the authority to hire “their own” ECRs, university administrations attempt to make their available professorships attractive by keeping all the ECR positions associated with them easy to vacate.

4. Occasionally, it is argued that permanent employment would reduce the productivity of scientists (due to lower pressure). To the best of our knowledge, however, there is no empirical evidence supporting this claim, and the same argument could also be used against the permanent appointment of professors.

Recommendations

(7.1.) Use of US-style tenure-track positions as the standard path toward a tenured professorship. This may be accomplished using an approach in which an ever-increasing number of professors is appointed this way (DGPs Board of Directors, 2022). The main goal in

this would be to provide ECRs with more predictability in planning their careers. A positive evaluation should lead to a permanent position, either with or without the title of professor.

(7.2.) Increase the number of permanent positions with full or even increased teaching load. The holders of these permanent positions may also contribute to relieving professors from some of their more peripheral tasks (see 6.1.).

(7.3.) Make 4 years the regular contract duration for PhD positions and/or take measures to shorten the typical time-span needed for completing a PhD to 3 years. The latter should become easier when also implementing recommendations 7.4. (relieving PhD students from data-collection work) and 8.6 (making a single manuscript the basis for a PhD).

(7.4.) Pure data collection for research projects can and should primarily be carried out by research assistants (RA), not by PhD students. This would relieve the latter of a task for which they are usually overqualified and give them time to concentrate on scientific tasks that they cannot delegate (e.g., reading, designing studies, analyzing data, attending conferences, writing manuscripts). It would also make the people who carry out the data collection less vulnerable to abuses of power by principal investigators, since the dependencies associated with PhD completion may not be used as leverage against them.

Problem Area 8. Unnecessarily Strong Power Imbalance (P8)

Depending on the position, jobs in academia offer varying degrees of opportunity to exercise power over others. The power gap between those who hold a professorship and most other people in the system (especially ECRs and students) is particularly large (Scherpenberg et al., 2021): A German professorship entails the power to make decisions over resources (e.g., laboratory time, research infrastructure such as computers, congress fees, travel expenses), to evaluate the work of others (e.g., grading of theses), and to hire people or discontinue their contracts. With a majority of votes in hiring and promotion committees (see P10), professors also decide who else will be appointed to a professorship. As reviewers and editors for scientific journals, they decide whose work is published where (see P9), and who will be awarded a research grant. As teachers, professors basically have complete autonomy in determining the content (including difficulty level) of their classes and exams.

In addition, more informal sources of power play an important role. For example, given their expertise and standing in the scientific community, professors are able freely to decide whether to help push the careers of junior scientists by (e.g.) collaborating with them, introducing them to influential colleagues (i.e., other potential collaborators), or recommending them for leadership positions (e.g., on editorial boards).

Unfortunately, the existing power imbalance exposes those who do not hold a professorship to potential power abuse (see P2), and research suggests that different forms of power abuse may actually be fairly common in German academia (e.g., Scherpenberg et al., 2021). For example, the power associated with a professorship may be used to coerce employees into applying questionable scientific practices (QRPs) (see P1). Speaking more generally, the integrity of the scientific discourse itself is likely to suffer when discussions between professors and ECRs take place within relationships characterized by so much dependency of the latter on the former (Kuhnt et al., 2022). In a recent survey, about 40 percent of German professors agreed with the statement that the existing hierarchies have a detrimental effect on the autonomy of ECRs (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), 2024).

Causes / Contributing Factors

1. The unnecessarily steep power imbalance between professors and most other people in the academic system results primarily from the accumulation of various functions in the hands of the former. Particularly problematic for ECRs (probably the most vulnerable group) may be the fact that their closest scientific collaborators are also their immediate superiors, combined with the common practice of short-term employment (P7). There is a very concrete risk that the need to have one's contract renewed repeatedly may be used as leverage against ECRs (Kuhnt et al., 2022).

2. Another possible source of leverage against ECRs is rooted in the alleged or actual power of professors to give or deny access to the research data that are collected as part of a project. For most ECRs, having access to these data is an indispensable prerequisite for being able to work on their research questions at all. However, the legal situation in this area ("who owns the data?") is largely unclear to most of those involved. Most often, it is simply assumed that the data are owned by the head of the respective division (i.e., the professor).

3. Moreover, most German professors are public servants ("Beamte"), which gives them an extraordinary level of job security (e.g., it is comparatively

difficult to fire them; e.g., Löwe and Valet, 2023), and their relatively fair salaries give them the advantage of being able to afford good legal counsel in cases of conflict – a privilege not enjoyed by most ECRs.

Recommendations

(8.1.) Disentangle the roles as scientific collaborator and superior of ECRs, both of which are currently held by professors, and instead assign personnel responsibility to the department as a whole. This means that professorships will no longer have ECR positions formally assigned to them (Specht et al., 2017). Rather, let junior and senior scientists agree on working together rather informally, and make it clear that such arrangements may be modified in the course of a junior scientist's association with the department (see 3.4.). In order to prevent an undue concentration of power in the dean's office, personnel decisions (e.g., over premature contract termination) should be made jointly by a standing committee with a regularly changing composition.

(8.2.) Make exemplary solutions in this regard visible (e.g., by reporting on them on the DGPs website), as a model for other faculties.

(8.3.) Authorship, access to data, and access to financial and material research resources (e.g., laboratory hours, travel funds) should be explicitly and bindingly regulated at the beginning of any scientific cooperation. For ECRs, this should be made part of a supervision agreement (see 3.4.).

(8.4.) Delegate (at least part of) the coaching function currently associated with professorships to persons who are particularly qualified for this (e.g., specialized counselors at grad schools). These counselors also need a budget for financing training workshops for ECRs.

(8.5.) Make an external assessment of all scientific work the rule: Persons who were themselves involved in the preparation of some piece of scientific output (such as a manuscript) should not be involved in the assessment of that work. The same goes for all of their close collaborators.

(8.6.) A single manuscript that could be submitted to a scientific journal for review should be sufficient as the basis for a doctorate. This rule takes into account the significantly higher demands of research in times of Open Science, as well as the general goal of giving more weight to the quality (as opposed to the quantity) of research.

(8.7.) This manuscript should be reviewed externally (i.e., by persons not part of the same research team or faculty) and independently of any peer review by a scientific journal that may also take place. Otherwise ECRs might be forced to endure long review processes for which they themselves are not in any way responsible. External (preferably international) reviewers may be selected by an independent body (e.g., thesis advisory committee). This approach implies that the completion of a doctorate is explicitly decoupled from any publication success.

(8.8.) Abolish the grading of doctoral theses. Evaluations should instead be made in terms of these four categories: "passed", "significant formal revision necessary", "significant revision of content necessary", "failed".

(8.9.) Consider a more equal representation of the four major academic status groups (professors, ECRs, students, administration) in the self-governing bodies of academic institutions. This is likely to require changes to state law.

(8.10.) Provide ECRs with budgets for congress participation and travel that they may use at their own discretion.

(8.11.) Establish and provide sustainable financing for a position that offers independent legal advice for ECRs at no cost to them (up to a certain limit).

(8.12.) Develop and publish a legally sound guideline on the topic of "Rights to Research Data". This is a task for the DGPs (and/or DFG) leadership. The commission recommends taking into account the relevant experiences documented by the German Research Ombudsman (Czesnick, 2020).

Problem Area 9. Ineffective Peer-Review System (P9)

Peer review (i.e., the assessment of research ideas and research results by colleagues) is supposed to function as the primary quality-assurance mechanism of the academic world. It is necessary because competently assessing some scientific achievement usually requires a high level of expertise, which only other researchers working in the same field have to a sufficient extent. The commission assumes that peer review must and will continue to play the decisive role in making such assessments in the future.

However, it has become clear that peer review often does not properly fulfill its quality-assurance function (Rennie, 2016; Tennant & Ross-Hellauer, 2020; Vazire & Holcombe, 2021). This applies both to reviews of scientific manuscripts and to reviews of grant proposals. The following concerns are commonly voiced in that regard:

(a) Peer review is hardly able to detect actual fraud (see P1).

