

# “The Grievance Studies Affair” Project: Reconstructing and Assessing the Experimental Design

Science, Technology, & Human Values  
1-23

© The Author(s) 2020

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0162243920923087

[journals.sagepub.com/home/sth](https://journals.sagepub.com/home/sth)Mikko Lagerspetz<sup>1</sup> 

## Abstract

Recently, high media visibility was reached by an experiment that involved “hoaxlike deception” of journals within humanities and social sciences. Its aim was to provide evidence of “inadequate” quality standards especially within gender studies. The article discusses the project in the context of both previous systematic studies of peer reviewing and scientific hoaxes and analyzes its possible empirical outcomes. Despite claims to the contrary, the highly political, both ethically and methodologically flawed “experiment” failed to provide the evidence it sought. The experiences can be summed up as follows: (1) journals with higher impact factors were more likely to reject papers submitted as part of the project; (2) the chances were better, if the manuscript was allegedly based on empirical data; (3) peer reviews can be an important asset in the process of revising a manuscript; and (4) when the

---

<sup>1</sup>Åbo Akademi University, Faculty of Social Sciences, Business and Economics, Finland

## Corresponding Author:

Mikko Lagerspetz, Åbo Akademi University, Faculty of Social Sciences, Business and Economics, Fänriksgatan 3 A2, Åbo 20500, Finland.

Email: [mlagersp@abo.fi](mailto:mlagersp@abo.fi)

project authors, with academic education from neighboring disciplines, closely followed the reviewers' advice, they were able to learn relatively quickly what is needed for writing an acceptable article. The boundary between a seriously written paper and a "hoax" gradually became blurred. Finally (5), the way the project ended showed that in the long run, the scientific community will uncover fraudulent practices.

### **Keywords**

peer reviewing, hoax, research ethics, gender studies, culture wars

## **Introduction**

The present article discusses a recent case of what can be termed "hoaxlike deception in science" (Schnabel 1994, p. 459). The project is already widely known by the authors' self-chosen name, "The Grievance Studies Affair." In 2017-2018, three authors, Helen Pluckrose (MA in early modern studies), James A. Lindsay (with a doctorate in mathematics), and Peter Boghossian (assistant professor of philosophy) compiled a series of hoax articles and submitted them to journals of gender studies, sociology, ethnography, and related fields. Boghossian and Lindsay had previously (in 2017) gained visibility with a hoax article, "The Conceptual Penis as a Social Construct," submitted for publication to a peer-reviewed gender studies journal. It was refused but then published in a fee-charging Open Access journal. Despite the actual failure, they presented the hoax as evidence of inferior quality control within gender studies.

With the new project, now the three authors wished to show that "especially in certain fields within the humanities"—which they refer to as "grievance studies"—scholarship is "corrupt" (Pluckrose, Lindsay, and Boghossian 2018). They claim a need to "begin a thorough review of these areas of study (gender studies, critical race theory, postcolonial theory, and other 'theory'-based fields in the humanities and reaching into the social sciences, especially including sociology and anthropology)." The name of the project, with its connotative relations to "The Dreyfus Affair" and beyond, seems to signal the authors' ambition to initiate a media debate of high visibility and, maybe, divisiveness.

The authors published their "experiment" on October 2, 2018. The media coverage so far, including the project's own media communication

(Pluckrose, Lindsay, and Boghossian 2018), explicitly makes a connection with the ongoing US and global “cultural war,” part of which is a political distrust of some fields of research, especially gender studies. The problem (the authors claim) is “leaking” to other disciplines—education, politics, public discussion, activism, and so on—and “needs to be dealt with.” According to them, radical constructivist theory has reached an authoritative position in these fields:

This problem is most easily summarized as an overarching (almost or fully sacralized) belief that many common features of experience and society are socially constructed. These constructions are seen as being nearly entirely dependent upon power dynamics between groups of people, often dictated by sex, race, or sexual or gender identification. All kinds of things accepted as having a basis in reality due to evidence are instead believed to have been created by the intentional and unintentional machinations of powerful groups in order to maintain power over marginalized ones. (Pluckrose, Lindsay, and Boghossian 2018)

To begin with, framing constructivism (or constructionism) as some kind of conspiracy theory misrepresents it almost as its exact opposite. Nor is the point of social constructionist analysis to make judgments about what exists and what does not; rather, it is about entering a dialogue on taken-for-granted social realities (Gergen 1999, 114; Weinberg 2014). My aim in this paper is to assess the “Grievance Studies Affair”: did it provide “strong evidence” of its own claims? If it did, then what did that evidence demonstrate? Using the material linked on the project’s website (the writers’ own texts, reviewers’ statements, editorial decisions), I reconstruct the activity as the experimental study that it claims to be. In other words, I give the project the benefit of the doubt and assess the project as a piece of research, trying to identify the relevant hypotheses and empirical results. I will start with an overview of some earlier studies and “hoaxes” that have used deception as a means of testing editorial practices of scholarly journals. I will then cite some of the project’s media coverage. After that, my own analysis will follow. I will examine the authors’ actual success in receiving positive editorial decisions for their manuscripts, with reference to the type of text submitted and to the impact factor of the journals targeted. After that, I will discuss the extent to which a reviewer can reasonably be expected to recognize “a hoax,” the ethical considerations involved, and the view of science and scientific communication behind the project. My conclusions are about the lessons to be learned.

## **Studies of Editorial Practice in Academic Publications**

An experimental way of testing the publishing criteria of the editors and reviewers of academic publications is to submit purpose-written texts for evaluation. Armstrong (1997) provides an overview of seventy studies on peer review, among them twelve experiments (or quasi-experiments) made between 1975 and 1997. Bornmann's (2011) overview concentrates on research from the first decade of the 2000s. The experimental studies explore biases, that is, the fairness of the peer-review process, and the reliability of the peer-review process, mostly measured as the rate of inter-reviewer agreement.

