

Chapter 6

Falsifying



Contents

6.1	Introduction.....	118
6.1.1	Sloppy Science.....	118
6.1.2	Falsifying.....	119
6.2	Bias at the Start of the Research Process: Asking Critical Questions.....	120
6.2.1	Confirmation.....	120
6.2.2	Challenging Bias.....	121
6.3	Bending the Empirical Cycle: Manipulations During Research.....	122
6.3.1	'Lies, Damned Lies, and Statistics'.....	122
6.3.2	Questionable Research Practices.....	122
6.3.3	Image Manipulation.....	126
6.4	Bias in Disseminating Research: Publication Bias.....	128
6.4.1	File-Drawer Problem and False Positives.....	128
6.4.2	Reviewer and Editorial Bias.....	129
6.5	Self-Deception.....	131
6.5.1	Mistakes Happen.....	131
6.5.2	To Remember or Not to Remember (That Is the Question).....	133
6.6	Science's Self-Correction.....	134
6.6.1	Self-Correction.....	134
6.6.2	Beyond Retraction?.....	135
6.7	Conclusions.....	136
6.7.1	Summary.....	136
6.7.2	Discussion.....	137
	Case Study: Yanomami Violence and the Ethics of Anthropological Practices.....	137
	Assignment.....	140
	Case Study: Fraud or Fiction? Diary of a Teenage Girl.....	140
	Assignment.....	143
	Suggested Reading.....	143
	References.....	144

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-48415-6_6) contains supplementary material, which is available to authorized users.

After Reading This Chapter, You Will:

- Know exactly what falsifying is
- Be able to distinguish falsification from other forms of fraud
- Understand how falsifying impacts the social sciences
- Develop strategies to address falsification

Keywords Conformation bias · Deep data diving · Dichotomization · Disconfirmation dilemma · Editorial bias · False positive · Falsification · Falsifying · File drawer problem · Image manipulation · Impact factor · HARKing · My side bias · Outliers · Pathological science · Peer review · p-hacking · Post-publication · Protheus phenomenon · Publication ethics · Publication bias · Questionable research practice · Replication · Retraction · Reviewer bias · Self-correction · Self-deception · Sloppy science · Submission bias

6.1 Introduction

6.1.1 Sloppy Science

In 2012, Anthropologist Mart Bax of Free University Amsterdam was already in retirement when suspicion arose about the validity of his field work, some of which dated back to the 1970s and 80s. It was thought that if he hadn't fabricated his research outright, then at the very least he had manipulated it. That is to say, he was accused of (among other things) having altered and even removed crucial details in his data. He had ascribed statements to untraceable participants and staged actions that could not be verified.

An integrity commission was set to work and concluded the following year that Bax was guilty of scientific misconduct. He had presented 'improbable events as "historical facts," embedded in research that systematically obscures names of persons and places and muzzles sources and contains inaccuracies in a large number of places' (Baud et al. 2013, p. 39).

The lack of openness and transparency Bax exhibited has become an exemplar of what we now call 'sloppy science' – carefree and negligent research practices that include both intended and unintended violations of scientific norms. In this case, the researcher seemed to have placed little value into verifiability and transparency, which sit as cornerstones of appropriate scientific practice.

When sloppy research veers into falsehoods, we speak of 'falsifying' or 'falsification.' To avoid confusion with the terminology used by philosopher Karl Popper (see below), we stick to 'falsifying.' Falsification in Popper's sense means actively seeking to *disconfirm* a hypothesis, falsifying effectively amounts to the opposite.

6.1.2 *Falsifying*

The term falsifying, literally meaning ‘rendering false,’ entails forms of manipulation that allow researchers to use a dataset that supports biased or even erroneous claims. It includes ‘trimming’ (leaving out certain findings) and ‘massaging’ (slightly changing) data, as well as altering images, misrepresenting results, and simply *not* reporting findings.

If fabrication (presenting fake or non-existent research data) and plagiarism (literary theft) are sciences’ deadly sins, punishable by severe penalties, then falsifying is its *daily* sin. It is a less visible, less spectacular form of misconduct, and the scientific community tends to view it more tolerably than its lethal counterparts. Unjustly so, says Köbben (2012), who warns that the accumulation of these smaller sins represents a far greater danger to science than the (isolated) larger ones. Overtime, and left unchecked, this accumulation will result in the large-scale pollution of scientific research.

This being said, some considerations must be first addressed. Not all manipulations represent a researcher’s intent to deceive, nor is every form of manipulation prohibited, as we shall see below. Also, one must distinguish between deliberate manipulations and honest errors. Furthermore, these two should not be confused with scientific disagreement (researchers challenging the conclusions of one another). Thus, although falsifying is considered ‘misconduct,’ it can be difficult to assess exactly *when* acceptable research practices lapse into dubious ones. It is on this note we enter into the heart of the problem of academic fraud – which is less about demarcating right from wrong, and more about ethical reflection and decision-making. The aim of this chapter is to raise awareness of these issues by exploring several dimensions of falsifying in research practices.

By and large, in this chapter we follow the research process itself. We start with the forms of bias that appear at the first stage of the process, when research questions are posed. Following the selection of questions, a discussion will take place regarding the falsities that result from (slight) alterations, or the act of manipulation during data collection and analysis.

We finish with a discussion of the biases often present when research conclusions are reported and disseminated in a skewed or one-sided way. Though this is referred to as *publication ethics*, it relates to our subject of research ethics because it discloses disturbances in the research process.

In two separate sections, we discuss the problem of *self-deception* (falsifying by not being critical enough) and a possible remedy against the falsehoods inherent to science.

6.2 Bias at the Start of the Research Process: Asking Critical Questions

6.2.1 Confirmation

Research starts with asking questions. But, as any student knows, asking good questions demands a self-critical attitude, and a readiness to address and counter one's own preconceptions. In fact, scientists should actively look for information that disconfirms their opinions about the world, an action for which the term 'falsification' is reserved (as explained above).

The reality is, this is much more difficult than it appears. There is a long-identified experimental effect known as *confirmation bias* (sometimes called *myside bias*, see; Perkins 1985; Toplak and Stanovich 2003). Confirmation bias consists of the tendency of individuals to judge new information in a way consistent with their preexisting ideas or convictions. People thus prefer supporting information rather than conflicting information and tend to overlook or disregard information that does not fit into their worldview (Jonas et al. 2001).

Notorious examples are found in 'psychic studies' (studies into paranormal activities) and psychoanalysis (studies into the unconscious mind). In both traditions, there exists a strong tendency toward confirming what was theoretically hypothesized. But it is far from restricted to just these domains and has been observed in more empirically oriented research traditions a well.

As a case in point, Greenwald et al. (1986) examined empirical research into a phenomenon known as the *sleeper effect*. This is the counter-intuitive finding that a persuasive message accompanied by a 'discounting cue' (a prompt that indicates the message is untrustworthy) tends to develop more impact over time. For example, viewers watching a 'smear campaign' against one political candidate, paid for by the opposite candidate, will develop a more favorable attitude towards the message weeks *after* being exposed to the message, rather than immediately afterwards, despite being aware that the source is biased.

To explain this effect, it was hypothesized that over time, the discounting cue becomes dissociated from the original message, and therefore ceases to be effective in countering it (this is known as the *dissociation hypothesis*). Research into the sleeper effect has not been able to confirm this hypothesis, however that did not deter the researchers from investigating it. Only much later did researchers realize that the 'dissociation hypothesis' was incorrect, and that an entirely different explanation was required. This fixation on a single hypothesis, and the resulting neglect of alternative theories has obstructed scientific understanding on the subject for some 25 years, Greenwald et al. observed.

Box 6.1: ‘Disconfirmation Dilemma’

How are we to deal with disconfirmation in the empirical process? If the results of an experiment don’t confirm theoretical expectations, researchers are confronted with what Greenwald et al. call a ‘disconfirmation dilemma.’ Researchers can decide to either (a) reanalyze the data, (b) revise the procedures, or (c) reformulate a different prediction (based on the same theory). Rarely do researchers decide for option (d), to publish disconfirming results.

When researchers resolve the disconfirmation dilemma by repeatedly retesting predictions, instead of reporting disconfirmation, they may be accused of some form of ‘falsifying’ because they resort to theory-confirmation rather than theory testing (Fig. 6.1).