(b) The peer-review system is susceptible to nepotism and (positive as well as negative) reciprocity effects. This is because academic fields are often small (people know who is working on what), roles of authors and reviewers change continuously in both directions, and

the supposed anonymity of authors and reviewers is often questionable, or even actively broken (“guess who reviewed your grant proposal recently?”), without this ever becoming transparent or being regulated.

(c) Peer review may be strongly influenced by political and strategic considerations (e.g., attempts to promote or impede a particular research subject, or a particular "school of thought” on a given subject), with this being hidden behind pseudo-scientific arguments.

(d) Some reviewers and editors use their positions of power in the peer-review process to pressure authors into citing them, or to steal others’ ideas, or to prevent/delay publications by others that could (supposedly) stand in the way of their own academic success (e.g., because the reviewed work challenges the reviewers’/editors’ own work).

(e) Reviews are sometimes negligent and superficial, which is a problem given how much people’s career prospects depend on publication success. For example, actual checks of reproducibility (i.e., running the associated code on the associated data and comparing the results to those reported in an article) are still not conducted at most journals.

(f) Some reviewers use inappropriately condescending, hostile, or hurtful language.

(g) The system is inefficient because articles rejected by one journal keep getting submitted to other journals until they are eventually published somewhere.

(h) The fact that so many findings cannot be replicated shows that the current peer-review system is not a sufficient gatekeeping mechanism against the publication of low-quality research.

The main victims of a dysfunctional peer-review system are those researchers who value and produce high-quality research, as they cannot be sure that their efforts will be appreciated and rewarded. Good publication success may be achieved with easily-produced and methodologically inferior research if non-scientific (e.g., personal, political) influences play a decisive role in the publication process. This may then translate into unfair disadvantages in hiring procedures (see P10). Other researchers working in the same field may also be negatively affected if they must build their own research on published studies of dubious quality.

Causes / Contributing Factors

1. The number of items (i.e., grant proposals and journal manuscripts) to be reviewed is constantly very high, which has led to a situation in which it becomes more and more difficult to find qualified reviewers (so-called "reviewer fatigue"). One of the main reasons for this high number of submissions is to be found in misaligned incentives (see P5): Researchers are primarily rewarded for publishing as much as possible as quickly as possible, and for raising and spending as much third-party funding as possible.

2. Most reviewers are professors, who tend to be overburdened with tasks (see P6). The constant distraction by numerous other tasks makes it less likely that they are willing and able to invest the necessary diligence into their reviewing work.

3. There are hardly any extrinsic incentives for taking one’s reviewing work seriously and investing the time and diligence that would be needed.

4. The primary interest of commercial publishers is to generate revenue from selling articles. More thorough reviews would prevent many articles from ever being published, and they would slow down the publication process for the remaining ones. There is a clear conflict of interest here between the publishers’ commercial interest and quality assurance.

5. Due to a common lack of transparency on the side of the authors, reviewers have a hard time getting a complete and unadorned picture of the research, and thus have to rely on the validity of the information that the authors are willing to make available.

6. Traditionally, peer reviews are not made public. This leaves room for editors’ decisions to be “based on peer review” with little accountability for the quality of the respective information base (i.e., the reviews), for how carefully the soundness of the reviewers’ arguments was considered, and whether appropriate measures for preventing undue (e.g., political) influences on the review process were taken.

Recommendations

(9.1.) Public declaration by professional societies and research institutions that a significant reduction of the publication output in terms of quantity is needed.

(9.2.) Reduce the number of manuscripts under review by significantly increasing the minimal methodological requirements (e.g., mandatory pre-registration and replication) and desk-rejecting manuscripts in case of non-compliance with these requirements. In this, special study features such as particularly expensive / sophisticated designs or participant samples that are difficult to recruit must be taken into consideration.

(9.3.) Move to not-for-profit scientific publishing as the standard model, as quickly as possible. There

should be no commercial interests that may interfere with the original purposes of scientific publishing. Platforms affording not-for-profit publishing do already exist and continue to emerge (e.g., the *PsychOpen GOLD* program at ZPID; *Meta-Psychology* hosted at Linnaeus University in Sweden). Not-for-profit publishing should lead to a smaller overall publication volume and a better signal-to-noise ratio in the published literature. It thus bears significant potential for improving efficiency in the spending of public resources.

(9.4.) The Commission recommends that the transition to not-for-profit publishing be publicly declared a key strategic goal by professional societies and universities.

(9.5.) As a general rule, the contents of peer reviews should be made public. This would provide some urgently needed transparency in regard to the soundness of the arguments on whose grounds journal editors ultimately reject or accept a paper.

(9.6.) Get specialized experts involved who – in addition to the traditional content review – evaluate submitted manuscripts for their more technical aspects (e.g., compliance with pre-registrations; reproducibility of results using open data and code).

(9.7.) Peer reviews should be signed by the reviewers, if possible. This would increase transparency and accountability with regard to possible influences of (hidden) networks and reciprocities, as well as reviewers' vested interests and biases for or against certain subjects or methods. In addition, such unblinding may help encourage reviewers to adopt a constructive tone. There may, however, still be situations in which anonymous peer review is appropriate.

(9.8.) Note, however, that unblinding may inadvertently strengthen the aforementioned reciprocity effects: Authors may wish to "pay back" reviewers whose identities they know, when the roles of author and reviewer are reversed some time in the future (Lynam et al., 2019). A further risk may lie in a generally reduced willingness of reviewers to express appropriate criticism publicly. These risks must therefore be kept in mind and monitored. This monitoring will become easier when the content of reviews is routinely published (see 9.5.).

(9.9.) Peer reviews should be recognized as a publication format in its own right, receive DOIs, and thus become citable. This should help increase the motivation of reviewers to be thorough in their work. The commission proposes that journals maintain (or link to) a database containing all submitted manuscripts as preprints, as well as the citable reviews that were written about each submission.

Note, however, that this approach does entail some potential for power abuse, because editors may prefer to invite certain reviewers and thus provide them with more opportunities for getting their reviews published (and cited). To prevent this, (a) it should be made transparent who has invited whom to be a reviewer for a manuscript, (b) such decisions should not be made by a single person, and (c) the number of such invitations to the same reviewer should be limited (e.g., to 3 per year and journal). Note that DOI allocation is also conceivable for anonymous reviews.

(9.10.) Promote registered reports as the gold standard in publishing empirical studies. This publishing model automatically entails pre-registration, helps fight publication bias, and makes the expertise of reviewers available at earlier stages of the research process, so it may actually contribute to better study planning. Submission of registered reports could be incentivized by offering fast-tracking for them.

Problem Area 10. Questionable Assessment Practices in Hiring Professors (P10)

For years, a need for improvement has been articulated with regard to selection procedures in hiring personnel for leadership positions in German academia (e.g., Abele-Brehm and Bühner, 2016a, 2016b). Common complaints include a lack of due diligence, the use of data (e.g., metrics) of questionable validity, failure to capture highly relevant personality characteristics (e.g., integrity, leadership skills), and an overly strong influence of local micropolitics. All of this may result in suboptimal decisions that have a variety of negative effects: (a) Students may be affected by poor teaching (see P4), (b) scientific personnel may be affected by problematic leadership behavior, both on the professional level (e.g., demanding the use of QRPs) and on the interpersonal level (e.g., bullying), (c) to the extent that unethical leaders give ECRs who tolerate such behavior preferential treatment (e.g., by furthering their careers), questionable scientific and interpersonal values may be passed on to the next generation and multiplied in their problematic effects, (d) appointing scientific leaders with questionable integrity is likely to lead to poor integrity of the scientific output, to the detriment of other researchers and the general public (see P1), (e) the appointment of persons with low integrity to academic leadership positions creates a considerable reputational risk for the institutions concerned ("scandals").