Researchers in psychology and related fields were the first to focus on different forms of reviewer bias. Abramowitz, Gomes, and Abramowitz (1975) prepared a brief bogus manuscript on student activism in two versions with different political implications. They sent it for review to psychology scholars with different assumed political leanings and found some support for their hypothesis about reviewers' tendency to favor papers whose political tone corresponded with their own (p. 193). Mahoney (1977), and later on, Epstein (1990), approached the impact of confirmatory bias, that is, of granting credibility primarily to findings that seem to confirm the editors' and reviewers' previous beliefs. Both used as experimental stimuli bogus articles in two versions, where the empirical results seemed either to confirm or disconfirm the perspective originally chosen. Both found evidence of reviewers' "confirmation bias." Whereas Abramowitz, Gomes, and Abramowitz (1975) and Mahoney (1977) asked for reviews from selected scholars, Epstein's (1990) study included the editorial decisions as well. He approached 146 journals within social work and neighboring disciplines. In addition to a statistical analysis of acceptance rates for the two versions, he also made a qualitative analysis of the reviewers' statements, identifying the issues raised, and the use of subjective and objective arguments. His overall impression was that the statements were generally of poor quality. They were often subjective and focused on formalities. The best ones came from the "allied disciplines," not from within the social work field itself (Epstein 1990, 24-25). Other possible sources of bias include the gender of the reviewer and of the author, and the latter's academic position, institutional affiliation, and prestige. Both have been studied by experiments, but the results are mixed (Bornmann 2011, 215).

All studies of reliability show low inter-referee agreement rates, typically falling in the range from 0.2 to 0.4 when corrected for chance (Bornmann 2011, 207). A comparison between natural and physical sciences,

humanities, and social sciences shows that this is the case irrespective of discipline (Bornmann, Mutz, and Daniel 2010, 6). Moreover, peer review rarely recognizes fraud or conflict of interest (Cowley 2015, 4). Recent discussions have, however, paid attention to its other possible functions. If the reviews are substantial and helpful, they can contribute to the final shaping of the paper. Peer reviewing also seems to create and reshape networks of collaboration (Dondio et al. 2019). The most strongly held argument for new, “open peer review” practices (where neither authors nor reviewers are anonymous) is about increased interaction, which in turn is believed to result in better publications (Ross-Hellauer, Deppe, and Schmidt 2017, 17). In short, the debate about peer reviewing is currently less about its ability to determine the quality of scholarship and more about its contribution to dialogue.

### **“Hoaxlike” Deception in Science**

When discussing the “Grievance Studies Affair” project, systematic experiments are, however, not the most obvious context of comparison. There are examples of another type of “experiment,” for which the term “hoaxlike deception in science” (Schnabel 1994) seems adequate. It consists of manipulative deception of a researcher or a research-related institution with the aim of demonstrating their putative incompetence. Based on five historical cases from the twentieth century, Schnabel (1994) concluded that the “success,” or the academic and media reception of the hoax, was heavily influenced by the previous position of the target vis-à-vis prevailing scientific orthodoxy. The weaker the institutional backing, the less weight was attributed to the target person’s counterarguments about research ethics, lacking validity of the experiment, and so on.

What in the strict language of research ethics would be called fraud is often, in a more conciliatory manner, referred to as “hoax.” The perpetrators themselves tend to favor the term “experiment,” implying both knowledge creation and a context that can be argued to justify manipulative practices. However, there are important differences between hoaxes and experiments worth the name. Even when submitting bogus papers in order to test the reviewers for biases and the review process for reliability, the systematic experimenters’ aim is not to embarrass but to find out how these institutional practices function. In contrast, the hoax is done explicitly in order to discredit its targets. There are many signs of this: sensationalism in reporting, lack of systematic experimental design, and the fact that we seldom hear about those hoaxes that have failed.

A well-known case of hoaxlike deception was performed by the physicist Alan Sokal (1996). Sokal managed to publish an article criticizing his own discipline in *Social Text*, a non-peer-reviewed cultural studies journal. Immediately after publication, Sokal revealed the hoax in another journal and received loads of mostly positive media coverage. Hilgartner's (1997) comparison of the "Sokal Affair" with Epstein's (1990) study and its reception has been a major source of inspiration to the present article. In Sokal's "experiment," no peer-review process was involved; Sokal actually allowed the journal to publish the parody. His experiment had no clear hypotheses, controls, or explicit analyses of what had been achieved; and the population experimented with consisted of just one avant-garde journal outside the academic mainstream of its research field (Hilgartner 1997, 517). The most interesting difference was, however, the reception. At the time Epstein conducted his study, he was an individual social worker with no university affiliation (Hilgartner 1997, 516). He had targeted a well-organized professional field and was summoned to defend himself in front of an ethics panel. Sokal, holding a prestigious academic position in a prestigious discipline, targeted "downward" at a nonprestigious journal in a field under constant political attack. The hoax propelled his career toward the status of an academic celebrity (Hilgartner 1997, 515-19). In a way, Sokal thus proved a point in science and technology studies, which he himself had set out to criticize (cf. Schnabel 1994). As Hilgartner (1997, 519) puts it, he ended up producing "a parody not only of cultural studies but also a parody of himself."

The emergence of Open Access, online publications has created, among other things, (sometimes-justified) suspicion about editorial standards. This concerns above all the so-called predatory journals, which charge author's fees and publish texts unselectively. In 2013, fee-charging Open Access journals were targeted by an experiment devised by the science journalist John Bohannon (2013), who sent a fabricated and clearly flawed pharmaceutical study to 304 journals with a biological, chemical, or medical title. Out of 255 journals that announced their final decision during the time span of the project, 157 accepted the paper and 98 rejected it. Bohannon always withdrew his paper before publication. Finally, there is a piece of computer software called SCIfgen (Stribling, Krohn, and Aguayo 2005), developed in 2005 by three MIT students, to produce meaningless fake computer-science papers. It is freely accessible and has been used countless times for fooling fake conference organizers and journals. "Predatory journals" in all disciplines seem easy prey for such hoaxes.

## The Grievance Studies Affair: The Media Response

Within the project calling itself “The Grievance Studies Affair,” the three members of the group authored and offered for the publication of nineteen, twenty, or possibly twenty-one<sup>1</sup> manuscripts, all of which are available through a link on their website (with the exception of two that are said to be rewritten adaptations of other papers). Of the decisions received from journal editors and reviewers, the first is dated September 22, 2017, and the last September 6, 2018. These documents are likewise available through a link. The group went public with the project on October 2, 2018, when one of the published articles had attracted media attention, including suspicions of fraud.