6.2.2 Challenging Bias

Effectively, confirmation bias undermines open and critical thinking and runs against creativity. This poses a serious challenge for science. Open-mindedness and creativity are two of science’s most crucial features. How can scientists avoid or

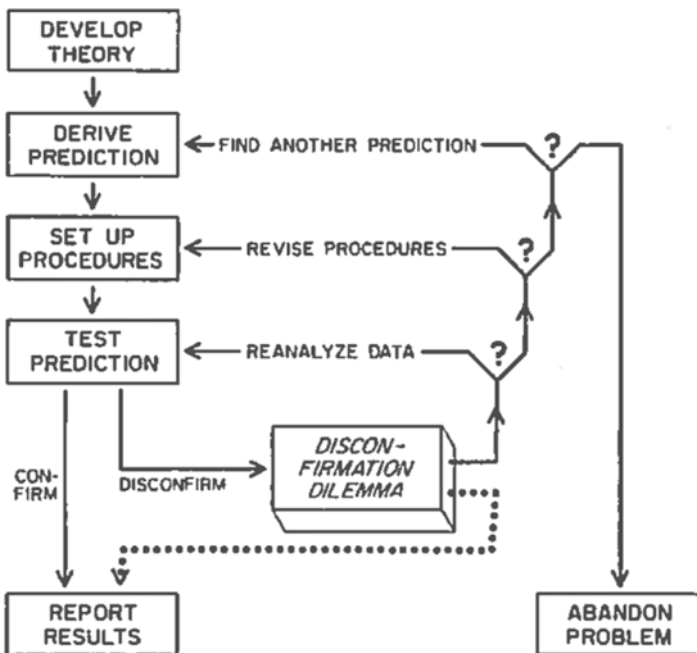


Fig. 6.1 Disconfirmation Dilemma. (Adopted from Greenwald et al. 1986, p. 220. The dotted line represents a route infrequently taken)

counter this type of bias? Is it possible, for example, to train scientists to consider *both* sides of an argument? There is evidence that this may be possible, at least to a degree.

Wolfe and Britt (2008) found that when students were assigned to one side of a (somewhat controversial) topic and received instructions to search for as much information on the topic as they saw fit, they would display confirmation bias. However, when instructed to search specifically for balanced information, confirmation bias was significantly reduced. Similarly, Macpherson and Stanovich (2007) found that *decontextualization instructions* (instructions to put aside one's own convictions and consider the issue from opposite sides) helped reduce confirmation bias. These findings, preliminary as they are, point to the importance of making explicit one's expectations (Box 6.1).

6.3 Bending the Empirical Cycle: Manipulations During Research

6.3.1 '*Lies, Damned Lies, and Statistics*'

The next step in the research process consists of setting up a design in order to test hypotheses. In the social sciences, significance testing of null hypotheses is an omnipresent tool, and will be the focus of the next three sections.

The simplest example of significance testing is the evaluation of null hypothesis $\mu_E = \mu_C$ against the alternative hypothesis $\mu_E \neq \mu_C$. Here μ_E denotes the mean of the outcome variable in an experimental group, and μ_C the mean of the outcome variable in the control group. If the p-value for testing the two hypotheses against each other is smaller than .05, the null hypothesis is rejected. If it is larger than .05, then it means there is no significant difference between the two groups, and the null hypothesis is accepted on account that no evidence of an experimental effect was found.

Scientific journals have long tended to only publish results that have shown the experimental condition to be 'effective,' that is, if the p-value is smaller than .05. It is therefore crucial for researchers to obtain low p-values, otherwise their effort, time, and money is wasted. For some, obtaining small p-values has become a goal in itself. This raises some ethical questions which will be explored below.

6.3.2 *Questionable Research Practices*

Can research outcomes be manipulated such that lower p-values are obtained? The answer is yes, for example, by removing so-called 'outliers.' Outliers are extreme scores, and removing them heightens the chance of getting significant results. It sounds like cheating but that need not be the case, there can be good reason to

remove outliers. Outliers can result from data errors (incorrectly recorded data) or because respondents may have failed to understand their role, or the questions asked. For instance, one survey gathered data on nurses' hourly wages. While on average respondents reported to earn \$12.00 an hour, with a standard deviation of \$2.00, one nurse reported an hourly wage of \$42,000.00, which was clearly erroneous (it was more likely their annual income). Not removing this number would influence the true outcome (see Osborne and Overbay 2004).

However, consider the case of Dirk Smeesters, professor of consumer behavior and society at the Rotterdam School of Management. His work attracted the attention of fellow researcher Uri Simonsohn from Wharton University in 2011. He had read some of Smeesters' work and suspected foul play. Simonsohn believed Smeesters' studies were 'too clean to be the result of random sampling' (quoted in Chamber 2017, p. 81). He requested and obtained Smeesters' dataset and discovered anomalies. It seemed that Smeesters had removed participants from his data when they led his hypotheses toward not being confirmed. Smeesters responded that the participants 'had not understood the instructions' (quoted in Kolfshoeten 2012, p. 270).

This did not satisfy Simonsohn. An integrity commission investigated the case and ruled that this reversal of logic, by which outliers are removed to boost significance, should be understood as 'data massaging.' Smeesters confirmed that he had acted 'erroneously' but denied that he had committed fraud: 'What I have done was to give a study, which was already almost good, a push in the right direction' (Kolfshoeten 2012, p. 270). That didn't help his case. Seven papers he co-authored were retracted and Smeesters resigned from his position in 2012.

What Smeesters engaged in are called *questionable research practices*, or QRPs for short. QRPs have become serious concerns in the academic community. Simmons (2011) noted that 'flexibility in data collection and analysis allows researchers to present almost anything as significant.'

QRPs take many shapes and forms. To name a few; failing to report all dependent measures, selective reporting (only submitting studies that were successful), and excluding data after looking at the impact (as Smeesters had done).

Evidence of QRPs on a large-scale were found by Masicampo and Lalande (2012). They collected the reported p-values from three high-level psychological journals and compared their distribution. Given that smaller p-values are more appreciated, one would expect to see a steady decline in reports with larger p-values. What they found instead was a steady decline, followed by a peculiar peak of p-values just below .05 (see Fig. 6.2).

Many take this as evidence for the existence of 'falsifying,' because it appears that researchers have manipulated their data to ensure their data falls within an acceptable p-value of below .05.

Further qualitative evidence of QRPs on a large-scale was found by Leslie John and his collaborators. John et al. (2012) surveyed over two thousand psychologists and found a majority of psychologists admitted to engaging in a variety of such behaviors. In their widely circulated article, it was estimated that some questionable

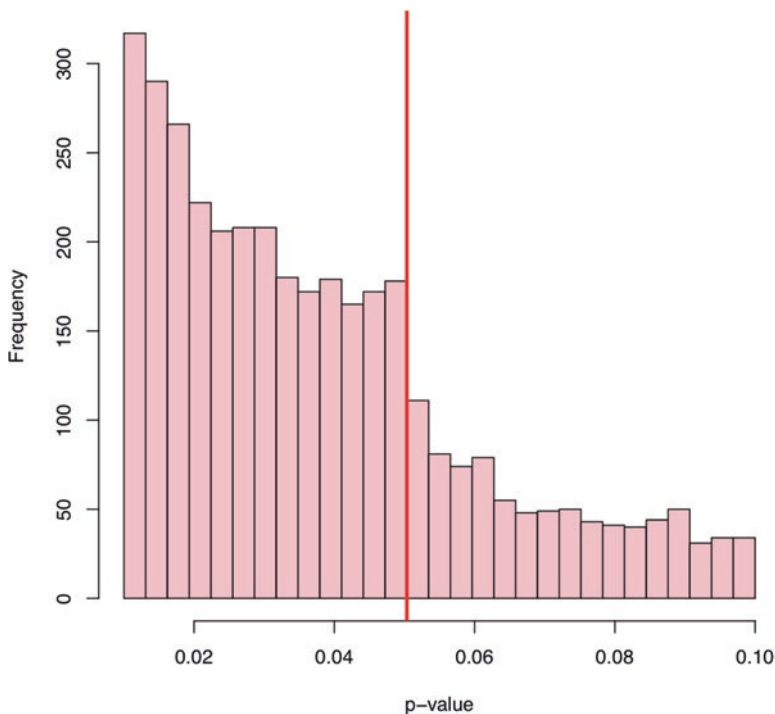


Fig. 6.2 ‘A peculiar prevalence of p-values just below .05’. Figure by Larry Wasserman, based on the data collected by Masicampo and Lalande (2012). Used with permission from the author. (Source: Normal Deviate, entry August 16, 2012)

research practices are so widespread that it must be assumed that virtually *everyone* uses them.