Causes / Contributing Factors

1. Validly assessing candidates for professorships not only requires (a) basic diagnostic competencies (e.g., knowledge of concepts such as reliability,

predictive and construct validity, as well as typical errors of judgment) and (b) specific expertise for judging the quality of an applicant's scientific contributions (e.g., publications) (instead of just counting them, for example), but also (c) the time and willingness to use these competencies conscientiously and thoroughly. However, time is something that German professors are notoriously short of, due to the vast range of tasks that they are supposed to handle simultaneously (see P6). The potential long-term benefits that may be incurred by doing diligent committee work (e.g., more productive collaborations with new colleagues, fewer scandals) may thus often be trumped by more immediate concerns over where best to allocate one's time and energy in the next few weeks or months.

2. As an alternative to a thorough qualitative examination of the research performance of individual applicants, the use of mainly quantitative metrics (e.g., number of publications and citations, third-party funding, h-indices, impact factors) is still very common. These metrics not only appear to be comparatively objective, but are also very easy to determine. Their use may thus save committee members time and effort, which makes it very attractive. However, it has become more and more obvious how easily these metrics may be manipulated, and how readily some researchers do manipulate them (e.g., Chapman et al., 2019). For example, guest and "honorary" authorships continue to be among the most common forms of scientific misconduct (Fong & Wilhite, 2017; Pruschak & Hopp, 2022), despite the fact that these practices have been denounced as unethical for decades. The lengths of publication lists and the impact factors of the journals one publishes in continue to play important roles in selecting academic leadership personnel. Taken together, this may lead to a selection of the most ruthless individuals for professorships (i.e., the ones who have the least objections against maximizing their own numeric indices of research productivity at any cost).

3. With regard to the assessment of teaching competence, it must be asked whether the typical 15-minute live-teaching sample is a valid indicator of those characteristics that one is looking for in a professor (e.g., the ability to explain difficult matters in ways that foster interest and understanding in students). The use of course evaluations by students as indicators of teaching quality is similarly questionable, as it is known that such data strongly reflect relatively extraneous factors such as affection for the teacher and the difficulty of the subject (Kreitzer & Sweet-Cushman, 2021; Stroebe, 2020).

4. As pointed out above, the commission assumes that problems such as abuse of power (see P2) and scientific misconduct (see P1) result from interactions between environmental conditions and the personality dispositions of certain individuals. A thorough assessment of such dispositions should thus be part of any hiring procedure for permanent positions in academia. However, at present there is still a gross disparity between their obvious relevance (Forster & Lund, 2018; Peus et al., 2015) and the rather marginal role that personality dispositions typically play in assessing candidates. More generally speaking, there is too little clarity regarding standards of good leadership behavior in German academia.

5. The Commission is not aware of any systematic studies addressing the influence of local micropolitics on hiring procedures in academia. However, a large body of anecdotal evidence suggests that the personal interests of current faculty members (e.g., fear of competition or critical scrutiny of one's own work) do often play a rather decisive role in the selection of candidates.

Recommendations

(10.1.) Members of hiring committees need diagnostic expertise. Concepts such as reliability and validity, typical judgment errors and their causes, as well as ways to avoid them are familiar to most trained psychologists. However, most hiring committees include many members who lack such expertise. It thus seems advisable to train those people for the task.

(10.2.) Members of hiring committees need time and energy for a thorough review of applications. Therefore, their efforts must be compensated (e.g., by relieving them of some of their teaching duties, or by granting them additional personnel and material resources for their research). This likewise applies to members of committees making tenure decisions or decisions on the hiring of staff for non-professorial but permanent positions.

(10.3.) Explicitly use Good Scientific Practice criteria when evaluating the scientific output and potential of applicants (Gärtner et al., 2025; Leising et al., 2022b, 2022a; Schönbrodt et al., 2025).

(10.4.) Develop and use more valid indicators of teaching skills. For shortlisted candidates, live or

recorded teaching samples could be assessed by experts with regard to correctness of content, didactics, and appropriateness for the intended audience, as well as the applicants' ability to engage with questions constructively, and to promote debate and knowledge transfer. Comprehension tests for participants (pre-post) would enable a more objective assessment of learning progress than is possible, for example, with global ratings of student satisfaction.

(10.5.) Relevant personality traits and competencies (e.g., ethical leadership, mentoring, and conflict management) must be explicitly recognized in their importance and be accounted for by assessment procedures (Abele-Brehm & Bühner, 2016b). Mere self-assessments of these characteristics must be assumed to have very limited utility, due to their strong social desirability. A personality variable of utmost importance is personal integrity, which needs to be distinguished from agreeableness, because integrity is most clearly demonstrated in instances where a person openly dares to object to unethical behavior even if it is shown by almost everybody else. More research into viable ways of validly assessing this domain is needed. Based on a review of the available literature, the commission recommends looking more closely into the following methods: assessment centers (Müller, 2015); structured interviews including biographical elements and critical scenarios (Strobel et al., 2018); Situational Judgment Tests (Corstjens et al., 2017; Kasten et al., 2020); integrity tests (Marcus et al., 2007; Ones & Viswesvaran, 2001; Ones et al., 1993, 2012; Van Iddekinge et al., 2012). The level of candidates' voluntary transparency about their own research processes (as demonstrated by their using Open Science practices) may function as a proxy measure of integrity in this domain. The RESQUE Collector App (Gärtner et al., 2025) may be used to assess this.

(10.6.) Significantly increase the proportion of external members in hiring committees, and have them selected in ways that ensure their independence (of the local faculty, and each other). For example, external members of hiring committees may be selected by a group that is appointed for that purpose by a professional association. This approach may not only help limit unwanted influences of local micro-politics, but also add some relevant (e.g., diagnostic) expertise.

Problem Area 11. Lack of Clarity of, and Low Commitment to, Ethical Standards (P11)

The existence of ethical norms is a necessary prerequisite for being able to judge people's actions as more or less morally acceptable. It is therefore central to all twelve problem areas that are addressed in the current paper. To establish an ethical organizational climate and culture, however, not only do norms have to exist, but members of the organization need to be aware of them, take them seriously, and identify with them.

In recent years, the concept of an organization's "ethical climate" has been applied to academic institutions. Work climate is defined as organizational norms and conventions that are seen by an organization's members as existing within its structures and procedures (Martin & Cullen, 2006). It is assumed that an organization's climate is amenable to planned change (e.g., Viđak et al., 2021). Richardson and Healy (2019) point out the relevance of an overarching framework for promoting an ethical climate at institutions of higher education, and the leadership's responsibility to assess the current ethical climate, in order to understand its impact on the staff's behavior.

Empirical studies suggest that many who work in science perceive the gap between theory and practice in that regard as being quite large (Chapman et al., 2019). For example, although most scientists are probably aware of the difference between acceptable and unacceptable research practices, use of the latter seems to be fairly common (see P1). When members of an organization perceive its alleged ethical principles as merely a shallow facade (because they keep getting violated by the actual behavior of the organization's members), this may breed cynicism and alienation (Mete, 2013). With regard to the use vs. abuse of power (see P2), even the relevant norms themselves seem not to be entirely clear. The commission's extensive discussions with various actors in the system revealed that many of them seem to be unsure about why certain actors are given certain powers, and where the line should be drawn between legitimate uses of those powers and illegitimate abuses.

There also seems to be little knowledge regarding forms of destructive leadership that may be fueled by a leader's narcissistic or psychopathic personality characteristics (see P10), and most employees know very little about their own rights under labor law. Such a lack of clarity may be conducive to misconduct. The commission thus recognizes a need for raising awareness in these regards.

Causes / Contributing Factors

1. A more consistent implementation of Good Scientific Practice is severely impeded by the current incentive system, which rewards quantity over quality (see P5). Injunctive norms (i.e., officially declared values and guidelines) often stand in direct opposition to descriptive norms (i.e., actual behavior).

2. In academia, issues of power and power abuse

in the workplace have only recently begun to be taken seriously. The commission suspects that the previous neglect of these issues was rooted in a widespread, hyper-competitive mindset that prioritized (the appearance of) productivity and achievement over almost all other aspects of work (e.g., the well-being of employees). It may also be a reflection of more general avoidance tendencies: Most people – within and outside of academia – are reluctant actively to confront the question of how well they are living up to (their own or others') moral standards, and to deal with such unpleasant topics as manipulation, deception, exploitation, corruption, and boundary violations.