Within days, the project web page was referred to by both US and European<sup>2</sup> mainstream media, by fashionable alt-right intellectuals, and even by blog posts of serious academics. As al-Gharbi (2018) summarized on October 10, 2018:

the incident has become something of a Rorschach Test within academic and media spaces: Those already disposed towards skepticism of “critical” scholarship on race, gender or sexuality view the incident as damning proof of a deep rot within these subfields, and perhaps with social research or academia more broadly. Those more sympathetic to “critical studies” instead see a deeply flawed and limited experiment—and accuse the authors of overstating their findings, speaking beyond their data and, intentionally or not, feeding into the agenda of right-wing reactionaries. There have been relatively few unexpected validators or critics.

Many of the issues that will be focused later on in this article were raised immediately: the project authors’ exaggeration of what was actually shown (Afinogenov 2018; Engber 2018), ethical considerations (Bergstrom 2018), misrepresentation of reviewers’ reports (Schieber 2018), lack of controls in the experimental design (Drezner 2018), and missing information about the funding of the project (Essig and Moorti 2018). Despite claiming to target a wide range of disciplines, the project actually did not manage to publish anything within race studies or sociology. al-Gharbi (2018) pointed out that the hoax only had some success within gender studies, a field which actually neighbors those of Pluckrose and Boghossian, and which “The Conceptual Penis” paper had previously tried to attack. Both al-Gharbi (2018) and Drezner (2018) made a point that will be discussed further on in this article: the experiment included not only hoaxing the reviewers and editors but also learning from them.

The more positive responses either echoed the project authors' claims about corrupt scholarship in the academia (e.g., Mounk 2018b) or dismissed the use of scientific or ethical standards when evaluating the project (Mounk 2018a; Smith 2018).

## **An Attempt at Empirical Analysis**

The project website (Pluckrose, Lindsay, and Boghossian 2018) presents almost no analysis of the results. It focuses instead on the motifs behind the project and on colorful descriptions of the texts written (both those published and those refused). However, the website provides links to most of the manuscripts and to the correspondence with those journals that used peer review. The manuscripts are presented in their most recent versions, sometimes rewritten after several submissions. This means that it is not possible to trace the revisions undertaken. What an outsider can analyze, however, are several editorial decisions, review processes, and the finalized manuscripts.

If the aim is to test the probability of a journal accepting an article proposal, the proper unit of analysis is one final decision (accept/reject) about one manuscript by one journal. In other words, cases of resubmission of a revised manuscript to a journal after initial rejection by the same journal count together as one single case. The seven unresolved cases (submissions without final decision) are left outside the analysis. This leaves us with twenty-one cases, that is, twenty-one final decisions concerning seventeen manuscripts (see Table 1). Seven final decisions were positive (with four texts actually published—"Dog Park" [i.e., "Wilson" 2018], "Fat Bodybuilding," "Dildos," "Hooters," all of them later retracted), fourteen were negative. The authors argue on their website that if the project had been allowed to continue, some of the seven unresolved cases would have yielded a positive decision; however, the present analysis only concerns the outcomes actually achieved (and reported; see Note 2). In Table 1, the manuscripts are referred to by the abbreviations that appear on the project website. I have decided not to give the names of the journals targeted by the deception or full references to the texts. Exception is made for two texts, which I will discuss in detail and refer to as "Wilson" (2018) and "Baldwin" (2018). As my discussion will show, I do not find the fact that they were accepted for publication as particularly embarrassing for the journals that did so.

The submitted texts show extreme variation in the issues discussed (from dog parks to political satire, from martial arts to masturbation), styles, and

**Table 1.** All Cases, by Decision Date, Abbreviation, Method, Journal Impact Factor, Number of Reviews, and Decision.

Decision Date (dd/mm/yy)	Article	Method	Two-Year Journal Impact Factor (Median = 1.086)	Number of Reviews	Decision
22/09/17	BJJ	T	2.453	—	N
19/11/17	BJJ	T	2.115	1	N
14/02/18	BJJ	T	1.636	2	N
19/02/18	Dog Park <sup>a</sup>	E	1.542	3	Y
25/02/18	Moon Meetings	P	0.486	—	Y
02/03/18	Feminist Bodybuilding	T	1.250	2	N
13/03/18	Fat Bodybuilding	T	0.458	3	Y
21/03/18	FMK	T	1.250	2	N
25/05/18	Hooters	E	2.115	3	N
09/06/18	Dildos	E	1.028	3	Y
16/07/18	CisNorm	E	2.453	4	N
31/07/18	HoH2 <sup>b</sup>	T	0.777	3	Y
09/08/18	Masturbation	T	2.030	2	N
09/08/18	WMK	T	N/A	3	N
21/08/18	FMK	T	0.861	4	Y
06/09/18	Hooters	E	1.086	2	Y
N/A	HoHI	T	0.952	—	N
N/A	The Autoethnography	T	0.796	—	N
N/A	As SZE	T	1.034	—	N
N/A	FemMK	T	0.911	—	N
N/A	Queering Plato	T	N/A	—	N

Note: N = 21. T = theoretical; E = empirical; P = “poetic inquiry”; N = declined; Y = accepted; N/A = not available.

<sup>a</sup>Discussed below as “Wilson” (2018).

<sup>b</sup>Discussed below as “Baldwin” (2018).

methods. Hence, no strict comparison of the journals’ standards is possible. Nor is it possible to compare the peer-review statements in the way done by Epstein (1990). One cannot tell whether all manuscripts (stimuli) were equally “inferior” or “good.” Some manuscripts predictably call for more, some for less, attention on, for example, theory, the conceptual apparatus, the empirical method, the conclusions, the language and formalities of writing, and so on.