This raises the question as to whether these practices constitute a new scientific norm (John et al. 2012). Discussing this controversial conclusion, Fiedler and Schwarz (2016) warn against an inflation in the usage of the term QRPs precisely because of the suggestion of normalization. Some of the reported practices, they argue, are merely ambiguous, not ‘questionable,’ while others may or may not be justifiable depending on the specifics of the case (p. 50).

We do not propose that QRPs are the ‘new norm.’ On the contrary, there is a serious danger that the scientific literature becomes polluted with ‘breakthroughs’ (significant findings) which are not breakthroughs at all. Indeed, in the June 1st, 2011 issue of *Scientific American*, John Ioannidis argues that exaggerated results in peer-reviewed scientific studies have reached ‘epidemic proportions’ in recent years (Box 6.2).

Box 6.2: ‘P-hacking and HARKing’

The following case, discussed in several entries on the critical website *Retraction Watch* in 2017, provides a rare glimpse into how the academic community responds to questionable research practices.

Several years ago Brian Wansink, a world-renowned food researcher at Cornell University, provided a visiting PhD student with the complete set of data of a self-funded study that failed to produce any notable results. He told the student that it is well worth the effort to search for overlooked patterns, further stating that ‘there’s got to be something here we can salvage because it’s a cool (rich & unique) data set.’ The student set to work and managed to produce five articles in just six months using the dataset. In a now deleted blog post of November 2016 (‘The Grad Student Who Never Said “No”’), published on his personal website, Wansink proudly reported on this student’s success, presenting it as a ‘lesson in productivity’.

His readers were less impressed. ‘This is a great piece that perfectly sums up the perverse incentives that create bad science. I’d eat my hat if any of those findings could be reproduced in preregistered replication studies.’ Another reader commented, saying that what was described in the blog sounded suspiciously like p-hacking and HARKing (entry at *Retraction Watch*, 2.2.2017).

P-hacking (also called ‘phishing’) is a term used to describe how researchers try to uncover statistically significant patterns in a data set without having a specific hypothesis. They just hope to find statistically significant results. *HARKing* is the flipside of this coin (HARK stands for Hypothesizing After Results are Known). It consists of presenting a post hoc hypothesis in a research report as if it were, in fact, an ‘a priori’ (earlier formulated) hypothesis.

Had Wansink been ‘bending the rules of the game’ by letting his student go through raw data in the hopes of unearthing something (*anything*), which then would be presented as a ‘finding’?

When confronted with the accusation of p-hacking, Wansink retorted that testing the null hypothesis had been his ‘plan A.’ It was when he didn’t find anything that he turned to ‘plan B.’ As Wansink explained: ‘P-hacking shouldn’t be confused with deep data dives – with figuring out why our results don’t look as perfect as we want. With field studies, hypotheses usually don’t “come out” on the first data run. But instead of dropping the study, a person contributes more to science by figuring out when the hypo worked and when it didn’t. This is Plan B’ (quoted in an entry on *Retraction Watch*, 2.2.2017).

Wansink’s rebuttal failed to convince readers of *Retraction Watch*. One wrote: ‘Deep dives are great, but they should be planned when the study is being constructed, not created after the fact in an attempt to “salvage” something from the experience’ (2.2.2017). Another sarcastically remarked: ‘Wansink’s use of the phrases “our results don’t look as perfect as we want” [...] pretty much speaks for itself’ (3.2.2017).

(continued)

Box 6.2 (continued)

However, not everyone saw wrongdoing. Another reader wrote in Wansink's defense, exclaiming that not all exploratory studies constitute 'p-hacking': '[There] is nothing wrong with a researcher honestly engaging in and presenting an exploratory analysis without a single pre-defined hypothesis. As long as these studies are presented honestly, they can provide useful insights and generate useful hypotheses that can later be verified (or debunked) through attempts to replicate, often by other researchers in the field' (19.2.2017).

The result of Wansink's actions? – By the summer of 2019, when this chapter was written, 17 of his papers were retracted (one even twice) and he resigned from his position. (See Resnick and Bellus 2018, for further discussion.) (Fig. 6.3)

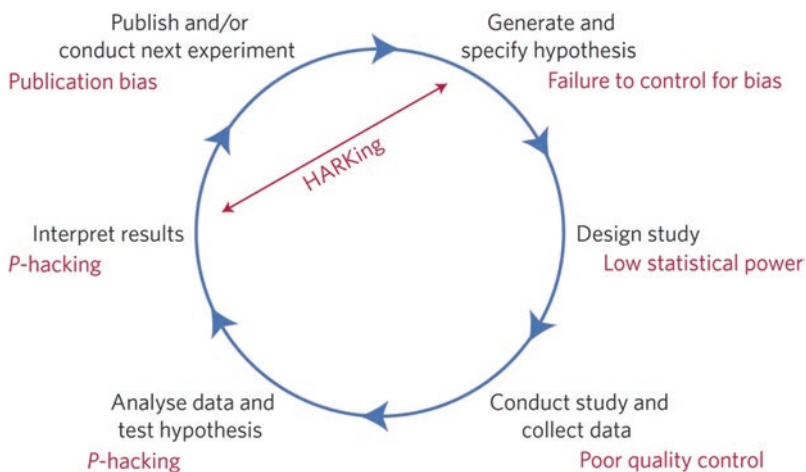


Fig. 6.3 Threats to reproducible science. (Adapted from Munafò et al. 2017)

6.3.3 Image Manipulation

Recent technological advances provide researchers with a wealth of opportunities for furthering their research in ways not available fifteen or twenty years ago. But as with any change, new ethical considerations emerge. In the case of digital images, there is a growing concern in the scientific community over how to properly handle them.

Today, there are multiple known cases of unethical manipulation of images that affected the interpretation of the data presented, a number of them having led to retractions. One such case, a 2009 paper published in the *Journal of Biological Chemistry* by Spanish researcher José G. Castaño, was retracted because (quoting from the notice published September 9th 2016 on *Retraction Watch*) ‘the same image was used to represent results of different experimental conditions’ on multiple occasions, adding further that the ‘background of one image had inappropriately been adjusted.’

A lack of awareness of what is considered an acceptable form of image manipulation calls for the creation of guidelines to help researchers distinguish between appropriate and inappropriate use of digital images. A set of such general principles is discussed by Cromey (2010), who compares these guidelines to what is already established practice in the field of photojournalism. A sampling of these guiding principles includes the following recommendations:

- Digital images should be acquired in a manner that does not intend to deceive the viewer or to obscure important information.
- Manipulation of digital images should be performed only on a copy of the image.
- Simple adjustments and cropping are acceptable but lossy (irreversible) compression should be avoided, and use of software filters to improve image quality is not recommended.
- Cloning or copying objects into a digital image is considered highly questionable (Box 6.3).

Box 6.3: ‘Consequences of Retraction’

With an increased awareness of research ethics in our day and age comes an increased awareness of the consequences of misconduct. We quote from an anonymous cry for help, posted on October 16th, 2013 on ‘Editage Insights’ (a platform for researchers, authors, publishers and academic societies): ‘I recently got an email from the editor of a journal in which my paper is published, requesting me to retract the paper because they found some errors in my data and statistical analysis. I am worried about my reputation if I have a retracted paper. I may not get a grant for my next study. Please advise me.’

In a response posted on March 30th, 2017: ‘I would encourage you to respond positively to the journal editor’s request and offer to have your paper retracted. If you do so, the journal’s retraction notice will inform readers that the paper has been retracted by agreement among the authors and the journal editor, owing to errors in data.’ (source: Editage Q&A).