Recommendations

(11.1.) Offer (continuous) training regarding topics with ethical relevance: power and abuse of power; Good Scientific Practice and Open Science; bullying and ostracism; sexual harassment; conformity and dissent; psychopathy and narcissism; leadership; corruption, including the applicable laws; dealing with and learning from mistakes; labor laws; quantity versus quality in research; whistleblowing and complaint mechanisms (e.g., Kohrs et al., 2023). Make participation in such training obligatory, if possible.

(11.2.) Such training should have an applied, psycho-educational character. Working on emotionally engaging case examples may be a particularly suitable method for this (Katsarov et al., 2022). The primary goal should be to make such topics discussable in the first place, to enable participants to recognize when ethical questions are involved, and to make more responsible decisions in that regard. Useful recommendations regarding course design may be found in van den Hoven et al. (2023) and in Prieß-Buchheit et al. (2020).

(11.3.) The target audience for such training should include students, professors, ECRs, staff, and university leadership – in other words, everyone who is somehow involved in research or teaching activities. A study covering three countries (Netherlands, Spain, Croatia) by Evans et al. (2022) shows the importance of an institution-wide approach.

(11.4.) For efficiency, the commission recommends that academic institutions work together in developing a shared curriculum for such this type of training.

(11.5.) Another possible measure would be the training of multipliers who are explicitly appointed as contact persons regarding ethical questions and problems. They might be tasked with proactively promoting the importance of ethical awareness and commitment (e.g., by giving short refresher talks on regular bases). Some tested material can be found online at https://www.path2integrity.eu/ and https://sops4ri.eu/.

Problem Area 12. Weak Control and Sanctioning Mechanisms (P12)

In many areas of the private sector (e.g., banking, mobility, pharmaceutical research), it has become accepted that the actions of organizations and the individuals in them have to be externally monitored in some way, especially if these actions have the potential adversely to affect public goods. In many fields of science (including psychology), the situation to date is very different: Most of the responsibility for supervision, controls, and possible sanctions lies with the academic institutions themselves. However, many case reports in the media as well as survey studies show that the current complaint systems are not very effective (Elson et al., 2021; Höbel et al., 2022; Scherpenberg et al., 2021; Schraudner et al., 2019). Potential complainants do not trust them, do not use them, and unethical conduct hardly ever has severe consequences for perpetrators.

Scientific misconduct harms the scientific community and the general public. Other forms of misconduct (exploitation, bullying, sexual harassment) harm the respective victims' physical and mental health. Moreover, in the rare cases where such misconduct is actually uncovered, it often causes immense reputational damage to those directly responsible and to the institutions they work for. This damage is all the greater if it also becomes apparent that others in the institution knew about the misconduct for a long time, did nothing to stop it, or even attempted to help cover it up.

Causes / Contributing Factors

Several different factors contribute to the current situation in which control and sanctioning mechanisms in academia fail to perform their functions adequately:

1. In contrast to many universities' public commitments to ethical research practices, the easiest path for them to create an appearance of research excellence is by rewarding researchers who help optimize the common metrics (see P5), while not asking too many questions regarding the means by which this is achieved.

2. Potential whistleblowers or complainants are not familiar with the respective procedures, or the relevant procedures and responsibilities have not even been sufficiently specified.

3. Potential whistleblowers or complainants lack confidence in the existing complaint channels.

This may be due to doubts about (a) the confidentiality and/or (b) the effectiveness of the designated procedures, and/or (c) the impartiality of the responsible office-holders. All three types of doubts seem to be justified, as media reports and personal testimonies to members of the commission suggest.

4. The investigative and sanctioning powers of the responsible offices and of the universities' leadership boards are very limited. For example, ombudspersons working for different academic institutions are not legally allowed to share their knowledge of individual cases with one another, and the most serious sanction that the German Research Foundation has for violations of its ethics code is a five-year ban on submitting grant proposals.

5. The responsible offices within German academic institutions tend to be poorly staffed. Ombudspersons, for example, are often academics who do the job on top of their everyday duties (see P6), who receive no (e.g., financial) compensation or even incentives for doing the job thoroughly, and who often lack the legal and psychological expertise necessary for doing it well. In addition, turnover is high, which makes it even more difficult to handle the often complex and lengthy cases properly.

6. If the accused are professors, the power imbalance between them and most potential complainants (see P8) tends to induce intense fear (e.g., of retaliation) in the latter. For example, complainants may be afraid that a professor could take revenge on them by attempting to ruin their career. Under German law, most professors are state employees and thus almost unfireable, whereas almost all ECRs work on fixed-term contracts with relatively short durations (see P7).

7. The effectiveness of internal complaint mechanisms may be severely diminished by conflicts of interest that arise from the existence of personal and/or professional relationships between the accused and (a) the officers responsible for dealing with complaints, and/or (b) others in the same institution who know of misconduct and might file a complaint. Often, taking reported problems seriously could lead to greater damage to relationships and reputations than attempting to conceal them.

8. The fact that so few cases of misconduct are reported is sometimes misinterpreted by decision-makers as an indication of a low prevalence of problems. There is a feedback loop in operation here, in which the factual weakness of oversight (with few complaints being made and even fewer sanctions imposed) is used as an argument against the need for oversight in the first place.

9. Lack of awareness and commitment regarding ethical standards (see P11) is also likely to contribute to the lack of effective oversight in academia. This is because (a) potential whistleblowers may simply not have a clear enough understanding of what is and what is not acceptable, and (b) members of the institution may decide to abstain from flagging unethical behavior that they see being tolerated all the time.

10. Finally, university professors enjoy a very high reputation with the German public (e.g., as intelligent, meticulous and responsible people whose primary aim is to benefit the public). Such views may make the necessity of oversight of these people's activities seem less than obvious for laypersons.

    Given all of these challenges – entirely to the detriment of potential whistleblowers – it is not surprising that successful complaints seem to be a rare exception rather than the rule in the science system (Elson et al., 2021; Höbel et al., 2022), and that many cases of even severe misconduct only come to light via the press.

Recommendations

(12.1.) Specify all reporting procedures and publish them on the same easy-to-find websites. It must be made clear which office is responsible for which type of complaint (scientific misconduct, bullying, corruption, sexual harassment, interpersonal conflict, supervision problems), how the relevant processes work, how the confidential handling of information (e.g., about the identity of persons providing information) is ensured, and what further protection measures are available to complainants.

(12.2.) Strengthen the responsible offices significantly, through adequate staffing, full-time payment, and granting of proper investigative powers (e.g., the right to inspect and – if necessary – secure the files and data-sets from research projects). The commission also considers the clarification of the legal situation in this area to be an important short-term goal.

(12.3.) Establish independent ethics offices at the state or federal level. The most important issue here is to decouple these offices from individual academic institutions, to reduce the aforementioned conflicts of interest. Such offices may be financed di-

rectly by the state, or by an association of academic institutions. The same requirements regarding resources and powers mentioned in point 12.2. apply here, as well. The commission has issued a separate document addressing these topics in more detail, incorporating advice from several members of the Network Against Power Abuse in Science e.V. (https://www.dgps.de/fileadmin/user_upload/PDF/Stellungnahmen/AMWF_Stellungnahme_Beschwerdeverfahren.pdf).

(12.4.) As a rule, get an independent office involved at the *beginning* of every complaint procedure. The decision as to whether and how a complaint should be followed up inside or outside the concerned institution must not rest with persons who are employed by that institution.

(12.5.) Staff ethics offices with several, rather than just one, persons at the executive level. This will establish some desirable redundancy and help mitigate potential biases and conflicts of interest.

(12.6.) Conduct regular proactive inquiries of employees. These may be anonymous surveys or confidential face-to-face interviews focusing on opportunities for improvement in organizational structures and processes. Approaching employees proactively should help lower the threshold for reporting problems. In this, aiming for representativeness is key to avoid interpretability problems due to possible (non-)participation bias.