The authors presume that they have unmasked the peer-review systems of the journals they experimented with as “inadequate” (Pluckrose, Lindsay, and Boghossian 2018). They are keen to stress that this does not concern the peer-review system across all disciplines. However, no control

**Table 2.** All Cases, by Decision, Journal Impact Factor, and Number of Submissions/Manuscript.

Cases	N	Two-Year Impact Factor 2017, Means <sup>a</sup>	Number of Submissions/ Manuscript, Means
All completed experiments	21	1.328	1,58
Accepted, all	7	0.891	2,57
By method			
Theoretical	3	0.699	2,67
Empirical	3	1.213	3,00
“Poetic inquiry”	1	0.486	1,00
Declined, all	14 (12) <sup>b</sup>	1.583 <sup>c</sup>	1,20
By method			
Theoretical	10	1.443	1,25
Empirical	2	2.284	1,00

<sup>a</sup>The impact factor (average number of yearly citations of one article published in the journal, counted on a two-year period) ranged from 0.458 to 2.453. The median was 1.086 (Source: SCImago, n.d.).

<sup>b</sup>The impact factor of two journals was not available.

<sup>c</sup>The difference between impact factors of journals which accepted and those which declined the manuscript was statistically significant (Student’s *t*-test,  $p = .0168$ ).

experiment was done with other fields of scholarship, nor were the results compared with previous similar experiments. In addition, not all journals targeted represent the academic mainstream of their respective fields.

The information provided on the project’s website and the documents linked to it make it possible to find information on some variables other than the discipline. Two-year impact factors, as of 2017, can be found for most of the journals involved. The figure used here is the average number of citations received by one article in the journal in question, calculated for the latest two years (SCImago n.d.). As to the manuscripts themselves, they vary with respect to one important variable: twelve manuscripts were “theoretical,” that is, not supported by original empirical data, while one was a “poetic inquiry,” not presenting itself as a research article in the first place. Five manuscripts were “empirical,” that is, referred to (fabricated) empirical data. Interestingly, three of the latter were among the seven manuscripts accepted for publication. As shown by Table 2, the “empirical” manuscripts were also more successful as to the impact factor of the journals that accepted them. In general, publication was more easily achieved when the impact factor was low. This negative correlation was statistically

significant (see remark to Table 2). The necessary conditions for publication were, thus, that the article was either empirical or the journal impact below median.

Of the twenty-one editorial final decisions, six were direct refusals by the editor and one was the acceptance without peer reviewing of the “poetic inquiry” by a journal specializing in such texts. A further fourteen decisions were based on peer reviews. The reviews concerned altogether eleven manuscripts, of which three were submitted to more than one journal. Also when resubmitted to the same journal, a manuscript could receive peer reviews twice or thrice. Altogether, there were thirty-seven reviews (see Table 1), totaling to 204,171 characters (without spaces) or 38,076 words. On average, one review contained 1,029 words; the average length was about the same for the published and for the refused manuscripts. For comparison, the reviews received in Epstein’s (1990, 19) study were typically just 150-300 words, and he cites previous research stating that referee reports in psychology averaged more than 500 words.

What can be concluded from the analysis is, first, the unsurprising fact that editors and peer reviewers were not equally demanding in all journals. Chances for being published were better in journals representing less well-established disciplines (such as fat studies), scarcer or zero within the established ones (such as sociology). Journals focusing on new and very specific fields also have lower impact factors. Secondly, the claim to present original empirical data clearly made the journals more inclined to publish. One cannot tell, of course, whether this applies only to the set of journals selected for the study or also to peer-reviewed journals in general. Finally, the peer reviewers’ role in the publication process appears essential in more than one way. They are not mere gatekeepers recommending publication or refusal. In the cases that resulted in publication, they first recognized the potential of developing the manuscript and then assisted the writer with feedback impressive already by its sheer volume. The eleven manuscripts that finally went through the review process received a total of ninety-seven suggestions of new literature. One of the articles, “FMK,” submitted to two journals, received in all six reviews. Similarly with the reviews of other articles, they addressed the manuscript contents in detail. Apparently, the authors also considered the advice. “I have seldom had the joy of working with a revision in which the reviewers’ feedback was so diligently and thoughtfully attended,” one reviewer puts it in her or his final statement before the article was accepted for publication.

## **How to Recognize a Hoax**

Both Mahoney (1977) and Epstein (1990) started their experiment with a hypothesis about confirmational bias. Their primary aim was not to test the reviewers' quality standards; they delivered credible papers that varied with reference to the central variable of the hypothesis ("confirming" or "disconfirming" the theory discussed). On the other hand, Sokal (1996) and the various hoaxes with predatory journals used obviously flawed or outright nonsensical papers as stimuli. The "Grievance Studies Affair" project seems to have done something in between. The authors describe their papers as "outlandish and intentionally broken in significant ways." At the same time, the authors were clearly keen to have their papers accepted, and (judged based on comments in peer reviews), they meticulously followed the advice given. In a way, they ended up experimenting with themselves: with their own ability (as trained academicians from neighboring disciplines) to amass "what appears to be significant evidence and sufficient expertise" (Pluckrose, Lindsay, and Boghossian 2018).

One of the papers was a combination of poetry with individual experiences interpreted in a theoretical framework. It was accepted without peer review. Was that one also a hoax? Maybe. However, I will leave it outside the present discussion, just noting that as to fiction literature, "beauty is in the eye of the beholder."

We saw in the previous section that reference to new empirical findings was an important factor in favor of publication. The project's flagship, the "Dog Park" paper, was about dogs' sexual behavior in a dog park, and about humans' reactions to that behavior. Commenting on that paper, the authors mention "incredibly implausible statistics." What could that mean? Obviously, statistical data can be "implausible" when highly surprising (and thus, potentially interesting); or whenever there is reason to suspect flawed methodology. The reviewers did, in fact, ask for clarifications about the method, which were duly added to the paper. According to "Wilson" (2018, 6), those statistics were based on "nearly 1000 h of public observations of dogs and their human companions." "She" also provides a close description of how the data were supposedly gathered ("Wilson" 2018, 8). Looking at the final version myself and not knowing that the data were fraudulent, I would probably also have accepted the methodology. Besides, there exists little research on the context addressed—the relationship between dogs and humans. As a peer reviewer, I would myself have expected more conclusions about the dogs' human companions' practice of attributing human feelings, thoughts, motivations, and beliefs to the behavior of dogs (cf.

