

How Metrics-Based Academic Evaluation Could Systematically Induce Academic Misconduct: A Case Study

Junhui Han and Zhengfeng Li

Received: 6 November 2016 / Accepted: 19 July 2017
© 2018 Ministry of Science and Technology, Taiwan

Abstract This article analyzes a case of postproduction misconduct, that is, the BioMed Central (BMC) retraction incident, which is the beginning of a series of massive retraction incidents that China has encountered in recent years. Our analysis echoes the pioneering research of STS scholar Mario Biagioli, who argues that our academic culture is shifting from “publish or perish” to “impact or perish.” Getting a good score on the metrics of academic evaluation becomes the goal of some scholars, leading to the emergency of postproduction misconduct. As revealed in the BMC retraction incident, commercial agencies that claimed to be able to facilitate academic publishing manipulated the peer-review process of academic papers by fabricating their peer reviews to help some clinicians meet the requirements of the title assessment system. This article advances Biagioli’s argument by expounding on the following two characteristics of the BMC retraction incident: first, peer reviews were fabricated by agencies instead of the authors themselves; second, the incident was induced systematically by the title assessment system, instead of particularly by individual factors of the authors. Through analyzing this case, we obtain important insight into postproduction

J. Han

Institute of Science, Technology, and Society, Tsinghua University, China.
e-mail: hanjh14@mail.tsinghua.edu.cn

Z. Li

Institute of Science, Technology, and Society, Tsinghua University, China.
e-mail: lizhf@mail.tsinghua.edu.cn

An earlier version of this article was presented in the Fifth Cross-Straits STS Academic Exchange Meeting in National Yang-Ming University, October 2015. We are grateful for the productive discussions there, which helped us formulate our arguments. Our thanks also go to the China Association for Science and Technology (CAST), which invited us to join the investigation team for several interviews and shared the reports based on all interviews with us. Two anonymous reviewers of *East Asian Science, Technology and Society* provided many valuable comments, which helped improve this article. The editors of the journal, Wen-Hua Kuo and Hee-je Bak, gave us many precious suggestions as well. In particular, Wen-Hua Kuo discussed this article with us several times, which helped a lot. Last but not least, we would like to thank Lili Dong, with whom we talked about this article and got helpful advice. Despite all the feedback we received from the researchers mentioned above, we take full responsibility for the views and arguments of this article.

misconduct, which is beneficial to more thoroughly understanding and then mitigating this new type of academic misconduct.

Keywords BMC retraction incident · postproduction misconduct · fake peer reviews · title assessment system

Abstract 本文分析了一个后生产性学术不端的案例,即 BioMed Central (BMC) 撤稿事件,该事件也是近年来中国遭遇的一系列大规模撤稿事件的开端。本文的分析响应了 STS 学者 Mario Biagioli 的开创性研究。Biagioli 认为,我们的学术文化正在从“发表,或者毁灭”(publish or perish) 向“有影响,或者毁灭”(impact or perish) 转变,在学术评价指标上获得一个好的得分成为一些学者的目标,从而导致了后生产性学术不端的出现。在 BMC 撤稿事件中,一些声称能够帮助进行学术发表的商业性中介通过伪造评审意见操纵了学术论文的同行评审过程,以帮助一些临床医生达到职称评审系统的要求。本文通过详细阐释 BMC 撤稿事件的以下两个特点推进了 Biagioli 的论述:首先,评审意见是由中介伪造的,而非作者自己;其次,该事件是由职称评审体系系统性地诱发的,而非由作者的个人因素个别地导致的。通过对此案例的分析,本文获得了对后生产性学术不端的一些重要认识,有助于更为深刻地理解这种新型的学术不端现象并减少其发生。

Keywords BMC 撤稿事件 · 后生产性学术不端 · 虚假评审意见 · 职称评审体系

1 Introduction

In April 2017, due to fake peer reviews, the science, technology, and medicine publisher Springer retracted more than one hundred articles authored by Chinese scholars, which in turn sparked a media frenzy. While this story dominated the Chinese media, it was not the first time that China had encountered a massive retraction incident in recent years. In 2015 BioMed Central (BMC) retracted forty-one Chinese publications, followed by Springer retracting sixty-four, Elsevier retracting nine, and Nature retracting three.¹ In each case, the articles were retracted for fabricating peer reviews. Clearly, this is a kind of academic misconduct, but it is quite different from the traditional type of misconduct, which mainly refers to fabrication, falsification, and plagiarism (NAS et al. 1992; Li and Li 2011). To distinguish from traditional misconduct, which usually occurs in the production process of the article, STS scholar Mario Biagioli (2016) calls this new type of academic misconduct “post-production misconduct.”