(12.7.) Test the viability and effectiveness of *anonymous* reporting of misconduct. Note, however, that German law requires complainants ultimately to come forward and identify themselves if their complaint is supposed to lead to an investigation. As the existing power imbalance (see P8) and the associated fear of retaliation may ultimately disable any reporting channel (anonymous or not), flatter hierarchies will still be needed.

(12.8.) Conduct evaluation research on the effectiveness of existing complaint procedures.

(12.9.) An ethics office that may function properly requires employees with legal, psychological, and administrative expertise, as well as some knowledge of the internal workings of the science system. This suggests it will be reasonable to fill these positions with people who have the necessary professional backgrounds (e.g., lawyers), and that a single position-holder will be unlikely to have all the necessary expertise.

Note that, since the first version of the current report was prepared, the new Whistleblower Protection Act (HinSchG) has come into force in Germany. It covers some of the same ground as the commission's recommendations in this chapter.

How Specific are these Problems for (German) Academia, and for Psychology?

We will conclude this article by briefly addressing the generality vs. specificity of our problem descriptions and recommendations for (a) academia, (b) Germany, and (c) the field of psychology. In this, we will not aim for comprehensiveness, but rather highlight what are some of the most prominent issues from our perspective.

First, it has been said above that many forms of unethical conduct are prevalent both within and outside of academia. This includes bullying, as well as sexual harassment and exploitation. Dhanani et al. (2021) reported that they could not find systematic differences in prevalence estimates between representative samples and convenience samples. This would mean that common prevalence estimates (approx. 20-40 percent) may be taken seriously. We are not aware of any studies comparing prevalence estimates between academia and other types of work environment. Therefore, the question of whether unethical behavior is more or less common in academia than outside of it has to remain unanswered for now. We do believe, however, that our analyses above show that the current academic system encourages and rewards unethical behavior to a certain degree.

Second, the German academic system does have some characteristics that are quite distinctive in international comparison. Among these is the fact that a typical German professor is not only the closest scientific collaborator of "her" PhD students and postdocs, but also their boss, and also involved in evaluating their scientific work. Compared to most other Western countries, this accumulation of power in the hands of a single individual is rather extreme (see P8).

In addition, Germany has a special law (the WissZeitVG) in place that legalizes giving ECRs a series of relatively short-term contracts (see P7). If an employer (typically a professor) is not satisfied with an ECR anymore, this law makes it legal effectively to terminate the ECR's employment by simply not renewing their contract. This possibility adds significantly to the existing power differential between senior and junior academics. The resulting leverage may be used by the former to exert all sorts of pressure on the latter (e.g., to work more, to work on tasks unrelated to their job, to somehow deliver seemingly good results, etc.).

The same dependencies that make the occurrence of these and other forms of unethical behavior more likely also make it less likely that any victims or witnesses will ever file complaints, due to fear of retaliation or being ostracized. The pervasive silence and fear among people working in the German academic sec-

tor have been repeatedly documented by investigative journalists (Illinger, 2019; Neumann, 2019; Wiarda, 2023). Furthermore, most German professors are employed by the state (Bundesland) in which they work, providing them high job security and strong legal protections against being laid off or sanctioned otherwise. All of this adds to the aforementioned power differential even more. The overall situation of German professors has been described as "an invitation to power abuse" (Winkler, 2023).

By international comparison, German professors have a relatively high teaching load. Most German universities are publicly funded and do not collect tuition fees. Even the most unfavorable teaching evaluations hardly ever have tangible consequences for a tenured professor. These factors combined may account for a relative neglect of teaching quality at German universities. We are not aware of any systematic investigations into these issues, however.

Third, above we highlighted some deficiencies of standard practice in psychological research, such as the use of underspecified theories and a general lack of coordination (e.g., regarding terminology and measurement practices) (see P1). These deficiencies continue to harm the credibility of the entire field. In fact, a recent German survey shows that researchers' trust in the solidity of the work that is being done in their own field is particularly low among psychologists (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), 2024). Given that at least some other fields have replicability rates that are not higher than those in psychology (e.g., Errington et al., 2021), we lean towards interpreting this as a sign of healthy realism rather than negativism.

To end on a more positive note, psychologists' growing awareness of these and other weaknesses of their own scientific work has undeniably led to a number of very fruitful initiatives for improvement (e.g., pre-registration; replication databases; theory-formalization efforts; large-scale data-collection and data-analysis collaborations; Diamond Open Access journals; transparent peer review). There can be little doubt that all of these are well-suited for improving scientific practice and for yielding more trustworthy results in more efficient ways. Even the greatest skeptic would have a hard time making the point that improvement is nowhere in sight. It is.

Author Contact

Corresponding author is Daniel Leising, TU Dresden (daniel.leising@tu-dresden.de).

ORCIDs

Nicole Bössel-Debbert 0000-0002-2578-6447 Annette Kluge 0000-0002-8123-0427 Daniel Leising 0000-0001-8503-5840 Dorothee Mischkowski 0000-0002-7563-402X Le Vy Phan 0000-0002-6470-8871 Franz Richter 0009-0004-6862-2962 Manfred Schmitt 0000-0002-9978-8357 Jutta Stahl 0000-0002-3252-180X

Acknowledgments

The members of the commission would like to express their gratitude to the following members of the two previous DGPs leadership boards (2020-2022 and 2022-2024) for their trust and continued support:

Markus Bühner, Anne Gärtner, Eva-Lotte Brakemeier, Stefan Schulz-Hardt, Karl-Heinz Renner, Anna-Lena Schubert, Matthias Nückles, Ulrich Ansorge, Christian Fiebach, Jens Bölte, Christina Bermeitinger.

The members of the commission would also like to thank the following individuals for their help with compiling the content of the present report, through their comments on earlier versions, in-depth discussions, and reading recommendations:

Gerhard Blickle, Sabine Chai, Hjördis Czesnick, Sibylle Detel, Malte Elson, Maximilian Frank, Kathrin Fucke, Tobias Heed, Gina Henze, Chris Hopwood, Uwe Kanning, Martin Kersting, Peter Kirsch, Jana Lasser, Seraina Lerch, Martin Kleinmann, Eva Korus, Stefan Krumm, Bernd Markus, Patrick Mussel, Rolf Nichelmann, Tuulia Ortner, Tilman Reitz, Stephan Rixen, Julia Rohrer, Tobias Rosefeldt, Dominik Schwarzinger, Jule Specht, Sonja Utz, Erich Witte, Joachim Wittkowski, Patrick Wöhrle, and the members of MaWi (the Network against Abuse of Power in Science).

The authors would like to thank Paula Knischewski for her help with the typesetting of this article.

Conflict of Interest and Funding

The authors declare that none of them has a conflict of interest associated with the publication of this work. No specific funding was associated with this work.

CRediT Roles

Authors are ordered alphabetically
Conceptualization: NBD, AK, DL, DM, LVP, MS, JS
Writing - Original Draft: DL
Writing - Review and Editing: NBD, AK, DL, DM, LVP, MS, FR, JS
Visualization: DL, JS, AK, DM, LVP

Open Science Practices

This article is conceptual in nature and as such is not eligible for Open Science badges. The entire editorial process, including the open reviews, is published in the online supplement.