Serpell 2003, 83-84) and less about the more distant political implications. I might nevertheless have welcomed, presumably with many others, this empirical addition to research on a poorly covered area.

Another published empirical article, the “Hooters” paper claimed to be based on weekly observations in a “breastaurant” during two years. The editor summarized the reviewers’ concerns about methodological rigor as such: “This then takes me to a core challenge [. . .]: trustworthiness. All three reviewers share my concern about the lack of demonstrated methodological integrity in the present paper.” The published version is then provided, in due course, with a detailed account of the method, including a mention about its “accordance with the ethical standards of the institution and with the 1964 Helsinki declaration and its later amendments [. . .].”

In theoretical studies, fraud is harder to define. Outright plagiarism is one type, misrepresentation of references is another one (which the authors claim not to have done). The papers “FMK,” “WMK,” and “FemMK” are said to be based on a passage in Hitler’s “Mein Kampf” (and this claim has greatly added to the project’s media visibility). According to a philosopher colleague, the passage in question discusses strategies of political communication and of forging alliances. However, everything historically specific in Hitler’s text (racism, references to the First World War, and so on) has been removed, and the remaining similarities between the texts are merely about the article structure (and besides, “Hitler was a better stylist”). Plagiarism is not the type of fraud applied here. The authors’ claim to have fooled the review process is based on the expectation that a competent reviewer would have uncovered the hoax. Nevertheless, when the authors heed the reviewers’ advice and adjust their papers accordingly, the border between a hoax and a seriously written paper gradually becomes more and more blurred.

## **Ethical Considerations**

In experimental studies, deception is sometimes regarded as defensible because of its benefit to scientific ends and, in turn, to society (Hegtvad 2007, 154). In the present case, the issue of deception enters the discussion from five different angles. (1) At least some of the papers were in fact fraudulent when presented and published as empirical research; (2) the experimental design was based on deceiving the editors and reviewers; (3) the authors motivated their mode of action by claiming that the journals and disciplines targeted were themselves deceptive, presenting political standpoints and academic gibberish as “scholarship”; (4) importantly, the

authors present their own project as a scientific experiment that has given “strong evidence” of that claim. Accordingly, also their reporting of the project must be judged by the standards of good scientific practice; (5) finally, reporting the sources of research funding is also a part of that practice.

On their website (Pluckrose, Lindsay, and Boghossian 2018), the authors do not waste many words on the issue of the ethics and integrity of their own project. They comment on the way their experiment ended:

With major journalistic outlets and (by then) two journals asking us to prove our authors’ identities, the ethics had shifted away from a defensible necessity of investigation and into outright lying. We did not feel right about this and decided the time had come to go public with the project.

Contrary to that statement, “outright lying”—to the editors, to the peer reviewers, and to those who possibly read the published articles—of course was a part of the project from the outset. The three “empirical” articles published (“Wilson” 2018; Hooters; Dildos) and the two rejected ones (CisNorm; Porn) present fabricated data, which counts as serious scientific misconduct (e.g., European Science Foundation 2011, 13). On the other hand, apart from misrepresenting the authors’ identity, the nonempirical manuscripts were not, as such, instances of misconduct, even if they were intentionally written to be “inferior” (at least in the beginning, before rewriting—see previous section).

The authors had no informed consent of the persons (editors, reviewers) with whom they were experimenting. A possible justification would be that no alternative ways of testing their practices were feasible (see Hegtvedt 2007, 154). Other studies of peer reviewing have also used deception. However, as a rule the researchers did not allow their experimental stimuli to be published; yet, Epstein still received heavy accusations of misconduct (Hilgartner 1997, 514). As to “The Grievance Studies Affair” project, one must ask whether the actual publication of the stimuli as journal articles added to any data possibly to be discovered by the experiment. The answer is clearly in the negative: publication added to journalistic appeal and to the embarrassment of the subjects experimented with, but not to its content.

Manuals and textbooks on research ethics stress the imperative of minimizing any possible harm to the research subjects. “The Grievance Studies Affair” project clearly intentionally aimed to harm certain journals and disciplines. The justification the authors give for that, however, is not easy to accept at face value. It includes no accusation of deceptive business

practices (such as can be used to justify attacks on “predatory journals”). The authors’ attack is on the purportedly “fatally flawed research” of others within fields influenced by “critical constructivism and radical skepticism.” They want to “push for universities to fix this problem” (Pluckrose, Lindsay, and Boghossian 2018). However, challenging theoretical standpoints or university politics does not require deceptive experiments (submitting one’s own intentionally fatally flawed research to journals to see what will happen); it can be more properly done in regular academic debate. Interestingly, one of the (refused) manuscripts written as a part of the project, “HoH1,” also formulates a matching justification: “hoaxes on unethical fields are morally justifiable, and hoaxes on ethical fields are unjustifiable” (p. 3). When the authors on their website describe this article and what it claims, they confess that it advocates “a blatant double standard.”

Research-based knowledge has a powerful position in public discourse. This calls for ethical standards higher than those acceptable in other media debate. Therefore, did the experiment bring about “strong evidence” of the “Corruption of Scholarship” in “especially some fields of the humanities,” as the authors claim? The present analysis has shown that it did not. If we treat the project as the kind of empirical experiment it claims to be, its external communication becomes a case of an “attempt to exaggerate the importance and practical applicability of the findings” (European Science Foundation 2011, 14).

The project obviously included time-consuming commitment by three researchers for more than a year, plus the work of a filmmaker making a documentary. Everything was funded by an anonymous “benefactor.” Secrecy about the funding source opens questions about conflicts of interest. Did it have effect on which journals were chosen as targets or on the way findings were reported? Political activism and research communication have become intermingled in a way that is potentially harmful to the credibility of research in general. Paradoxically, this is what the authors themselves claim to be against.