6.4 Bias in Disseminating Research: Publication Bias

6.4.1 *File-Drawer Problem and False Positives*

Imagine a researcher testing the effects of the new and promising treatment ‘X’ (say a particular form of cognitive behavioral therapy for a certain type of anxiety disorder). To the researcher’s disappointment, a comparison between the experimental group who received this form of therapy and a control group who received no therapy (or a different therapy), resulted in a p-value of .14. Since this is larger than .05, the hypothesis is rejected, and the results are not published. Another researcher (unaware of the first researcher’s work because it was not published) is interested in the same therapy. Their comparison results in a p-value of .37, and again the results are not published. Shortly thereafter, a third researcher evaluates treatment ‘X’. They find a p-value of .02, and as a result the findings are published.

Based on this one publication by researcher No. 3, the unsuspecting reader will conclude there is evidence for the effectiveness of treatment ‘X’. In reality, the effectiveness of treatment ‘X’ is a *false positive* because there is actually more evidence to the contrary – that it doesn’t work. The problem being, that evidence was never published. Our full understanding is obscured by the fact that studies that show no result are not published. Robert Rosenthal coined a term for this; the *file-drawer problem*.

‘False positives’ and the complementary ‘file-drawer problem’ relate not so much to theoretical or methodological issues in research, but to questions regarding *dissemination* (communication or non-communication of research findings), related to *publication ethics*.

If there are frequent false positives and/or numerous unpublished null results, any meta-analysis of a particular research subject will turn up corrupted. How serious is this problem?

One way of researching this question is by comparing the amount of research undertaken with the number of publications stemming from that research. In many fields of research, pre-registration is required (the researcher must catalogue his research protocol in advance, submitting hypotheses, methodology, and expected findings). This makes it easier to check for both p-hacking and HARKing, but it also allows for the questioning of *submission bias* (the tendency to only submit for publication studies that have ‘positive findings’). In the social sciences though, pre-registration is a recent phenomenon and not the norm, with a few exceptions.

One exception is the public registry TESS (Time-Sharing Experiments in the Social Science). Franco et al. (2014) followed studies registered in TESS over a ten-year period, to see how many of them were eventually published in peer-reviewed journals. It turned out that 80% of the registered studies were written up, but less than half (48%) were published. Unsurprisingly, there proved to be a strong relationship between the outcome of the study (whether or not the hypothesis was supported by the results) and it being published. Studies that had negative results were far less likely to be published, and even less likely to be written up at all.

Why do researchers opt to not write up ‘null results’? Franco et al. (2014) questioned a selective group of researchers by email and got answers that confirmed their suspicion. As one of their respondents reported: ‘I think this is an interesting null finding, but given the discipline’s strong preference for $p < 0.05$, I haven’t moved forward with it.’ (p. 1504).

6.4.2 *Reviewer and Editorial Bias*

Another source of publication bias is located in the process of peer review and journal editorship. Both peer reviewers and journal editors effectively function as gatekeepers, deciding whether an article is worthy of publication. We discuss both roles below, starting with peer reviewers.

The ‘peers’ in the peer review process are researchers themselves, often experts in a field from which they are recruited. They are asked to assess the quality of manuscripts sent to a journal. The review procedure in which they participate is as a rule blind. That is to say, the author is not aware of the reviewer’s identity (single blind), but often the reviewer doesn’t know the identity of the author either (double blind). Peer reviewers do not get paid for their work, should have no interests involved, be unconnected to the authors, and act solely on the desire to guarantee objectivity and impartiality in science.

But does it really work like that? Some argue that peer review is indeed the best system that we have, providing impartial quality control. Others contend it may have functioned as such at one point but it longer does in today’s society, where science cannot permit the luxury of operating from an ivory tower any longer (we turn to this discussion in greater detail in Chap. 9). And then there are those who argue that the peer review system has never guaranteed quality control in the first place.

They point to the fact that some of the most important and groundbreaking works in the history of science were never peer reviewed, that some of these work were initially rejected by peer reviewers, and that vice versa, flawed, non-sensical or even absurd papers were accepted by them (see Box 5.7 on ‘hoaxing’).

If we cannot rely on peer reviewers to detect errors, identify misconduct, or spot what is truly innovative, then the peer review system does not safeguard quality. But perhaps it is even worse. There is reason to believe that the peer review system is biased in at least two ways.

1. Peer reviewers may be too conservative. Reviewers are believed to be biased against new findings and new ideas. Also, they focus too much on finding weaknesses in manuscripts and not on the positive contributions therein. Suls and Martin (2009, p. 43) argue this may be so because ‘appearing to be too lenient seems worse than appearing to be too harsh.’
2. Reviewers seem prejudiced in favor of prestigious research institutions and established authors. Peters and Ceci (1982) found confirmation of this suspicion in a small-scale study they performed. They selected 12 previously published

articles, originally written by researchers from prestigious American psychology departments, and resubmitted them to the same (highly prestigious) journals that had previously published them, but under fictitious names and fictitious institutions. Only three articles were detected as ‘resubmissions’; eight out of the nine remaining articles were rejected on ground of insufficient quality (some were even critiqued for having ‘serious methodological flaws’).

Consider next the role of journal editors. As gatekeepers, editors not only have to safeguard quality, but also to present interesting, original, and novel findings to their readership. This may lead to a bias against replication studies because they don’t offer anything new, despite replication being the ‘gold standard’ in science (see Møller and Jennions 2001). Kerr, and later Roney and Zenisek (quoted in Hubbard and Armstrong 1994) conducted a survey among editors and review board members of both management and psychology journals, and indeed found confirmation of *editorial bias* against replication studies.

The editor’s obligation to present novel and interesting results can furthermore lead to an effect known as the *proteus phenomenon*. In essence, whenever positive results are published, a window of opportunity quickly opens for researchers to publish findings that contradict these results. Editors often publish these findings because they too are a ‘novelty.’ The net effect is a tendency for journals to rapidly publish conflicting results (see Pfeiffer et al. 2011) (Box 6.4).

Box 6.4: ‘Adjusting the Data? A Dilemma’

On the ‘r/AskAcademia’ Reddit community, a student identified as ‘Throwinbin’ (henceforth ‘T’) published a telling post. T referenced a supervisor who requested that they make use of a dataset in which a specific approach was implied. The problem was, the data (provided by a third party) did not fit the format that the approach required. In T’s words: ‘Using [my supervisor’s] method will involve removing whole articles from our data set and changing the (important, central) main attribute of the set. It’s basically massaging the data until it fits the model of his method. I’m not comfortable doing this as I strongly believe that it’s going to give us false results [...]. How to raise this with [my supervisor] without sounding really bad?’

Below are the (edited) exchanges between T and three community members who responded to the post. From these exchanges, it becomes clear that ‘data massaging’ is not the only issue at stake, and a number of other ethical dimensions are in play. Consider how T managed the situation:

Respondent A: ‘Is there a person in your department you could consult (with no stake in your publications)?’

T: ‘I have a “second supervisor;” but I’m not keen to take this concern elsewhere in the department at the moment. I like my supervisor and I don’t want to harm his career.’

(continued)

Box 6.4 (continued)

Respondent B: ‘If you do want to approach your supervisor, then ask how to organize the data to make them fit (rather than the approach of saying it is wrong), and maybe he will explain something you hadn’t thought about.’

T: ‘I’ve done this – which is how I now have in emails him telling me to remove certain rows from the data, and later reorganize whole columns without making sure the changes carry over the entire dataset. I’ve made the changes he asked for and ran his method and it looks pants [not good].’

Respondent C: ‘Could you not go down the ‘play dumb’ route? – “I’m confused, maybe I’m just stupid, I’m not sure how your [method] is entirely relevant. Can you explain how it’s better than x”?’

T: ‘I do understand his need for this method well; he wants to move on from our current institution and this will look good on a CV. I sympathize with him as I’m not very happy with the situation in our department either and would be looking to move on if I could. I’ve admitted defeat and made our data work his model. Results so far are rubbish, so I’m going to take it to him and put the ball in his court – though if he insists on it going into a paper I don’t want my name anywhere near it.’

You: What advice would you give ‘T’?