References

Abele-Brehm, A. E., & Bühner, M. (2016a). Überlegungen zur Optimierung von Berufungsverfahren in der Psychologie. *Psychologische Rundschau*, *67*(4), 262–268. https://doi.org/10.1026/0033-3042/a000333

Abele-Brehm, A. E., & Bühner, M. (2016b). Wer soll die Professur bekommen? [Who shall get the professorship?] *Psychologische Rundschau*, *67*(4), 250–261. https://doi.org/10.1026/0033-3042/a000335

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature*, *508*(7496), 312–313. https://doi.org/10.1038/508312a

Anvari, F., Alsalti, T., Oehler, L. A., Zach, M., Hussey, I., Elson, M., & Arslan, R. C. (2025). A fragmented field: Construct and measure proliferation in psychology. *Advances in Methods and Practices in Psychological Science*, *8*(3), Article 25152459251360642. https://doi.org/10.1177/25152459251360642

Arcudi, A., Cumurovic, A., Gotter, C., Graeber, D., Joly, P., Ott, V., Schanze, J.-L., Thater, S., Weltin, M., & Yenikent, S. (2019). *Doctoral Researchers in the Leibniz Association: Final Report of the 2017 Leibniz PhD Survey*. Leibniz PhD Network. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-61363-9

Basile, K. C., D'Inverno, A. S., & Wang, J. (2020). National prevalence of sexual violence by a workplace-related perpetrator. *American Journal of Preventive Medicine*, *58*(2), 216–223. https://doi.org/10.1016/j.amepre.2019.09.011

Beadle, B., Do, S., El Youssoufi, D., Felder, D., Gorenflos López, J., Jahn, A., Pérez-Bosch Quesada, E., Rottleb, T., Rüter, F., Schanze, J.-L., Stroppe, A.-K., Thater, S., Verrière, A., & Weltin, M. (2020). *Being a doctoral researcher in the Leibniz Association: 2019 Leibniz PhD Network Survey Report*. Leibniz PhD Network. https://www.ssoar.info/ssoar/handle/document/69403

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*(2), 187–215. https://doi.org/10.1037/0033-2909.117.2.187

Borkenau, P. (2012). Not all authorships are created equal. *Measurement: Interdisciplinary Research and Perspectives*, *10*(3), 147–148. https://doi.org/10.1080/15366367.2012.720192

Brachem, J., Frank, M., Kvetnaya, T., Schramm, L. F., & Volz, L. (2022). Replication crisis, p-hacking and Open Science: A survey of questionable research practices in student projects and impulses for teaching. *Psychologische Rundschau*, *73*(1), 1–17. https://doi.org/10.1026/0033-3042/a000562

Cain, N. M., Pincus, A. L., & Ansell, E. B. (2008). Narcissism at the crossroads: Phenotypic description of pathological narcissism across clinical theory, social/personality psychology, and psychiatric diagnosis. *Clinical Psychology Review*, *28*(4), 638–656. https://doi.org/10.1016/j.cpr.2007.09.006

Chapman, C. A., Bicca-Marques, J. C., Calvignac-Spencer, S., Fan, P., Fashing, P. J., Gogarten, J., Guo, S., Hemingway, C. A., Leendertz, F., Li, B., Matsuda, I., Hou, R., Serio-Silva, J. C., & Stenseth, N. C. (2019). Games academics play and their consequences: How authorship, h-index and journal impact factors are shaping the future of academia. *Proceedings of the Royal Society B*, *286*(1916), 20192047. https://doi.org/10.1098/rspb.2019.2047

Christian, K., Johnstone, C., Larkins, J.-a., Wright, W., & Doran, M. R. (2021). A survey of early-career researchers in Australia. *eLife*, *10*. https://doi.org/10.7554/eLife.60613

Cohen, A., & Baruch, Y. (2022). Abuse and exploitation of doctoral students: A conceptual model for traversing a long and winding road to academia. *Journal of Business Ethics*, *180*, 505–522. https://doi.org/10.1007/s10551-021-04905-1

Corstjens, J., Lievens, F., & Krumm, S. (2017). Situational judgement tests for selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection, and Employee Retention* (pp. 226–246). Wiley-Blackwell. https://doi.org/10.1002/9781118972472.ch11

Czesnick, H. (2020). Konflikte im Zusammenhang mit Autorschaften und Datennutzung – Erfahrungen des "Ombudsman für die Wissenschaft". https://doi.org/10.5281/zenodo.4723737

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). (2024). *Barometer für die Wissenschaft: Ergebnisse der Wissenschaftsbefragung 2023*. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung. https://www.bmbf.de/SharedDocs/Downloads/DE/2023/dzhw-wissenschaftsbefragung.pdf?__blob=publicationFile&v=4

DGPs Board of Directors. (2022). #IchBinHanna: Tenure-Track anstatt Befristung (Pressemitteilung). https://www.dgps.de/aktuelles/details/ichbinhanna-tenure-track-anstatt-befristung/

Dhanani, L. Y., LaPalme, M. L., & Joseph, D. L. (2021). How prevalent is workplace mistreatment? A meta-analytic investigation. *Journal of Organizational Behavior*, *42*(8), 1082–1098. https://doi.org/10.1002/job.2534

Elson, M. (2024). Pay researchers to spot errors in published papers. *Nature*, *629*. https://doi.org/10.1038/d41586-024-01465-y

Elson, M., Fiedler, S., Kirsch, P., & Stahl, J. (2021). Bericht des Ombudsgremiums der Deutschen Gesellschaft für Psychologie – DGPs [OSF Preprint]. https://doi.org/10.17605/OSF.IO/MK9ST

Errington, T. M., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Reproducibility in cancer biology: Challenges for assessing replicability in preclinical cancer biology. *eLife*, *10*, e67995. https://doi.org/10.7554/eLife.67995

European Commission, Directorate-General for Research and Innovation. (2021). *Towards a reform of the research assessment system: Scoping report*. European Commission. https://data.europa.eu/doi/10.2777/707440

Eva, N., Howard, J. L., Liden, R. C., Morin, A. J. S., & Schwarz, G. (2025). An inconvenient truth: A comprehensive examination of the added value (or lack thereof) of leadership measures. *Journal of Management Studies*, *62*, 3072–3117. https://doi.org/10.1111/joms.13156

Evans, N., Buljan, I., Valenti, E., et al. (2022). Stakeholders' experiences of research integrity support in universities: A qualitative study in three European countries. *Science and Engineering Ethics*, *28*, 43. https://doi.org/10.1007/s11948-022-00390-5

Falagas, M. E., & Alexiou, V. G. (2008). The top-ten in journal impact factor manipulation. *Archivum Immunologiae et Therapiae Experimentalis*, *56*, 223–226. https://doi.org/10.1007/s00005-008-0024-5

Fanelli, D. (2009). How many scientists fabricate and falsify research? a systematic review and meta-analysis of survey data. *PLoS ONE*, *4*(5), e5738. https://doi.org/10.1371/journal.pone.0005738

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Fong, E. A., & Wilhite, A. W. (2017). Authorship and citation manipulation in academic research. *PLoS ONE*, *12*(12), e0187394. https://doi.org/10.1371/journal.pone.0187394

Forster, N., & Lund, D. W. (2018). Identifying and dealing with functional psychopathic behavior in higher education. *Global Business and Organizational Excellence*, *38*(1), 22–31. https://doi.org/10.1002/joe.21897

Frank, M., Miller, B., Vosskuhl, J., Zänkert, S., Hesselmann, F., & Joliffe, J. (2024). Error needs culture! Exploring the relationship between error culture and openness [Preprint]. https://doi.org/10.31219/osf.io/kruhn_v2

Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences*, *24*(1), 1–2. https://doi.org/10.1016/j.tics.2019.10.007

Gärtner, A., Leising, D., Freyer, N., Musfeld, P., Lange, J., & Schönbrodt, F. D. (2025). Responsible Research Assessment II: A specific proposal for hiring and promotion in psychology. *MetaPsychology*, *9*, MP.2024.4604. https://doi.org/10.15626/MP.2024.4604

Gasparyan, A. Y., Ayvazyan, L., & Kitas, G. D. (2013). Authorship problems in scholarly journals: Considerations for authors, peer reviewers and editors. *Rheumatology International*, *33*, 277–284. https://doi.org/10.1007/s00296-012-2582-2

Hachmeister, C.-D. (2019). *Im Blickpunkt: Promotionen als Indikator für die Leistung von Hochschulen. Auswertung von Daten des Statistischen Bundesamtes und des CHE Rankings 2019/20*. Centrum für Hochschulentwicklung. Gütersloh, Germany.