## **Gatekeeping and Scientific Dialogue**

Finally, what is the view of science and scientific communication behind the project? The authors criticize (their own representation of) constructivism, and the alternative they offer is simple: “scholarship [...] must be rigorous [...] we should be able to rely upon research journals, scholars, and universities upholding academic, philosophical, and scientific rigor” (Pluckrose, Lindsay, and Boghossian 2018). Authors of “The Grievance

Studies Affair” tried to show the lack of “rigor” by submitting for review manuscripts they depict as “nutty” and as including “some little bit of lunacy or depravity.” They are about issues seldom addressed by research: masturbation, dogs’ sexual behavior, dildos, “breastaurants,” and so on. The argument seems to be that journals publishing, or even considering, texts on such unserious issues are themselves not to be taken seriously. However, the very novelty of a research topic or an argument often led to favorable comments from the reviewers. Reviewer #1 of the published “Dildos” article comments,

I think it is true that there is “virtually no rigorous scholarly work [that] investigates the topic of improving straight male partner sensitivity by means of receptive anal eroticism,” but I do wonder what such a study would look like. Right now, it seems, that much of this is speculative, but nonetheless interesting and provocative.

To write on less-researched issues looks in fact like an effective strategy and not only in the journals and disciplines targeted by the project. Bourdieu (1975, 22) remarks that willingness to deal with less legitimate and less prestigious research topics may be profitable because of less intense competition. This does not mean that all “lunacy or depravity” is equally welcome. Contrary to what the authors suggested, also purportedly “radical constructivist” claims were often met with criticism. Consider this polite statement by Reviewer #1 of the refused “Feminist Bodybuilding” article:

To say that woman bodybuilders (and athletes more broadly) and their capacity to build large muscles have been limited by societal standards of femininity is a fair argument. But it is not tantamount to saying socialization is ALL that has restricted women and biology plays no role, which is a much more difficult argument to make. The latter seems to be the author’s claim, and it isn’t adequately substantiated.

One of the papers (“Baldwin” 2018) was finally published after three rounds of reviews. It discussed political satire. The paper focused on one important point: the difference between satire aiming “upward” and “downward”; on prevailing (hegemonic) discourses or on nonhegemonic ones. For a reader’s point of view, I contacted a colleague—a sociologist who has done research on related issues. I sent him or her the paper without revealing its background. Her or his reading was as follows (quoted here with relevant permission):

My initial feeling was confusion about why everything needs to be discussed in such convoluted ways. At the same time, I have to say that what I interpreted as the point [of the text] is important: that humor has different meanings (as subversive or hegemony supporting) depending on power relations and positionality of those involved. That is how I still think. The point is of course not a new one, and I wondered why the author did not refer to the usual classical treatises on the functions of humor (e.g., Aristotle, or Freud). [...] I was in a somewhat suspicious mood when reading, but I could see a lot of humor in the article itself, too. I hope that also in a more usual situation, I would have recognized the ironic distance.

This reader points at a function that even a less-than-perfect article can have: it can include a point still worth discussing. When presenting the reviewers and editors as gullible, the authors fail to see the dialogical character of the endeavor called science: it is not a mere reservoir of knowledge, but an ongoing debate (cf. Popper 1963, 127). To that, any carefully argued point of view can contribute. New fields of study emerge, and new research questions arise, sometimes on less prestigious arenas to begin with. Eventually, they will be integrated into the body of mainstream scholarship or pass into oblivion. One should also not forget that the experiment in fact was ended because of suspicions of fraud aired in public debate. Even that is a part of the empirical outcome. The debate in and around science is, among other things, a mechanism of quality control.

The project authors accuse the peer reviewers for bias and laxity of quality standards. As we have seen, such criticism is not new and applies to all disciplines. At the same time as its deficiencies as a quality screening mechanism have become clearer, peer reviewing has paradoxically reached growing importance in the industry of measuring “excellence” and ranking universities, research groups, and researchers. Within research politics, to claim shortcomings of the process within some specific field thus amounts to questioning the field’s academic legitimacy and accordingly, that of any financing it receives.

## Conclusions

The “Grievance Studies Affair” project is part of the more general phenomena of the US culture wars and anti-intellectualism, by now turned global. The project’s primary focus on gender studies might reflect its authors’ personal frustration with that specific field, but it also coincides with concerted political action both in the United States and in Europe (Kuhar and

Paternotte 2017). Before learning more about the project's funding, it is too early to say whether this connection is incidental. The present analysis focuses on the question of what the experiment demonstrated empirically and its ethics.

First, the study lacked the comparative element that could make it possible to argue that the disciplines that were targeted have laxer quality standards than those that were not. The percentage of submissions with positive decisions (33 percent) is close to that of the social work journals studied by Epstein (1990), but a comparison is impossible because of the different character of the stimuli. In addition, some of the project's later communication suggests that the total number of submissions and refusals during the project's entire life span was in fact substantially higher than during the later phase that the project website reports about, and the real percentage of positive decisions thus substantially lower (see Note 2). The experiment did not demonstrate the peer reviewers' negligence. Instead, the reviews linked to the project's website are thorough (which is also shown by their length compared with those referred to by earlier studies of peer reviewing).

The journals targeted by the project were not chosen in any systematic way to be representative of all journals in their respective fields. Accordingly, all possible conclusions will concern the project's haphazardly chosen set of data only. No external validity can be claimed. With these severe limitations, experiences from the project can be summed up as follows: (1) journals with higher impact factors were more likely to reject papers submitted as part of the Grievance Studies Affair project; (2) the chances of having a submission accepted were much better if the manuscript claimed to be based on an empirical study; (3) peer reviews are an important asset for the process of revising a manuscript; (4) closely following the reviewers' advice, the project authors, with academic education from neighboring disciplines, were able to learn relatively quickly what is needed for writing an acceptable article; finally, (5) the way the Grievance Studies Affair project ended suggests that the scientific community will ultimately uncover fraudulent practices.

The first conclusion needs little comment. Less established journals in new and specific fields of research have fewer manuscripts to choose from, which may lead to lower thresholds for publication. However, from the positive side, this may also mean more risk-taking, which can prove worthwhile in the end. Even a less-than-perfect article may include an interesting point that deserves to enter the scientific discussion. The second conclusion about the relative priority given to empirical studies seems important.