6.5 Self-Deception

6.5.1 Mistakes Happen

Earlier we noted that various forms of falsifying must not be confused with ‘honest mistakes.’ But there is one type of ‘honest mistake’ that *should* be considered a form of falsifying, even if the researcher had no intention to mislead. This is *self-deception*. Self-deception occurs when the researcher is so strongly convinced that a particular model or theory is correct that they are unable to accept evidence to the contrary. We will discuss a few forms of self-deception from here.

Perhaps the strongest form of self-deception consists of discovering information that doesn’t exist, a phenomenon ironically dubbed *pathological science*. The discovery of so called ‘N-rays’ (an alternative to X-rays) by French physicist René-Prosper Blondlot in 1903 counts as one such example (Grant 2008, pp. 88–89). Blondlot built a device that enabled him to ‘see’ these rays. He gave demonstrations and others, if trained properly (or told what to look for), would see them too. The non-existence of N-ray was exposed when a sceptic secretly turned off the device and Blondlot still claimed to ‘see’ the rays.

The biography of Wilhelm Reich, a former Freudian who had gone astray (Sharaf 1983) offers a similar story. Reich was convinced he had discovered a new form of energy, which he called ‘orgone.’ He built a device in which orgone would accumulate, and while testing his device, he found a constant temperature difference of 2 °C inside the ‘orgone accumulator.’ Believing this to prove the validity of his discovery beyond a reasonable doubt, he contacted Einstein, who kindly agreed to study his device. Two weeks later, Reich received word from Einstein, who stated that his assistant had come up with a simpler explanation for the temperature difference – lack of air circulation. Reich was unswayed and maintained faith in his ‘discovery.’

In the field of parapsychology, Alfred Russell Wallace, a British naturalist and the co-discoverer of natural selection, offers another interesting example. During his investigations, Wallace accepted certain observations as evidence for the existence of extra-perceptual phenomena. In his autobiography, he wrote about his conversion to ‘Spiritism’ after having attended a series of séances with a ‘medium’: ‘I was so thorough and confirmed a materialist that I could not at that time find a place in my mind for the conception of spiritual existence, or for any other agencies in the universe than matter and force. Facts, however, are stubborn things. [...] Facts became more and more assured, more and more varied, more and more removed from anything that modern science taught, or modern philosophy speculated on. The facts beat me’ (quoted in Shemer 2002, p. 192).

The irony of ‘facts’ having ‘beaten’ Wallace is probably not lost on the reader, for the séances were in reality very likely carefully orchestrated performances of frauds. Wallace himself, however, was not a fraud – he was taken in by the performance (Fig. 6.4).



Fig. 6.4 Carried away by self-deception

6.5.2 *To Remember or Not to Remember (That Is the Question)*

The above discussed cases of self-deception may be comical illustration of how scientists were able to fool themselves in the past. Yet the question arises whether we can be sure that some of our own present day discoveries are not also instances of self-deception.

In the last two decades of the twentieth century, a debate emerged within psychology over whether or not a repressed childhood memory of sexual assault could be recovered through the aid of specific forms of therapy (Pezdek and Banks 1996). Proponents of recovered memory therapy argue that children who go through such assaults ‘dissociate’; meaning they repress all memories of such traumatic experiences and will not remember them, unless aided in some way.

Recovered memory therapy became prominent in the 1990s. Certain recipients of the therapy recalled highly bizarre satanic-abuse memories. In some cases, these testimonies led to prison sentences for men accused of these crimes. However, it turned out that at least some of these accusations were false and the ‘perpetrators’ were released. This prompted critics to question the validity of recalled memories (for a full discussion see Loftus and Ketcham 1994).

Did the ‘recovered memory movement’ find in the testimonies of their clients what they *wanted* to hear, or had they unearthed a new phenomenon which mainstream science *refused to accept* because it was too controversial? (Box 6.5)

Box 6.5: ‘Not Sure If It’s Research Misconduct’

A PhD student turned February 2020 to ‘r/AskAcademia’ discussion platform on the website *Reddit*, for advice, writing how that it seemed as if a professor in had been engaged in research misconduct, and considered taking it to the board.

When I opened the file [of my professor], I noticed that a lot of the figures have been altered. In several cases he took a bar graph (looks like a screenshot from a Prism file) and then covered one of the bars with a different bar. The new bar would have a different height and number of significance asterisks than the original one. I could move over the replacement bar and see the original figure – the replacement was clearly cropped out of a different figure and pasted onto this one. [...] I talked to a few other students about this. They think it’s possible he’s just lazy or doesn’t know how to use Prism. Like maybe his students repeated an experiment and that changed the results, but he didn’t want to (or know how to) update the figure in Prism so he just pasted the new bar on top? This seems sketchy to me, and I don’t think it explains the difference from his published figure either. [...] I’m hesitant to ask him directly. If he actually is falsifying data, it’s not like he’s going to admit it to me. I’d prefer to speak with the department chair and see what she recommends.

(continued)

Box 6.5 (continued)

However my classmates think going over his head without first asking for an explanation would be wrong. I'm really at a loss for what to do.

Here are some of the replies this PhD student received:

1. I think it's quite weird that you are worried that it may reflect badly on you if you ask him directly. Yet, you think that the more drastic approach of going to the chair is less worrying.
2. You don't have evidence of misconduct and there are only downsides to yourself from making accusations. I'd say forget about it.
3. I think a good and non-accusatory way to go about it is (if initially via email): 'Hi X, I've noticed that there are revisions to the graphs in the PowerPoint. Did you happen to obtain more evidence/data changing the original graphs and supporting your conclusions? If so, what steps or projects are you pursuing after the new information?'

Which of these advices do you prefer? Or would you consider a different approach?

Source: Reddit, AskAcademia, 'Not Sure if What I'm Seeing is Research Misconduct?'

6.6 Science's Self-Correction

6.6.1 Self-Correction

Discoveries of falsehoods in research are traditionally met with a defensive system of self-correction known as *retractions*. This quite simply means that a 'contaminated' publication is flagged but not withdrawn from the public domain. A note is attached to the paper that states it has been 'retracted.' Retraction can take place with or without the author's consent and can be argued on the grounds of methodological or theoretical flaws, or because research misconduct was identified.

Retractions are not to be taken lightly. Until quite recently, misconduct and subsequent retraction of a publication remained an internal matter, known to only a few parties. However, with the rise of digital publications, retractions have become much more public (and visible). For example, the website *Retraction Watch* is dedicated exclusively to highlighting misconduct, fraud, and retractions in science (across all disciplines). It keeps track of virtually all that is going on in the academic world in a very public manner, posting the full names and affiliations of all parties involved. A retracted article, though 'withdrawn,' not only remains visible, it effectively becomes a permanent stain on an author's reputation (see Box 6.3 for an impression of this consequence).

Apart from the personal consequences, there remains the question of what damage fraudulent articles can cause. After all, an undetected (unflagged) fraudulent paper will remain in the public domain, continuing to act as a source of pollution in future literature. This is important to consider, as a great deal of time may pass before a fraudulent study is retracted. Interestingly, in the last few decades, the retraction process has quickened, with the number of retracted papers increasing in lock step. From this, three questions can be raised: (1) Is this increase a good thing or not, and how to account for it?; (2) Are retractions the right answer to the problem of QRPs?; and (3) Are there better alternatives? We will very briefly touch on these issues in the sections to come.

6.6.2 *Beyond Retraction?*

In an often-cited article, Daniele Fanelli (2013) investigates retractions in scientific literature. Scouring through data from the Web of Science (a publisher-independent global citation database) for the entire twentieth century, Fanelli notes a sudden increase in retracted papers per year since the 1980s of some 20%. He then compared this increase to the number of 'corrections' applied to articles in the same period, which did not see a similar increase.

Fanelli proposed two hypotheses, that both have a radically different outlook on the question of whether or not the increase in retractions signifies a positive development. One attaches the increase to growing misconduct within the academic community, and thus sees the growing number of retractions as a bad sign. The other states that the system has become more resilient, and thus the increased number of retractions signifies something good.

Fanelli argues that the evidence in his study suggests that the 'stronger system hypothesis' is more likely to explain the rise in retractions than the hypothesis that scientists have become more fraudulent. Peers, editors, and the scientific community at large seem to have become more sensitive to and aware of misconduct, and consequently, have become more proactive about it (see furthermore Ioannidis 2012; Fanelli 2018).