Höbel, M., Durglishvili, A., Reinold, J., & Leising, D. (2022). Sexual harassment and coercion in German academia: A large-scale survey study. *Sexual Offending: Theory, Research, and Prevention*, *17*, 1–34. https://doi.org/10.5964/sotrap.9349

Hodson, G., Book, A., Visser, B. A., Volk, A. A., Ashton, M. C., & Lee, K. (2018). Is the Dark Triad

common factor distinct from low Honesty-Humility? *Journal of Research in Personality*, *73*, 123–129. https://doi.org/10.1016/j.jrp.2017.11.012

Hoffmann, B., Beyer, M., Rohe, J., Gensichen, J., & Gerlach, F. M. (2008). 'Every error counts': A web-based incident reporting and learning system for general practice. *Quality and Safety in Health Care*, *17*(4), 307–312. https://doi.org/10.1136/qshc.2006.018440

Horbach, S. P. J. M., Breit, E., Halffman, W., & Mamelund, S.-E. (2020). On the willingness to report and the consequences of reporting research misconduct: The role of power relations. *Science and Engineering Ethics*, *26*(3), 1595–1623. https://doi.org/10.1007/s11948-020-00202-8

Illinger, P. (2019). Widerstand gegen die Wahrheit [Resistance against the truth]. https://www.sueddeutsche.de/wissen/niels-birbaumer-locked-in-syndrom-als-1.4478914

Institut für Demoskopie Allensbach. (2019). *Forschungsfreiheit an deutschen Universitäten: Ergebnisse einer Online-Umfrage unter Professoren und wissenschaftlichen Mitarbeitern* (IfD-Bericht Nr. 8724). Allensbacher Archiv.

Jackenroll, B., & Julmi, C. (2018). Sind deutsche Professoren ausgebrannt? Burnout und dessen Folgen für die Leistungserstellung in deutschen Universitäten [Are German professors burnt out? Burnout and its consequences for performance in German universities]. *Hochschulmanagement*, *13*, 34–40.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kasten, N., Freund, P. A., & Staufenbiel, T. (2020). "Sweet little lies": An in-depth analysis of faking behaviour on Situational Judgment Tests compared to personality questionnaires. *European Journal of Psychological Assessment*, *36*(1), 136–148. https://doi.org/10.1027/1015-5759/a000479

Katsarov, J., Andorno, R., Krom, A., et al. (2022). Effective strategies for research integrity training—a meta-analysis. *Educational Psychology Review*, *34*, 935–955. https://doi.org/10.1007/s10648-021-09630-9

Kis, A., Tur, E. M., Lakens, D., Vaesen, K., & Houkes, W. (2022). Leaving academia: PhD attrition and unhealthy research environments. *PLoS ONE*, *17*(10), e0274976. https://doi.org/10.1371/journal.pone.0274976

Kohrs, F. E., Auer, S., Bannach-Brown, A., Fiedler, S., Haven, T. L., Heise, V., Weissgerber, T. L., et al. (2023). Eleven strategies for making reproducible research and Open Science training the norm at research institutions [Preprint]. https://doi.org/10.31219/osf.io/kcvra

Kreitzer, R. J., & Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, *20*(1), 73–84. https://doi.org/10.1007/s10805-021-09400-w

Kuhnt, M., Reitz, T., & Wöhrle, P. (2022). *Arbeiten unter dem Wissenschaftszeitvertragsgesetz: Eine Evaluation von Befristungsrecht und -Realität an deutschen Universitäten*. Qucosa. Dresden. https://mittelbau.net/evaluation-des-wisszeitvg/

Lasser, J., Bultema, L., Jahn, A., Löffler, M., Minneker, V., & van Scherpenberg, C. (2021). Power abuse and anonymous accusations in academia: Perspectives from early career researchers and recommendations for improvement. *Contributions to Higher Education Research*, *43*(1-2), 48–61.

Leising, D., Gärtner, A., & Schönbrodt, F. (2025). Responsible research assessments (parts I and II): Responses to the commentaries. *MetaPsychology*, *9*, MP.2024.4603. https://doi.org/10.15626/MP.2024.4603

Leising, D., Liesefeld, H., Buecker, S., Glöckner, A., & Lortsch, S. (2024). A tentative roadmap for consensus building processes. *Personality Science*, *5*. https://doi.org/10.1177/27000710241298610

Leising, D., Thielmann, I., Glöckner, A., Gärtner, A., & Schönbrodt, F. (2022a). Ten steps towards a better personality science – how quality may be rewarded more in research evaluation. *Personality Science*, *3*, 1–44. https://doi.org/10.5964/ps.6029

Leising, D., Thielmann, I., Glöckner, A., Gärtner, A., & Schönbrodt, F. (2022b). Ten steps towards a better personality science – a rejoinder to the comments. *Personality Science*, *3*, 1–15. https://doi.org/10.5964/ps.7961

Lemser, S. (2019). *Das Für und Wider einer Karriere in der Wissenschaft: Geschlechtsunterschiede bei Studierenden und Promovierenden am Bereich MatNat der TU Dresden [The pros and cons of a career in science: Gender differences in students and ECRs at TU Dresden's school of science]*

[Master's thesis, Technische Universität Dresden] [Unpublished Master's thesis].

Löwe, P. S., & Valet, P. (2023). Ist der öffentliche Dienst noch ein attraktiver Arbeitgeber? Veränderte Wahrnehmungen zu Jobsicherheit, langfristiger Beschäftigung und Karriereperspektiven von Beschäftigten im öffentlichen Dienst und der Privatwirtschaft von 1985 bis 2018. *Zeitschrift für Sozialreform*, *69*(1), 1–30. https://doi.org/10.1515/zsr-2022-0105

Lynam, D. R., Hyatt, C. S., Hopwood, C. J., Wright, A. G. C., & Miller, J. D. (2019). Should psychologists sign their reviews? Some thoughts and some data. *Journal of Abnormal Psychology*, *128*(6), 541–546. https://doi.org/10.1037/abn0000426

Mackie, S. A., & Bates, G. W. (2018). Contribution of the doctoral education environment to PhD candidates' mental health problems: A scoping review. *Higher Education Research & Development*, *38*(3), 565–578. https://doi.org/10.1080/07294360.2018.1556620

Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behaviour: Big five, or one in addition? *Personnel Psychology*, *60*(1), 1–34. https://doi.org/10.1111/j.1744-6570.2007.00063.x

Martin, K. D., & Cullen, J. B. (2006). Continuities and extensions of ethical climate theory: A meta-analytic review. *Journal of Business Ethics*, *69*, 175–194. https://doi.org/10.1007/s10551-006-9084-7

Mete, Y. A. (2013). Relationship between organizational cynicism and ethical leadership behaviour: A study at higher education. *Procedia - Social and Behavioral Sciences*, *89*, 476–483. https://doi.org/10.1016/j.sbspro.2013.08.880

Miller, J. D., Phillips, N. L., & Lynam, D. R. (2025). Questionable research practices violate the American Psychological Association's Code of Ethics. *Journal of Psychopathology and Clinical Science*, *134*(2), 113–114. https://doi.org/10.1037/abn0000974

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. https://doi.org/10.1037/rev0000111

Müller, W. (2015). Assessment Center in Berufungsverfahren der Universität Bremen. In C. Peus, S. Braun, T. Hentschel, & D. Frey (Eds.), *Personalauswahl in der wissenschaft* (pp. 261–267). Springer.

Neumann, R. (2019). Bitte publik machen, aber ohne meinen Namen! [Please publicize, but without my name]. https://www.laborjournal.de/blog/?p=10331

Nielsen, M. B., Matthiesen, S. B., & Einarsen, S. (2010). The impact of methodological moderators on prevalence rates of workplace bullying: A meta-analysis. *Journal of Occupational and Organizational Psychology*, *83*, 955–979. https://doi.org/10.1348/096317909X481256

Niemann, Y. F., Muhs, G. G. y., & Gonzalez, C. G. (Eds.). (2020). *Presumed incompetent II: Race, class, power, and resistance of women in academia*. Utah State University Press.