Studies of peer-review seldom address this issue. The project exploited one important vulnerability of the review process: The reviewers have most often no realistic possibility of uncovering the fabrication of data. They can only assess what the writer decides to tell them about his or her methodology and, when in doubt, demand clarification.

The project's experiences with peer reviews run contrary to some of the recent criticism that has been aired in discussions about academic publishing (e.g., Wagenknecht 2018), in which authors have said they found reviews unhelpful. In this project, they were long, detailed, and obviously written by dedicated people. It is possible that some shortcomings of the peer-review process are more poignant within other disciplines than those targeted. With substantial help from the reviewers, the authors were in fact able to produce articles that are not recognizably amateurish, as was shown by my own little "experiment" of asking an independent opinion about "Baldwin" (2018). Sokal's memorable hoax illustrated, if anything, that journals tend to be benevolent toward an author with well-established academic credentials. At least some cases within the Grievance Studies Affair project point at benevolence of another kind: the editors and reviewers were not barring the entry to their research field of unknown authors from other disciplines but went through the (unpaid) voluminous process of providing beginners with thorough comments and advice.<sup>3</sup>

The final uncovering of the hoax by journals and media shows one more point that is important: scientific studies and articles are part of an ongoing discussion, not separate pieces of knowledge. Fraud tends to be unmasked. For this, dialogue is more important a function than gatekeeping. On the other hand, even less "rigorous" articles may add something to the realm of academic debate. A far-fetched treatment of Plato (as in the unpublished "Queering Plato" manuscript) shows at least the possibility of even that interpretation.

### **Acknowledgments**

The author has received help and support from the sociological research seminar of Åbo Akademi University, from two other colleagues (a philosopher and a sociologist), and helpful comments from the anonymous reviewers of this journal.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Mikko Lagerspetz  <https://orcid.org/0000-0001-8639-6971>

## Notes

1. On their web page, they give the number of manuscripts as twenty. They list and describe twenty-one manuscripts, but two are said to be “rewritten” versions of other ones. However, a tweet from James Lindsay on October 10, 2018 (link cited by al-Gharbi 2018) mentions forty-eight submissions. This figure includes the early stage of the project and is not reported on the project website, where the number of submissions totals to twenty-eight.
2. Such as *Svenska Dagbladet* (October 3, 2019), *The Times* (October 4, 2019), *Helsingin Sanomat* (October 5, 2019), *Le Figaro* (October 9, 2019), *Neue Zürcher Zeitung* (October 11, 2019), *Frankfurter Allgemeine Zeitung* (October 15, 2019), and so on. These first reports quite closely reproduced the project’s own media communication.
3. Schieber (2018) was one of the reviewers targeted.

## References

- Abramowitz, Stephen I., Beverly Gomes, and Christine V. Abramowitz. 1975. “Publish or Politic: Referee Bias in Manuscript Review.” *Journal of Applied Psychology* 5 (3): 187-200.
- Afinogenov, Greg. 2018. “Orthodoxed! On ‘Sokal Squared’.” *n+1*, October 4. Accessed July 22, 2018. <https://nplusonemag.com/online-only/online-only/orthodoxed/#rf1-10307>.
- al-Gharbi, Musa. 2018, October 10. “Academic Grievance Studies and the Corruption of Scholarship.” Accessed July 26, 2019. <https://heterodoxacademy.org/academic-grievance-studies/>.
- Armstrong, J. Scott. 1997. “Peer Review for Journals: Evidence on Quality Control, Fairness, and Innovation.” *Science and Engineering Ethics* 1997 (3): 63-84.
- Baldwin, Richard. 2018. “When the Joke Is on You: A Feminist Perspective on How Positionality Influences Satire.” *Hypatia: A Journal of Feminist Philosophy* [Retracted]. Accessed October 4, 2018. [https://drive.google.com/drive/folders/1cJLr\\_o04R-zpHcMNaIWPGs7Ue\\_i-tkCw](https://drive.google.com/drive/folders/1cJLr_o04R-zpHcMNaIWPGs7Ue_i-tkCw).
- Bergstrom, Carl T. 2018. “A Hollow Exercise in Mean-spirited Mockery.” *The Chronicle of Higher Education*, October 9. Accessed August 21, 2019. <https://www.chronicle.com/article/What-the-Grievance/244753>.

- Bohannon, John. 2013. "Who's Afraid of Peer Review?" *Science* 342 (6154): 60-65.
- Bornmann, Lutz. 2011. "Scientific Peer Review." *Annual Review of Information Science and Technology* 45 (1): 199-245.
- Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2010. "A Reliability-generalization Study of Journal Peer Reviews: A Multilevel Meta-analysis of Inter-rater Reliability and Its Determinants." *PLoS One* 5 (12): e14331. Accessed July 19, 2017. <https://doi.org/10.1371/journal.pone.0014331>.
- Bourdieu, Pierre. 1975. "The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason." *Social Science Information* 14 (6): 19-47.
- Cowley, Stephen J. 2015. "How Peer-review Constrains Cognition: On the Frontline in the Knowledge Sector." *Frontiers in Psychology* 6 (2015): Article 1706.
- Dondio, Pierpaolo, Niccolò Casici, Francisco Grimaldo, Nigel Gilbert, and Flaminio Squazzoni. 2019. "The 'Invisible Hand' of Peer Review: The Implications of Author-referee Networks on Peer Review in a Scholarly Journal." *Journal of Informetrics* 13 (2019): 708-16.
- Drezner, Daniel W. 2018. "A Paper That Would Never Have Gotten Past Peer Review Criticizes the Academy." *The Washington Post*, October 4. Accessed July 16, 2019. [https://www.washingtonpost.com/outlook/2018/10/04/paper-that-would-never-have-gotten-past-peer-review-criticizes-academy-film/?noredirect&utm\\_term=.d22b3df2d11e](https://www.washingtonpost.com/outlook/2018/10/04/paper-that-would-never-have-gotten-past-peer-review-criticizes-academy-film/?noredirect&utm_term=.d22b3df2d11e).
- Engber, Daniel. 2018. "What the 'Grievance Studies' Hoax Actually Reveals." Accessed July 19, 2018. <https://slate.com/technology/2018/10/grievance-studies-hoax-not-academic-scandal.html>.
- Epstein, William M. 1990. "Confirmational Response Bias among Social Work Journals." *Science, Technology, & Human Values* 15 (1): 9-38.
- Essig, Laurie, and Sujata Moorti. 2018. "Only a Rube Would Believe Gender Studies Has Produced Nothing of Value." *The Chronicle of Higher Education*, October 9. Accessed August 21, 2019. <https://www.chronicle.com/article/What-the-Grievance/244753>.
- European Science Foundation. 2011. *The European Code of Conduct for Research Integrity*. Strasbourg, France: European Science Foundation. Accessed November 6, 2018. [http://archives.esf.org/index.php?eID=tx\\_nawsecured1&u=0&g=0&t=1541603456&hash=f46061077b0ea0af892b47bf3a39b02f6c08ae0b&file=/fileadmin/be\\_user/CEO\\_Unit/MO\\_FORA/MOFORUM\\_ResearchIntegrity/Code\\_Conduct\\_ResearchIntegrity.pdf](http://archives.esf.org/index.php?eID=tx_nawsecured1&u=0&g=0&t=1541603456&hash=f46061077b0ea0af892b47bf3a39b02f6c08ae0b&file=/fileadmin/be_user/CEO_Unit/MO_FORA/MOFORUM_ResearchIntegrity/Code_Conduct_ResearchIntegrity.pdf).
- Gergen, Kenneth J. 1999. "Agency: Social Construction and Relational Action." *Theory & Psychology* 9 (1): 113-15.

