Behind the scenes of scientific articles: defining categories of fraud and regulating cases

David Pontille
Didier Torny

Working Paper 13-MS-03

January 2013
Introduction

On March 7, 2012, the blog Retraction Watch\(^1\) announced that Yoshitaka Fujii, Associate Professor of Anesthesiology, had just been fired from the University of Toho for having published nine clinical studies that did not comply with the current ethical guidelines. The blog also announced that one of the journals concerned, Anesthesia & Analgesia, whose subtitle is “the gold standard in anesthesiology,” revealed in a communiqué that a study of the integrity of Dr. Fujii’s research had been launched in 2010 (Schafer 2012). The journal’s American editor-in-chief indicated that 24 other articles could be the subject of scientific fraud. He also stated that in 2000 his journal had published a letter from German researchers presenting a meta-analysis of 47 articles jointly signed by Dr. Fujii (Kranke et al. 2012). These authors underscore the remarkably identical frequency of headaches as side effects of anesthesia in all studies and concluded that only an underlying influence could explain this stupefying consistency.

Dr. Fujii had responded with a short jointly published letter, simply indicating that he had reported empirical observations and that they were expected given the state of knowledge. This response was sufficiently convincing to the International Anesthesia Research Society that Dr. Fujii was able to publish 12 additional articles over the following decade. Yet the German researchers had pursued their meta-analyses, in 2001 showing in Acta Anaesthesiologica Scandinavica that 64% of the clinical data published for an antiemetic originated from Dr. Fujii’s team, even though his dosage results diverged substantially from those produced by all of the other laboratories (Kranke et al. 2001). Following this

---

* Both authors contributed equally to data collection and processing, to the analysis, and to writing the article. This working paper has been published in French in Revue d’Epidémiologie et de Santé Publique, 2012, vol. 60(4), p. 247-253.

\(^1\) http://retractionwatch.wordpress.com/
publication, the authors alerted the Food and Drug Administration and the Japanese regulatory agency, apparently to no avail.

In March 2010, an editorial in the British journal *Anaesthesia* described the obstacles to demonstrating falsification and the invention of clinical data, taking as an example the suspicions weighing against Dr. Fujii’s data (Moore et al. 2010). This editorial provoked a large number of responses, leading editorial boards to engage in a reflection on their role in combatting scientific fraud (Wagner 2012). Also, in 2011 when Dr. Fujii submitted an article to the *Canadian Journal of Anesthesia*, where he had already published 39 articles, the journal opened an inquiry with the University of Toho, which revealed that there was no Committee for the protection of human research subjects, prompting investigations on all of his research production (Miller 2012). At the same time, a meta-analysis of 169 controlled trials cosigned by Dr. Fujii, initiated following the 2010 editorial and verified at length, was published as a special article in *Anaesthesia* on March 8, 2012. It concluded that the distribution of 28 of the variables published did not correspond to what could be expected, with the probability of occurrence under the hypothesis of independence even reaching $10^{-33}$ for one of them (Carlisle 2012).

Finally, on April 9, 2012, a joint letter from 22 editorial directors was addressed to the leaders of nine institutions that had hosted Dr. Fujii’s research and simultaneously published on some of the journals’ web sites. It informed them of the publication of this study and announced that the articles would be retracted, unless the institutions involved could provide proof that the articles were trustworthy: “for each study concerned, we request that your institution declare 1/ that the study was conducted as presented in the article, 2/ that you have examined the primary data and verified that they were authentic, and 3/ that the appropriate ethical framework was clearly established for the study” (Collectif 2012).

Answering that call, the Japanese Society of Anesthesiologists led an inquiry and concluded, in its 29th June 2012 report, that 172 articles jointly signed by Dr Fujii contained falsified data and should be retracted, which is an unequaled record for a single author.

How can one explain that Dr. Fujii was able to publish apparently falsified data for such a long time? Is this an isolated researcher who was particularly skilled in fabricating forgeries corresponding to the expectations of the scientific community, including those of his coauthors? Is this a local institutional problem related to his mandarin status, as in the case of Professor W.S. Hwang, a Korean national hero and fraudulent inventor of human cloning (Gottweis & Kim 2009)? Are we dealing with a disciplinary idiosyncrasy, since two other anesthesiologists have recently undergone similar waves of retracted articles and that 4% of the articles submitted to *Anaesthesia* have turned out to be plagiarized (Yentis 2010)? Or should we consider that any academic production is potentially tarnished, implicating all scientific institutions (authors, financers, editors, journals) (Ioannidis 2005)?