This begs the question, even if editors have become more aware of the issue, can we trust that science will be able to rectify (all of) its mistakes this way? Stroebe et al., reviewing a number of recent examples of misconduct, are not overly optimistic. Science is based on trust, they argue, and as such 'scientists do not expect their colleagues to falsify their data, and do not look for signs of fraud' (2012, p. 680). What would really help, Stroebe et al. argue, is to fortify the position of the whistleblowers, who, after all, have been responsible for detecting the majority of falsities in the first place.

Furthering this line of thinking, consider *Post-publication Peer Review* (PPPR). PPPR is a commenting system that allows publications to be reviewed and discussed online, on platforms such as *PubPeer* and *Open Review* after they have been published – on a (mostly) permanent basis.

Appraising this approach, Jaime Teixeira da Silva (2015, p. 37) considers that the advantage of PPPR is that it ‘makes authors, editors, peers, journals and publishers accountable for what they have published or approved of publishing in the framework of their publishing models.’

However, the question is whether PPPR should consist of anonymous reviews (comparable with traditional peer review) or not. Teixeira da Silva is a vocal opponent of anonymity in peer review and a severe critic of *PubPeer*, which publishes anonymous reviews and allows unchecked accusations with little or no accountability.

Evidently, PPPR invites questions about the quality of those peers, but it also points to a new direction science is taking. In the twenty-first century, research in the social sciences is no longer considered an isolated effort of one individual (or a small group of individuals), but rather that of whole networks. Using the strength of collectives (networks) while simultaneously answering the increasing call for greater transparency, we find a growing inclination among social scientists to use open repositories to deposit and share data, pre-registration of protocols, and the commissioning of experts to monitor and review research. Thus, in the social sciences (modeling the medical sciences), ethical review boards have attained a progressively more important function in research.

While many of these initiatives further the social sciences in becoming more open and more accountable, aiding it in diminishing publication bias and forms of sloppy science, it does little to overcome confirmation bias, which still looms over the field, mainly because scientists will still only publish ‘significant’ results. In an attempt to address this problem, Ioannidis (2012) and van Assen et al. (2014), among other advocates, propose that journals should no longer focus on novel findings. Let them instead publish *everything*, including null results. They argue this change will make the scientific record complete, rather than fragmented.

6.7 Conclusions

6.7.1 Summary

In this chapter, we’ve followed the empirical cycle from beginning to end, exploring the various ways bias may disturb or corrupt our findings. We found that research does not always reveal what was intended or desired, leading to the danger of misrepresentation, one-sidedness, or even the production of downright falsehoods.

We showcased how the questions we ask may be biased towards the confirmation of what we already know. *Confirmation bias* (AKA *myside bias*) effectively obstructs creativity and progress in science and impedes more objective or at least impartial explorations from taking place.

With a strong incentive to publish research that show significant results, the danger of *questionable research practices* (QRPs) was introduced. *Data massaging*, *p-hacking*, *HARKing*, and other tricks meant to lower the p-value and thus ‘heighten’

the validity of research outcomes have the potential of polluting research findings on a large-scale, and endangers science's credibility.

The *file-drawer problem* and *false positives* point to the danger of bias during the dissemination process and reside under *publication ethics*. The tendency to report only what is significant, and to avoid reporting null findings creates a distorted view of reality, further enhanced by *editorial* and *review bias*, and the dangers of *self-deception*.

Increased retractions of 'contaminated' (fraudulent) papers show that science is able to self-correct, but the question is whether this is enough. Some argue that the system is strong enough to correct itself in the long run, whereas others believe more drastic measures are called for, including *post-publication peer review* (PPPR), *pre-registration*, and new journal policies to publish everything instead of only 'interesting', 'novel', and 'significant' findings.

6.7.2 Discussion

Research falsifying clearly poses a threat to science's claims of objectivity, verifiability, and other core values of science (see Chap. 2). Part of the problem may be attributed to overly ambitious researchers not taking the standards seriously enough, but part of it cannot be attributed to willful misconduct. Confirmation bias may be the result of something that remains entirely unconscious, and the file-drawer problem may be more likely the result of a fault in the system than the fault of an individual researcher. Similarly, editorial bias seems ingrained in the larger dissemination process, and certainly requires further attention. What suggestions do you have for addressing these ever-present issues of falsifying?

Case Study: Yanomami Violence and the Ethics of Anthropological Practices

Toon van Meijl

The classic monograph by Napoleon Chagnon ([1968] 1983) about violence and warfare among the Yanomami Indians in the Amazon is one of the best-known ethnographic studies of a tribal society. In combination with a number of films he made alongside Timothy Ash, the monograph *Yanomamo: The Fierce People* offered a penetrating picture into a society that was intensely competitive and violent. In Yanomami society, one-third of all adult men were claimed to die as a result of violence (Fig. 6.5).

Fig. 6.5 Yanomami woman and her child at Homoxi, Brazil, June 1997. (Photo: Cmacauley)



Chagnon visited the Yanomami periodically over many years to examine his assumption that patterns of warfare and violence may best be explained in terms of man's inherent drive to have as many offspring as possible, which he labelled reproductive fitness. He argued that the most aggressive men win the most wives and have the most children, thus passing their aggressive genes on to future generations more abundantly than the peaceful genes of their nonaggressive rivals. For Chagnon, the Yanomami provided an excellent case of this sociobiological principle because in the 1960s, while they were exhibiting an intense competition for wives, they were still virtually unaffected by Western colonial expansion.

The assumption that Yanomami society had not been influenced by colonial contacts, however, has been criticized by the historical anthropologist R. Brian Ferguson (1995). Rather than viewing the Yanomami as innately violent, he interpreted the intense violence in the region as a direct consequence of changing relationships with the outside world. Although the villages visited by Chagnon may not have had contact with missionaries or colonial officers, their presence in the wider region had disturbed the balance in inter-communal relations, especially by the introduction of steel tools and weapons. As a consequence, the rivalries between villages intensified and fighting erupted in efforts to gain access to the increasingly important new goods available in the region. Accordingly, Ferguson contended that the fighting was a direct result of colonial circumstances rather than biological drivers.

Several years later, Chagnon's interpretation of violence in Yanomami society was also criticized on ethical grounds by the investigative journalist Patrick Tierney (2000b). He argued that the violence witnessed by Chagnon had not only been caused by indirect influences of colonial contact with westerners, as Ferguson had argued, but also by Chagnon's own fieldwork practices. He pointed out in great detail that Chagnon had contributed to disturbing the balance between communities

by providing steel goods, including weapons, to his informants, which in turn provoked numerous conflicts, raids, and wars. He was also accused of exploiting hostilities between factions and rival communities so he could document violent incidents for the films he and Timothy Ash produced. Finally, Chagnon was charged with transgressing Yanomami ethics by obtaining the names of dead relatives, which was considered taboo for surviving relatives. Thus, Chagnon's own fieldwork practices were argued to be a direct cause of the violence that he explained only in terms of genetics.

The publication of *Darkness in El Dorado* (Tierney 2000b) was preceded by a pre-publication in *The New Yorker* (Tierney 2000a), which appeared shortly before the annual meeting of the American Anthropological Association (AAA) in 2000. This piece highlighted an additional accusation, namely that Chagnon had collaborated with epidemiologist James Neel, who was claimed to have tested a new vaccine against measles among the Yanomami. As a consequence of Neel's work, hundreds of Yanomamis were said to have died because they never built up an immunity to the measles virus. To prevent a huge scandal that could severely harm the reputation of the entire discipline of cultural anthropology, a public debate was held at the AAA meeting about the ethical aspects of Chagnon's research practices. According to his critics, he had violated the ethics of ethnographic fieldwork in order to prove his sociobiological hypotheses about the genetic causes of violence and warfare (Turner 2001).

The debate about Chagnon did not only focus on the ethics of field research among a vulnerable group, but also on the professional responsibility of anthropologists. In this context, Chagnon was criticized for collaborating with a group of wealthy Venezuelans in order to obtain access to the living area of the Yanomami Indians, despite the Venezuelan government rejecting his application for a research visa. More importantly, however, Chagnon was criticized for not objecting to the use, or abuse, of his representation of the Yanomami as extremely violent and prone to warfare. Chagnon's characterization of the Yanomami was later used to prevent the establishment of a reservation by gold prospectors who joined into a coalition with politicians, military leaders, and journalists so they could continue their search for gold in the Amazon. A Brazilian organization of anthropologists submitted a form of protest about this to the AAA. This protest, in turn, caused the AAA to investigate the work of Chagnon and its dissemination.