- Hegtvedt, Karen A. 2007. "Ethics and Experiments." In *Laboratory Experiments in the Social Sciences*, edited by Murray Webster, Jr., and Jane Sell, 141-73. Amsterdam, the Netherlands: Elsevier.
- Hilgartner, Stephen. 1997. "The Sokal Affair in Context." *Science, Technology, & Human Values* 22 (4): 506-22.
- Kuhar, Roman, and David Paternotte. 2017. *Anti-gender Campaigns in Europe: Mobilizing against Equality*. London, UK: Rowman & Littlefield.
- Mahoney, Michael J. 1977. "Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System." *Cognitive Therapy and Research* 1 (2): 161-75.
- Mouk, Yascha. 2018a. "The Circling of the Academic Wagons." *The Chronicle of Higher Education*, October 9. Accessed August 21, 2019. <https://www.chronicle.com/article/What-the-Grievance/244753>.
- Mouk, Yascha. 2018b. "What an Audacious Hoax Reveals About Academia." *The Atlantic*, October 5. Accessed July 19, 2019. <https://www.theatlantic.com/idea/s/archive/2018/10/new-sokal-hoax/572212/>.
- Pluckrose, Helen, James A. Lindsay, and Peter Boghossian. 2018. "Academic Grievance Studies and the Corruption of Scholarship." *Areo*, October 2. Accessed October 4, 2018. <https://areomagazine.com/2018/10/02/academic-grievance-studies-and-the-corruption-of-scholarship/>.
- Popper, Karl R. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London, UK: Routledge and Kegan Paul.
- Ross-Hellauer, Tony, Arvid Deppe, and Birgit Schmidt. 2017. "Survey on Open Peer Review: Attitudes and Experience amongst Editors, Authors and Reviewers." *PLoS One* 12 (12): e0189311. Accessed November 7, 2018. <https://doi.org/10.1371/journal.pone.0189311>.
- Schieber, David. 2018. "A Strange Start to Peer Reviewing." *The Chronicle of Higher Education*, October 9. Accessed August 21, 2019. <https://www.chronicle.com/article/What-the-Grievance/244753>.
- Schnabel, Jim. 1994. "Puck in the Laboratory: The Construction and Deconstruction of Hoaxlike Deception in Science." *Science, Technology, & Human Values* 19 (4): 459-92.
- SCImago. n.d. *SJR—SCImago Journal & Country Rank* [Portal]. Accessed October 16, 2018. <http://www.scimagojr.com>.
- Serpell, James A. 2003. "Anthropomorphism and Anthropomorphic Selection: Beyond the 'Cute Response'." *Society & Animals* 11 (1): 83-100.
- Smith, Justin E. H. 2018. "In Defense of Hoaxes." *The Chronicle of Higher Education*, October 9. Accessed August 21, 2019. <https://www.chronicle.com/article/What-the-Grievance/244753>.

- Sokal, Alan. 1996. "A Physicist Experiments with Cultural Studies." *Lingua Franca* 6 (4): 62-64.
- Stribling, Jeremy, Max Krohn, and Dan Aguayo. 2005. *SCIgen—An Automatic CS Paper Generator*. Accessed October 31, 2018. <https://pdos.csail.mit.edu/archive/scigen/>.
- Wagenknecht, Thomas. 2018. "Unhelpful, Caustic and Slow: The Academic Community Should Rethink the Way Publications are Reviewed." *LSE Impact Blog*. Accessed November 7, 2018. <http://blogs.lse.ac.uk/impactofsocialsciences/2018/06/22/unhelpful-caustic-and-slow-the-academic-community-should-rethink-the-way-publications-are-reviewed/>.
- Weinberg, Darin. 2014. *Contemporary Social Constructionism: Key themes*. Philadelphia, PA: Temple University Press.
- Wilson, Helen. 2018, May 22. "Human Reactions to Rape Culture and Queer Performativity at Urban Dog Parks in Portland, Oregon." *Gender, Place & Culture* [Retracted]. Accessed October 4, 2018. [https://drive.google.com/drive/folders/1cJLr\\_o04R-zpHcMNaIWPGs7Ue\\_i-tkCw](https://drive.google.com/drive/folders/1cJLr_o04R-zpHcMNaIWPGs7Ue_i-tkCw).

### Author Biography

**Mikko Lagerspetz** is a professor of sociology at Åbo Akademi University, Finland's Swedish-language university. Earlier, he was a professor at Tallinn University, Estonia, and the director of its Center for Civil Society Research and Development. He was the President of the Estonian Association of Sociologists in 1998-2003. His research interests include civil society, minorities, sociology of knowledge, and the postsocialist change in Central and Eastern Europe.