Two major interpretations are generally contrasted in this type of situation. The first tolls the bell of scientific integrity: it is mainly based on the massive increase of the number of articles retracted (Van Noorden 2011), on the fact that these retractions are strongly correlated with the journals’ impact factor (Fang & Casadevall 2011) or that the retractions for fraud take

---

2 Let us not forget the brilliant physicist J.H. Schön, covered with prizes and the lead author of eight articles published in *Science* and seven articles published in *Nature* between 2000 and 2001, but then fired by Bell Labs and whose dissertation was revoked by the University of Constance.
place in journals with a higher impact factor than the retractions for experimental or
calculation errors (Steen 2011a). It also responds to the high frequency of researchers
declaring having observed colleagues involved in practices they deem contrary to scientific
ethics (Fanelli 2009). In contrast, the second interpretation ferociously defends tooth and nail
a model of self-regulation. It is founded, for example, on the usually brief time between
publication and retraction, the generally low number of citations related to these articles
(Furman et al. 2012), the appearance in some journals of temporary publication bans, and
considers the creation of specialized places such as the Abnormal Science and Retraction
Watch blogs as extensions of peer review (Marcus & Oransky 2011).

Our objective in this article is not to choose between these two positions, which stabilized
over the 1980s (Broad & Wade 1982; Lafolette 1992) and were regularly reactivated by other
actors with each new case detected. Instead, from a perspective informed by science and
technology studies, we wish to establish a general diagnosis on the regulation of publication
practices and suggest methods of analysis, by reviewing both the recent cases and others,
less recent, that they call to mind. First of all, this paper reviews how the practices deemed to
be deviant have been qualified in the past by distinguishing the categories stemming from
data integrity and the conditions of data production from those concerning the relationships
between the authors of publications and their content. We will then analyze three specific
problems and the institutional responses aiming to prevent them, channel them, and attempt
to solve them.

**Striking it rich with dubious data**

Medical research for the most part has been marked by a progressive rise in ethical
demands that strive toward defining the quality of the data produced. The need to collect
informed consent from the subjects participating in therapeutic research is a good example,
the subject of a vast literature on its concrete manifestations (Corrigan 2003). Similarly, in the
construction of all the instruments designed for direct or indirect use in epidemiological
studies, protecting the subjects’ anonymity is essential to conducting these investigations.
These diverse ethical demands frame the conditions of data production: any time compliance
is lacking, the reliability of the data is compromised for this reason alone, as seen with the
case of Dr. Fujii. But the materials on which the results published are founded can be subject
to tampering beyond these regulatory obligations.

Several ways of contravening “good practices” can be identified in the series of affairs
discussed herein. The first resides in the selection of relevant data to obtain more coherent
results, a smoother curve, a more readable image. R. Fisher, an important statistician,
geneticist, and support for eugenics, 80 years after the famous experiments by Gregor
Mendel on his peas, showed, for example, that the probabilities obtained were undoubtedly
too good to be true and stemmed more from a confirmation bias than from an observation.
His research has contributed to designing tools to remove the biases introduced by
experimenters, notably in controlled clinical trials (Marks 1999). In this process of selecting
relevant data, the question of ill will is secondary to all the possible means of deviating from
an experimental, observational, or statistical norm.

The second consists in conscious falsification activities, which can manifest as minor
arrangements of experimental material, falsification of data, or manipulation of images. D.
Das, a surgeon working on the protective effects of resveratrol, explaining in fine the probable role played by wine consumption in the French Paradox, was suspected of having massively doctored Western blot images that were the basis of dozens of articles that he published, leading to his being fired from the University of Connecticut after 3 years of inquiry (DeFrancesco 2012). The multiplication of cases raising questions on the falsification of images reveals just how fragile and uncertain are the graphic representations that are held to be tangible components of scientific proof in the end (Lynch & Woolgar 1990), and raises debate on the nature of the data necessary to certify their integrity.

Finally, the third manner stems from fabrication pure and simple. Falsification extends here to all materials and is based on the pure invention of data. The example of the physician and dentist Jon Sudbø comes to mind, who, in 2006, admitted to having literally created 900 patients in a Norwegian cohort, 250 of whom had the same date of birth. This allowed him to demonstrate spectacular effects of taking nonsteroidal anti-inflammatory drugs on oral cancer in heavy smokers and to publish in *The Lancet* the year before.

The distinction between selection, falsification, and fabrication is not self-evident and instead comes from *ex post* attribution when cases are revealed after investigations (Fanelli 2009). As in the affairs concerning Prof. Hwang, when there are no detailed admissions, it is impossible to distinguish between what has been authentically produced in the laboratory and what is instead voluntary or involuntary artifact. If the absence of reproducibility is often considered a clue to falsification, the opposite is not necessarily true: a highly skilled forger can produce a sort of retroplagiarism, arranging or fabricating data that others, through slow scientific work, will have actually produced and analyzed.