The report of the so-called El Dorado Task Force, however, is equally as controversial as Chagnon's work. Chagnon's critics argue the report is too weak, while his supporters argue it is too strong. The report did rehabilitate the reputation of epidemiologist James Neel, but Chagnon will likely be forever stuck in a widely contested ethical debate. The confusion about the report, however, has only increased since the membership of the AAA rejected it (Borofsky 2005). At the 2009 AAA meeting, a new panel was organized to discuss this controversy, which accused the AAA of scandalous behaviour by using Tierney's book to investigate Chagnon and his companion Neel, rather than defending these researchers against so-called false journalism by Tierney.

Assignment

1. Is it possible to use modern societies as ethnographic analogies to suggest how early prehistoric societies operated? Or should anthropological research always be situated in a specific social, political, and historical context?
2. What guidelines can we suggest to ensure that anthropological field research practices do not violate a code of ethics for research involving human participants?
3. How can we define the professional responsibility of anthropological researchers to influence the reception and use of their findings?
4. Do anthropologists have an obligation to protect the interests of their research participants, even when they are allegedly violent?

Case Study: Fraud or Fiction? Diary of a Teenage Girl

When psychoanalysis was still a developing discipline (early 1900s), a publishing house was founded that would issue psychoanalytic literature exclusively. Among its publications were the journals of a teenage girl named Gretl (*Tagebuch eines halbwüchsigen Mädchens*, 1919) (Fig. 6.6). She was 11 at the start of the journal, and 14 by the end. Her ‘case’ seemed to illustrate Freud’s theories on psychosexual development perfectly.

Gretl came from an upper middle-class family and was a typical teenage girl: she gossiped, quarreled with her friends and made up again, cried hot tears over silly things, and, of course – she came of age. More specifically, she became aware of her own sexuality. She discovered the difference between boys and girls and found out about the ‘great secret.’ Writing in an October 9th entry she exclaimed: ‘Now I know everything!! So that’s where little children come from.’

By the time she turned 14, her mother had died. At the funeral, she expressed feelings of hurt because her older sister Dora was allowed to walk besides her father in church, but she was not. Dora even said to her sister that the death of their mother was ‘God’s way of punishing their father’ because they (the sisters) had kept things hidden from their mother – a typical instance of ‘magical thinking’, as described by Freud.

The diary was supposedly authentic. Not a word was altered, the anonymous editor of the journals assured the reader, nor had grammatical errors been corrected (so presumably slight but meaningful slips of the pen could reveal the young girl’s true intentions).

The diary confirmed many psychoanalytic notions in detail (sexual anxiety, childhood jealousy, oedipal feelings, etc.). In fact, in an introductory note to the book, Sigmund Freud wrote: ‘The diary is a little gem. I really believe it has never before been possible to obtain such a clear and truthful view on the mental impulses

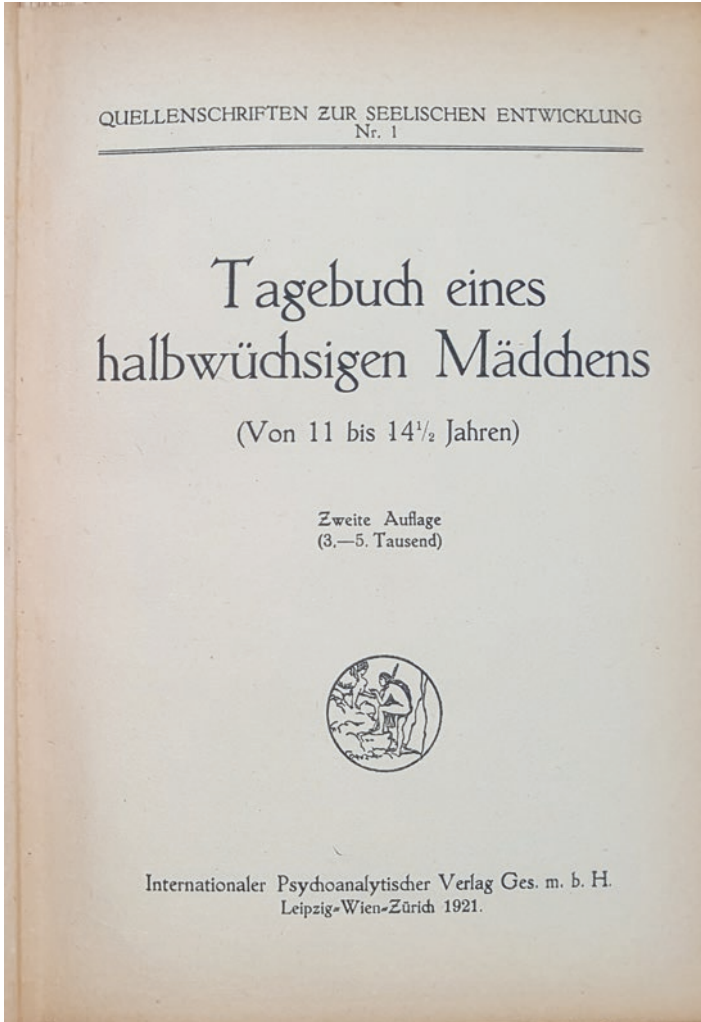


Fig. 6.6 Cover of *Diary of a Young Girl* (Tagebuch eines halbwüchsigen Mädchens), published in 1919

that characterize the development of a girl in our social and cultural stratum the years before puberty.’

A year after the journal’s arrival, Hermine Hug-Hellmuth, an early (and now forgotten) follower of Freud and practicing child analyst, confirmed rumors that it was she who had collected the young girls notes and published them (Fig. 6.7). In 1921, an English translation of the diary appeared, and it became a commercial success. Shortly thereafter however, accusations of fraud bubbled to the surface. According to a critic, Gretl’s journals were too sophisticated to be true. The critic?

Fig. 6.7 Hermine Hug-Hellmuth, an early follower of Freud. (Source: wikicommons)



Cyril Burt, then a young psychologist, who ironically would later be exposed as a fraud himself (see Chap. 5, Box 5.6 on Cyril Burt).

The editor of the journal denied all allegations, claiming that Gretl's published diary entries were 'authentic' and had not been 'touched up.' While the controversy raged on, Hug-Hellmuth tragically died (she was murdered by her nephew, whom she had partly raised and treated with psychoanalysis). In the years after her death, more incriminating details of fraudulent information surfaced. Critics revealed numerous chronological errors, including Gretl's mention of a grading system at her school which was introduced only after the diary had supposedly been written. Today, historians concur that the diaries are not authentic and, in all likelihood, were largely if not entirely made up by Hug-Hellmuth.

This case raises three important questions: (1) Why would someone want to publish a fictitious diary? (2) How did the psychoanalytic community respond to the affair when the diaries were exposed as fraudulent? And (3) How does a case like this reflect on the field of psychoanalysis in general?

It may not have been fame the author was looking for. Rather, as Appignasi and Forrester proposed in their review of the case (1992, p. 200), Hug-Hellmuth had merely meant to 'provide evidence for Freud's theories.' While this explanation is to a certain extent circular, it still gives us a hint as to her possible motives. Psychoanalysis was still a young science in the first few decades of the twentieth century and it was very much in need of confirmation, with many of Freud's followers struggling to find support for his concepts. Lacking a library of psychoanalytic cases in the field's early years, many enthusiasts turned to myths, stories, and historical figures for evidence. The 'Diaries of a Young Girl' seems to fit perfectly into this pattern of early 'missionary work' that was meant to give credibility to psychoanalysis.