Olsthoorn, L. H. M., Heckmann, L. A., Filippi, A., Vieira, R. M., Varanasi, R. S., Lasser, J., Bäuerle, F., Zeis, P., & Schulte-Sasse, R. (2020). *PhDnet Report 2019*. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3243876_4

Ones, D. S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, *9*(1-2), 31–39. https://doi.org/10.1111/1468-2389.00161

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, *78*(4), 679–703. https://doi.org/10.1037/0021-9010.78.4.679

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2012). Integrity tests predict counterproductive work behavior and job performance well: Comment on Van Iddekinge, Roth, Raymark, and Odle-Dusseau (2012). *Journal of Applied Psychology*, *97*(3), 537–542. https://doi.org/10.1037/a0024825

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

Palmer, J. C., Holmes, R. M., & Perrewé, P. L. (2020). The cascading effects of CEO dark triad personality on subordinate behaviour and firm performance: A multilevel theoretical model. *Group & Organization Management*, *45*(2), 143–180. https://doi.org/10.1177/1059601120905728

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in*

Personality, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Peukert, K., Jacobi, L., Geuer, J., Paredes Cisneros, I., Löffler, M., Lienig, T., Taylor, S., Gusic, M., Novakovic, N., Kuhl, T., Ordini, E., Runge, A., Samoylov, O., Härtel, M., Amend, A.-L., & Nagel, M. (2020). *Survey Report 2019*. Helmholtz Juniors. https://www.helmholtz.de/fileadmin/user_upload/06_jobs_talente/Helmholtz-Juniors/Survey_Report2019.pdf

Peus, C., Braun, S., Hentschel, T., & Frey, D. (Eds.). (2015). *Personnel Selection in Science*. Springer.

Prieß-Buchheit, J., Aro, A. R., Demirova, I., Lanzerath, D., Stoev, P., & Wilder, N. (2020). Rotatory role-playing and role-models to enhance the research integrity culture. *Research Ideas and Outcomes*, *6*, e53921. https://doi.org/10.3897/rio.6.e53921

Pruschak, G., & Hopp, C. (2022). And the credit goes to... ghost and honorary authorship among social scientists. *PLoS ONE*, *17*(5), e0267312. https://doi.org/10.1371/journal.pone.0267312

Rahal, R.-M., Fiedler, S., Adetula, A., Berntsson, R. P.-A., Dirnagl, U., Feld, G. B., Fiebach, C. J., Himi, S. A., Horner, A. J., & Lonsdorf, T. B. (2023). Quality research needs good working conditions. *Nature Human Behaviour*, *7*(2), 164–167. https://doi.org/10.1038/s41562-022-01508-2

Regler, B., Einhorn, L., Lasser, J., Vögele, M., Elizarova, S., Bäuerle, F., Wu, C., Förste, S., Shenolikar, J., & Group 2018, PhDnet. (2019). *PhDnet Report 2018*. Max Planck PhDNet. https://doi.org/10.17617/2.3052826

Rennie, D. (2016). Let's make peer review scientific. *Nature*, *535*, 31–33. https://doi.org/10.1038/535031a

Richardson, M., & Healy, M. (2019). Examining the ethical environment in higher education. *British Educational Research Journal*, *45*(6), 1089–1104.

Scheffel, C., Korb, F., Dörfel, D., Scherbaum, S., et al. (2023). Gute wissenschaftliche Praxis und Open Science im Empiriepraktikum: Wissenschaftlicher Kompetenzerwerb durch Replikationsstudien. *Psychologische Rundschau*, *74*(4), 241–243. https://doi.org/10.1026/0033-3042/a000643

Scherpenberg, C., Bultema, L., Jahn, A., Löffler, M., Mineker, V., & Lasser, J. (2021). Manifestations of power abuse in academia and how to prevent them. https://doi.org/10.5281/zenodo.4608392

Schönbrodt, F. D., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L. V., Schmitt, M., Scheel, A. M., Schubert, A.-L., Steinberg, U., & Leising, D. (2025). Responsible Research Assessment I: Implementing DORA and CoARA for hiring and promotion in psychology. *MetaPsychology*, *9*. https://doi.org/10.15626/MP.2024.4601

Schraudner, M., Striebing, C., & Hochfeld, K. (2019). *Work culture and work atmosphere in the Max Planck Society*. Fraunhofer IAO. https://www.mpg.de/14284109/mpg-arbeitskultur-ergebnisbericht-englisch.pdf

Sommer, J., Jongmanns, G., Book, A., & Rennert, C. (2022). *Evaluation des novellierten Wissenschaftszeitvertragsgesetzes*. Interval GmbH & Institut für Hochschulentwicklung e.V. https://www.bmbf.de/SharedDocs/Downloads/de/2022/abschlussbericht-evaluation-wisszeitvg.html

Specht, J., Hof, C., Tjus, J., Pernice, W., & Endesfelder, U. (2017). *Departments statt Lehrstühle: Moderne Personalstruktur für eine zukunftsfähige Wissenschaft*. AG Wissenschaftspolitik der Jungen Akademie.

Stahl, J., Bössel-Debbert, N., Elson, M., Fiedler, S., Kirsch, P., Kluge, A., Leising, D., Mischkowski, D., Phan, L., Richter, F., & Schmitt, M. (2024). *Machtmissbrauch an Universitäten: Strukturelle Ursachen und Ebenen potentieller Maßnahmen. Beitrag zur Ringvorlesung: Missbrauchte Macht*. Universität Würzburg. https://www.psychologie.uni-wuerzburg.de/fileadmin/0602-grk2660/2024/Stahl_Wuerzburg240124.pdf

Strobel, A., Franke-Bartholdt, L., & Kersting, M. (2018). Eignungsinterviews/direkte mündliche Befragungen. In *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430 (2018)* (pp. 66–83). Springer.

Stroebe, W. (2020). Student evaluations of teaching encourage poor teaching and contribute to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, *42*(4), 276–294. https://doi.org/10.1080/01973533.2020.1756817

Tennant, J. P., & Ross-Hellauer, T. (2020). The limitations to our understanding of peer review. *Research Integrity and Peer Review*, *5*, 6. https://doi.org/10.1186/s41073-020-00092-1

Vallbracht, C. (2024). Geldnot an Hochschulen führt zu Entfall an Stellen [Lack of funding leads to loss

of jobs] [Published February 2]. https://www.forschung-und-lehre.de/politik/geldnot-an-hochschulen-fuehrt-zu-entfall-von-stellen-6259

van den Hoven, M., Lindemann, T., Zollitsch, L., et al. (2023). A taxonomy for research integrity training: Design, conduct, and improvements in research integrity courses. *Science and Engineering Ethics*, *29*, 14. https://doi.org/10.1007/s11948-022-00425-x

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, *97*(3), 499–530. https://doi.org/10.1037/a0021196

Vazire, S., & Holcombe, A. O. (2021). Where are the self-correcting mechanisms in science? *Review of General Psychology*, *26*(2), 212–223. https://doi.org/10.1177/10892680211033912

Viđak, M., Barać, L., Tokalić, R., Buljan, I., & Maruić, A. (2021). Interventions for organizational climate and culture in academia: A scoping review. *Science and Engineering Ethics*, *27*, 1–23. https://doi.org/10.1007/s11948-021-00298-6

Weihs, C., Hernandez Rodriguez, T., Doeckel, M., Marty, C., & Wormer, H. (2018). Arbeitszeiten von Professorinnen und Professoren in Deutschland 2016. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *12*, 135–177. https://doi.org/10.1007/s11943-018-0227-y

Wiarda, J.-M. (2023). Das stille Leiden der Betroffenen [The victims' silent suffering]. https://www.jmwiarda.de/2023/11/20/das-stille-leiden-der-betroffenen/

Winkler, M. (2023). Machtmissbrauch an Universitäten [Power abuse at universities]. https://zeitgeschichte-online.de/themen/professorinnen-gegen-machtmissbrauch-universitaeten

Wissenschaftsrat. (2023). *Strukturen der Forschungsfinanzierung an deutschen Hochschulen*. Wissenschaftsrat. https://doi.org/10.57674/pms3-pr05