Circumventing proof of originality

One of the fundamental principles of scientific activity consists in producing original results, as indicated by the categorization of journal articles that distinguishes “original research” from “reviews”, “letters”, and “editorials.” According to this principle, the recognition of one’s peers and the awarding of prizes are strongly related to the priority of original discovery. No importance is supposedly accorded to second place. Among the different ways to get around this requirement, this section will emphasize the question of plagiarism and authorship.

Plagiarism and self-plagiarism

In these conditions, the temptation is sometimes great to cite only partially and sporadically the similar work that has preceded, or even to totally omit its mention. In our era of electronic publications and cut-and-paste, there is no simpler operation than to insert one or several paragraphs written and published earlier by others into an article being written, for example by copying a piece of text from an article on nosocomial infections in Brazil into an article on the same subject in Spain (Cisterna et al. 2011). The same clearly holds true for data, figures, and images.

At times fraud is not limited to partial plagiarism: at the end of the 1970s, E. Alsabti, an alleged researcher in oncoimmunology, without a doctorate, succeeded in climbing the ladder of an academic career by putting together an impressive list of publications in barely 3 years (Broad & Wade 1982, chap. 3). His tactic was simple: he copied entire articles that had already been published, modified the title, replaced the authors’ names with his own,
and submitted the manuscript to a less well-known journal. He succeeded in accumulating more than 60 publications in this way, in joining 11 scientific societies and in working in several prestigious American institutions.

Rather than borrowing from others’ work, it can be even simpler to dig into one’s own data. What researchers and clinicians have long designated as a “pressure to publish” (Maddox 1988) has turned into the proliferation of articles based on the same research, the spawning of similar results into a variety of texts and the preference for quantity to the detriment of quality (Hamilton 1990). The practice of self-plagiarism has developed to such an extent that it has progressively given rise to terms such as “least publishable unit” (Broad 1981), “salami slicing” (Huth 1986), and even “duplicate publication” (Susser & Yankauer 1993), and the effects on meta-analyses can be considerable (Tramèr et al. 1997). Even if, contrary to plagiarizing others, this censured concealment is interpreted in a variety of fashions, everyone agrees that the basis of the problem lies in the absence of a citation of the original study (Freund et al. 2012).

**Authorship**

An alternative manner to bypass the requirements for originality is to maneuver around the edges of authorship. In many domains, the guidelines regulating access to article byline place value on significant intellectual contributions. Yet the suspected articles often carry the name of researchers whose contribution is considered minimal, if not inexistent, and therefore these individuals have a very distant contact with the research conducted and the results published. Since the end of the 1980s, researchers and journal editors have produced, in what is generally called “honorary authorship”, three categories of authorship to denounce these situations that they judge unacceptable (Rennie & Flanagin 1994; Flanagin et al. 1998).

The first designates “guest authorship”, which separately treats people whose name is recognized: these are particular names that change the status accorded to an article by their presence alone. The presence of these researchers’ names potentially increases the chances of the article publication and its future visibility and thus figures as a sign of quality. This practice is current and arises in many cases. The affair of the cardiologist J.R. Darsee, one of the most closely followed and documented in the 1980s, brought this practice to the light of day: other than the presence of falsified and entirely invented data, the 55 publications that were retracted often carried the name of his mentor, E. Braunwald, who had little knowledge of their content, and who had not been worried during the detailed investigations of the NIH investigators (Stewart & Feder 1987).

Another category, which plays less on this distinction of status or renown, emphasizes more the offer of authorship governed by the principle of reciprocity, as a gift in view of a counter-gift (“gift authorship”): it is not exceptional that researchers accept to make room for someone so as not to offend a partner making the request, to encourage future collaborations, to maintain good cooperative relations, to thicken a list of publications, or to return the favor (Pontille 2004). In this context, the above-mentioned case of E. Alsabti, nonetheless revealed an extreme and particular form of gift at the time: some of his articles included the names of fictitious authors whom E. Alsabti, at the beginning of his fall, blamed entirely for the abuses for which he was denounced.
Finally, as with other types of writing, notably literary, the name of the writer is not systematically the name of those who appear on the byline (“ghost authorship”). This is notably the case when a researcher, who does the greatest share of the work and writes all or nearly all an article, allows only his students or less experienced colleagues to be the authors of the article so as to assist them in publishing in a prestigious journal and thus lengthen their publications list.