How did the psychoanalytic community respond to the allegations? The editors of the publishing house were still making desperate efforts to check the diaries' authenticity by the time Hug-Hellmuth died (in 1924) (Borch-Jacobson and Shamdasani 2012, p. 284). By 1927, the publishers decided to retract the book, directing bookstores to return any remaining copies without an accompanying rationale. The English translation, published in the UK, however, remained available and was reprinted several times, with no note of its fictitious nature. Additionally, a number of practicing psychoanalysts continued to defend the diaries. As an example, Helene Deutsch said she considered Hug-Hellmuth to be 'too imaginative to have recreated a childhood out of whole cloth' (quoted in Roazen 1985, p. 19). In sum, while the history of psychoanalysis is riddled with controversies, it appears that the case of the forged diary had little to no impact on the early reception of the field.

Assignment

Should the discovery of the fraudulent diaries have had a bigger impact on the early reception of psychoanalysis? Consider some of the possible reasons they didn't.

Personal factors – Hug-Hellmuth was a woman working in a field dominated by men; she was not considered a central figure in psychoanalysis.

Contextual factors – Hug-Hellmuth met an untimely death and could never be held accountable, nor could fraud be sufficiently established at the time.

Disciplinary factors – Psychoanalysis has often been accused of being a sect-like cult, not open to discussion.

1. Which of these factors do you think holds the most weight?
2. Think of a similar case of fraud (Diederik Stapel or Cyril Burt for example) and consider which of these factors impacted the field most. How so?

Suggested Reading

For a general introduction into the methodological problems in present-day science, we recommend John Staddon's 2017 highly readable *Scientific Method: How Science Works, Fails to Work or Pretends to Work*. We also recommend Fanelli's papers on retractions in scientific literature, and the question of whether or not they signify a positive trend (Fanelli 2013, 2018). A must read on the subject of false positives can be found in Ioannidis (2005) 'Why Most Published Research Findings Are False.' Finally, we recommend Stroebe and Spears' 2012 article 'Scientific Misconduct and the Myth of Self-Correction in Science,' which provides a crucial discussion of some of the proposed measures to counter the problems discussed in this chapter.

References

- Baud, M., Legêne, S., & Pels, P. (2013). *Draaien om de werkelijkheid* [Circling around reality]: Rapport over het antropologisch werk van prof. Em. M.M.G. Bax. Amsterdam University, 9 September 2013.
- Chamber, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton: Princeton University Press.
- Cromey, D. W. (2010). Avoiding twisted pixels: Ethical guidelines for the appropriate use and manipulation of scientific digital images. *Science and Engineering Ethics*, 16, 639–667. <https://doi.org/10.1007/s11948-010-9201-y>.
- Fanelli, D. (2013). Why growing retractions are (mostly) a good sign. *PLoS Medicine*, 10(12), e1001563. <https://doi.org/10.1371/journal.pmed.1001563>.
- Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, 115(11), 2628–2631. <https://doi.org/10.1073/pnas.1708272114>.
- Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Studies*, 7(1), 45–52. <https://doi.org/10.1177/1948550615612150>.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505. <https://doi.org/10.1126/science.1255484>.
- Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, 93(2), 216–229. <https://doi.org/10.1037/0033-295X.93.2.216>.
- Hubbard, R., & Armstrong, J. S. (1994). Replication and extension in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing*, 11(3), 233–248. [https://doi.org/10.1016/0167-8116\(94\)90003-5](https://doi.org/10.1016/0167-8116(94)90003-5).
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- Ioannidis, J. P. A. (2011). Quantifying selective reporting and the proteus phenomenon for multiple datasets with similar bias. *PLoS One*, 6(3), e18362. <https://doi.org/10.1371/journal.pone.0018362>.
- Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7(6), 645–654. <https://doi.org/10.1177/1745691612464056>.
- 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>.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Köbben, A.J.F. (2012). *Bedrog in wetenschap* [Fraud in science] Lecture before the department of humanities at the Royal Academy of Sciences, 9 January 2012.
- Kolfschooten, F. (2012). *Ontspoorde wetenschap. Over fraude, plagiaat en academische mores over fraude, plagiaat en academische mores [science derailed. On fraud, plagiarism, and academic Morales]*. Amsterdam: Uitgeverij de Kring.
- Loftus, E. A., & Ketcham, K. (1994). *The myth of repressed memory*. New York: St Martin's Press.
- Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences*, 17(2), 115–127. <https://doi.org/10.1016/j.lindif.2007.05.003>.
- Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of *p* values just below .05. *The Quarterly Journal of Experimental Psychology*, 65(11), 2271–2279. <https://doi.org/10.1080/017470218.2012.711335>.
- Møller, A. P., & Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends in Ecology & Evolution*, 16(10), 580–586. [https://doi.org/10.1016/S0169-5347\(01\)02235-2](https://doi.org/10.1016/S0169-5347(01)02235-2).

- Munafò, M., Nosek, B., Bishop, D., et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021. <https://doi.org/10.1038/s41562-016-0021>.
- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should ALWAYS check for them) check for them. *Practical Assessment, Research, and Evaluation*, 9(6), 1–8.
- Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, 77, 562–571.
- Pezdek, K., & Banks, W. P. (1996). *The recovered memory/false memory debate*. San Diego: Academic.
- Pfeiffer, T., Bertram, L. & Ioannidis, J.P.A. (2011, March 29). Quantifying Selective Reporting and the Proteus Phenomenon for Multiple Datasets with Similar Bias. *PlosOne*, <https://doi.org/10.1371/journal.pone.0018362>
- Resnick, B. & Bellus, J. (2018, October 24). *A top Cornell food researcher has had 15 studies retracted. That's a lot.* Retrieved from: <https://www.vox.com/science-and-health/2018/9/19/17879102/brian-wansink-cornell-food-brand-lab-retractions-jama>
- Sharaf, M. (1983). *Fury on earth: A biography of Wilhelm Reich*. New York: Da Capo Press.
- Shemer, M. (2002). *The borderlands of science. Where sense meets nonsense*. Oxford: Oxford University Press.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>.
- Staddon, J. (2017). *Scientific method: How science works, fails to work or pretends to work*. London: Taylor & Francis.
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7(6), 670–688. <https://doi.org/10.1177/1745691612460687>.
- Teixeira da Silva, J. A. (2015). Debunking post-publication peer review. *International Journal of Education and Information Technology*, 1(2), 34–37. <http://www.aiscience.org/journal/ijeit>.
- Toplak, M.E. & Stanovich, K.E. (2003), Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Appl. Cognit. Psychol.*, 17: 851–860. doi: “<https://doi.org/10.1002/acp.915>” <https://doi.org/10.1002/acp.915>.
- Van Assen, M. A. L. M., van Aert, R. C. M., Nuijten, M., & Wichert, J. M. (2014, January 17). *Why publishing everything is more effective than selective publishing of statistically significant results*. *PlosOne*. <https://doi.org/10.1371/journal.pone.0084896>.
- Wolfe, C. R., & Britt, M. A. (2008). Locus of the myside bias in written argumentation. *Thinking & Reasoning*, 14, 1–27.

References for Case Study: Yanomami Violence and the Ethics of Anthropological Practices

- Borofsky, R. (2005). *Yanomami: The fierce controversy and what we can learn from it*. Berkeley: University of California Press.
- Chagnon, N. A. (1983). *Yanomamö: The fierce people* (3rd ed.). New York: Holt, Rinehart, & Winston. (Original work published 1968).
- Ferguson, R. B. (1995). *Yanomami warfare: A political history*. Santa Fe: School of American Research Press.
- Tierney, P. (2000a). The fierce anthropologists. *The New Yorker*, October 9, pp. 50–61.
- Tierney, P. (2000b). *Darkness in el dorado: How scientists and journalists devastated the Amazon*. New York/London: Norton & Company.
- Turner, T. (2001). *The Yanomami and the ethics of anthropological practice*. Ithaca, NY: Cornell University, sponsored by the Latin American Studies Program, with support from the U.S. Department of Education Title VI.

References for Case Study: Fraud or Fiction? Diary of a Teenage Girl

- Appignasi, L., & Forrester, J. (1992). *Freud's women*. London: Weidenfeld & Nicolson.
- Borch-Jacobson, M., & Shamdasani, S. (2012). *The Freud files*. Cambridge: Cambridge University Press.
- Hug-Hellmuth, H. (1919). *Tagebuch eines halbwüchsigen Mädchens*. Vienna: Internationaler Psychoanalytischer Verlag.
- Roazen, P. (1985). *Helene Deutsch. A psychoanalyst's life*. New York: Meridian Books.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