Institutional regulations

Common reactions in the event of practices deemed deviant consist in reaffirming the general principles at the foundation of scientific integrity: disinterestedness, organized skepticism regularly held as cardinal virtues (Merton 1942; Weed & McKeown 1998) and stipulated in charters and professional agreements or through specialized structures within professional or learned societies (Coughlin 1997). These reactions also require more restrictive concrete operations aiming to prevent and limit these phenomena by indirect actions. We will examine here three specific operations relative to responsibility, negative results, and conflicts of interest.

Restore responsibility

In an evaluation system that values individual performance, the increase of the number of authors per article is regularly described. Although guest and gift authorship cast doubt on the attribution of scientific credit, other forms are even more surprising. Some cases, most particularly the above-mentioned J.R. Darsee affair, revealed that some authors were not aware that their name had been adjoined to the suspected publications. The journals consequently took control to defend their self-regulatory function. They sought to return responsibility to the authors in two major steps. The first resulted from the many cases of fraud that shook the 1980s: statements are required, signed by all of an article’s producers to ensure that they approve the content and that they indeed intended their name to appear on the article.

The second step played out between 1996 and 2000, following heated debates between researchers, journal editors, and research administrators (Biagioli 1998; Pontille 2001). They aimed to set up traceability of research operations so as to be able to reattribute responsibility in case of a future problem (Torny 2003). After several conferences and an experimental phase, journals opted for a new procedure, called “contributorship”, which promotes transparency: systematically describe the individual contributions in a section of the article specifically designed for this purpose. Before publication, each contributor is now invited to either describe his or her contribution in a written note, or to fill out a form that delineates a taxonomy of research operations. Thus the producers of articles should approve the final version of the text and date their engagement: each one dates and signs the letter or form manually. This procedure reinforces the judicial responsibility of the handwritten signature and forces each contributor to take responsibility for his or her contribution.

Publishing negative results

Whether the result of self-censure or fraudulent practices on the part of researchers, or the practice of journals for visibility or financial reasons (Smith 2005), publication biases favoring positive studies are well known (Simes 1986; Stern & Simes 1997). The additional effect of
this bias is relegating negative results to second-rank journals (Kanaan et al. 2011) or even non-publication. This has been demonstrated for antidepressants, for example (Turner et al. 2008), and for a number of substances approved by the Food and Drug Administration between 1998 and 2000 (Lee et al. 2008), to such an extent that certain researchers now take the rate of positive results published as a transdisciplinary indicator of fraud or bias (Fanelli 2010) and others speak of “evidence-biased medicine” (Melander et al. 2003).

To solve this problem, two complementary solutions have been proposed. For clinical studies, this means founding academic evaluation of drugs based on applications made to the regulatory agencies rather than on the published literature (Turnet et al. 2012). In addition, certain journals propose systematically publishing negative results for all types of research, such as the *Journal of Errology* or the *Journal of Negative Results in Biomedicine*, which has an “unofficial” impact factor of 1.1. These journals consider that unexpected and failed results are important in scientific knowledge, thus returning to one of the founding principles of experimental practice of the seventeenth century, initiated notably by Robert Boyle at the Royal Society of Sciences of London (Shapin & Schaffer 1985).

**Uncover and reduce conflicts of interest**

A third question brings together all of the problematic aspects described in the first sections: conflict of interest. This notion was imported into the scientific world in the 1960s to describe the increasing relations between public research and industry, and the problems that this could raise, in particular in expert consultations with health and environmental regulatory agencies (Gingras & Gosselin 2008). It is through this prism of conflict of interest, for example, that the accusations of scientific fraud against Dr. Needleman were interpreted by two of his colleagues while the Environmental Protection Agency had just adopted restrictive measures for the lead industry (Rosner 2005). More generally, in environmental epidemiology, the consequences in terms of regulating results in public health have led the chemical, tobacco, and asbestos industries to not only finance research that was favorable to them (Pearce 2008), but also to maintain uncertainty over the long term to fight against stabilized knowledge being established (Proctor & Schiebinger 2008).

With the ever-increasing meshing and merging of financing from the pharmaceutical industry and academics, the notion of conflict of interest in its contemporary sense truly appeared in the 1980s (Gingras & Gosselin 2008): it was thus repeatedly demonstrated that financing from industry or pharmaceutical laboratories influenced the results produced (Abraham 1994; Bekelman 2003; Friedman & Richter 2004; Lexchin et al. 2003), including on subtle modalities such as the favorable presentation of hormone replacement treatments well after their risk–benefit ratio was challenged at the beginning of the 2000s (Fugh-Berman et al. 2001). In this context, Elsevier was denounced for creating six seemingly scientific journals that were actually financed by pharmaceutical laboratories and designed to present compilations of articles that were favorable to them (Hansen 2012). Certain forms of ghostwriting also stem from this practice: through specialized companies, the pharmaceutical firms, once the research and articles had been completed, requested researchers who were reputed in their specialty to sign the publications in exchange for consequential financial compensation (Mirowski 2005; Sismondo 2007).

With reform in mind, institutions and journals began including the financing sources and conflict of interest disclosures as a solution to these problems, with, for example, regular
guidelines by the International Committee of Medical Journal Editors beginning in 1993\(^3\), even though their implementation took an exceedingly long time in the eyes of their detractors (Krimsky & Rothenberg 2001). Others have suggested radical reforms of complete separation between private interests and public research (Schafer 2004), a position recently found on expert testimony in the parliamentary discussions after the Mediator French scandal. However, everyone agrees that internal transformations of scientific production, most particularly biomedical research, has made a large number of researchers true entrepreneurs (Etzkowitz 1996). The case of A. Potti illustrates these tensions: as soon as his first articles on genomic markers of the sensitivity of different chemotherapies were published in *Nature Medicine* and the *New England Journal of Medicine* in 2006, Duke University and his team boasted of “personalized medicine” that they were going to develop so as to attract patients, grants, as well as the large pharmaceutical laboratories, so that they could conduct large-scale trials. But a few years later, Dr. Potti was implicated both from inside the science world by biostatisticians (Baggerly & Coombes 2009) and from the outside, through professional letters showing the many “enhancements” that had appeared on his resumé (Goldberg 2010), and then by his patients’ families attracted by the potentially revolutionary character of his practices, who had sued for having been deceived.

**Conclusion**

Emphasizing how long these problems have been around, we had a double objective: first, showing the emblematic value of certain cases, truly striking episodes that have produced preventive measures so that these practices would not continue; second, taking stock of the increase in the number of affairs revealed that have led to the retraction of articles as a clue to a “surveillance bias.” Just as certain fraudulent practices have become simpler to carry out, the detection and dissemination of behaviors considered problematic has been largely facilitated. For example, journals such as *Anaesthesia* routinely use plagiarism detection software, but it is also possible to construct tables that juxtapose text from several articles, showing how close they are, putting together a series of an author’s variable conflicts of interest, or anonymously denounce image falsification by circulating a video clip\(^4\).

The recurrent stakes in all of these problematic situations revolves around the scientific community’s ability to self-regulate. Nearly all of the institutional measures describe herein directly involve academic authors. Yet for the last 15 years, we have seen the journals themselves be targeted, notably in recurrent debates aiming to assess the independence of journal editors (Davis & Müllner 2002) and to reform peer review (Weicher 2008). What should be thought of the responsibility of A&A which for years published Dr. Fujii’s articles after the first meta-analyses? How should the process leading the *Journal of Clinical Microbiology* to publish a plagiarized article published a few years early in its own columns be qualified\(^5\)? Like researchers, journals are also subjected to conflicts of interest (Friedman

---

\(^3\) These guidelines were revised in 2000, and the member journals of the ICMJE took a position on conflicts of interest the following year in a common editorial titled “Sponsorship, Authorship, and Accountability” praising transparency and academic independence. A conflict of interest disclosure form common to all the ICMJE journals was established in 2010.

\(^4\) http://www.youtube.com/watch?v=FXaQwanWnU

& Richter 2005; Lundh et al. 2010) and to the imperatives of success, partly founded on their impact factor, that incite them to introduce references motivated solely by journal citation (Smith 1997), a practice recently qualified as “coercive citations” (Wilhite & Fong 2012).

Symmetrically, the case of Dr. Fujii reminds us that journals have a tremendous collective organizational power, widely implemented in the authorship and conflict of interest reforms. In addressing their letter to research institutions, journals have placed them before their own responsibility, thus attempting to reverse the burden of proof. However, the distribution of the responsibility between the actors of academia is also played out on another level: the responsibility with regard to patients, the eventual victims of deviant behaviors (Steen 2011b). Indeed, if research having consequences in terms of treatments, prevention modalities, or, more generally, public health is at the heart of conflicts of interest, opening it to nonacademic actors allows new criticisms to be expressed and taken into consideration through judicial and political action on the part of users and citizens.

Acknowledgements
We thank Anne Bertrand (EHESS) for her careful reading of the manuscript and her suggestions on an earlier version. Most of the recent affairs studied in this article stem directly or indirectly from their revelation in the blog Retraction Watch, maintained by Adam Marcus and Ivan Oransky. We thank them for their meticulous and nearly daily work assembling this information.

References