Biagioli argues that academic misconduct has been undergoing a radical qualitative transformation in recent years, adapting itself to modern metrics-based regimes of academic evaluation. Unlike traditional misconduct, which concerns fraudulent ways of producing scholarly publications, postproduction misconduct mainly targets the publication system itself. For instance, some authors provide fake e-mail addresses of recommended peer reviewers to the publisher and then offer supportive reviews through the fabricated e-mail addresses. Some authors and journals maximize their citations or impact by citing each other’s work, forming the so-called citation ring. Some even hack publishers’ databases to seek more invitations to review articles,

¹ These incidents were also widely reported and discussed by the Chinese media, two examples of the reports can be found at http://www.beijingreview.com.cn/keji/201511/t20151113_800042489.html and <http://www.jiemian.com/article/1011039.html>, accessed 24 May 2017.

which gives them an opportunity to insert more citations of their own work. According to Biagioli, the cause for this transformation of academic misconduct is the shift of our academic culture, from “publish or perish” to “impact or perish.” It is no longer enough for researchers to just publish their work. Instead, to secure their academic life, their work must be influential. Currently, academic articles are usually measured using a set of metrics that places emphasis on the publisher’s reputation and number of times the work is cited. Thus, researchers aim to obtain a favorable score on the metrics system, which can also incentivize academic misconduct in the process.

Despite a growing number of articles being retracted due to this emerging type of misconduct, there is very little empirical research on postproduction misconduct. In addition, there has been a proliferation of news reports about this phenomenon; however, they often lack details. In this article, we will gain important insight into postproduction misconduct by exploring the BMC retraction incident, which is the first massive retraction incident encountered by China in recent years. Unlike the writers of news reports, we have detailed information obtained through investigation.

In the BMC retraction incident the authors involved were predominately clinicians, distinguishing them from cases of postproduction misconduct seen previously in which the authors were academic researchers. The clinicians were faced with some difficulties in meeting the high publication requirements of the title assessment system (zhi cheng ping shen xi tong 职称评审体系), a system that assesses their professional levels based on a certain set of metrics. As a result, the clinicians sought consultation from “third-party agencies” (di san fang zhong jie 第三方中介), institutions or individuals that serve as intermediaries between authors and publishers and that provide publication-related services. Thus agencies became involved in the production of academic articles by helping clinicians organize, modify, or polish their articles. When it came time for the agencies to assist with the successful publication of the articles, they fabricated the peer reviews.

Similar to Biagioli, this article explores the postproduction misconduct but takes a more critical look at the fact that the peer reviews were fabricated by agencies, rather than by the authors themselves. Moreover, this article investigates the way the incident occurred on a systemic level incentivized by the metrics-based title assessment system as opposed to its being driven by individual factors of the authors. The extended analysis of these characteristics will contribute to the discussion about postproduction misconduct initiated by Biagioli.

The remainder of the article first discusses the BMC retraction incident as well as the research methods utilized in this work. An overview of the retracted articles is given, followed by an analysis of the role that the agencies played in the incident and how the incident was systematically induced. The article ends with important insight into postproduction misconduct from an STS perspective.

2 The BMC Retraction Incident

The BMC retraction incident was first reported by the *Washington Post* on 27 March 2015.² According to the report, BMC retracted forty-three articles due to fake peer

² The online version of this news report can be found at https://www.washingtonpost.com/news/morning-mix/wp/2015/03/27/fabricated-peer-reviews-prompt-scientific-journal-to-retract-43-papers-systematic-scheme-may-affect-other-journals/?utm_term=.294fd06d5494, accessed 24 May 2017.

reviews. The list of retracted articles showed that forty-one out of the forty-three were authored by Chinese researchers. The authors involved were affiliated with respected Chinese universities and hospitals, making the incident an enormous scandal for the Chinese academic community. While most lead authors were young researchers, many corresponding authors were viewed as esteemed scholars and authorities in their fields. Although BMC stated that this incident was not isolated to Chinese scholars, it stunned Chinese society, especially the Chinese academic community, after the story was reported by domestic media.

BMC is a major publisher of scholarly medical and science articles, specializing in open-access journal publication. When submitting to BMC journals, authors were normally required to recommend potential reviewers for their manuscripts, which is a standard used by many publishers. However, in this case the agencies exploited their ability to recommend potential reviewers by providing false e-mail addresses. While BMC remarked that the retracted articles were rather convincing, they became suspicious after finding an unusual pattern of e-mail addresses listed for the potential peer reviewers. In addition to the e-mail addresses, they also were alerted by the fact that the same researcher was recommended as a reviewer across different topics, which is uncommon in highly specialized fields. Finally, when they tracked down some of the researchers who were recommended, they learned that the individuals listed had not served as reviewers, indicating that someone else had used their names.

After investigation BMC concluded that the peer-review process of these articles had been manipulated by a third party. In the retraction notes attached to some of these articles, BMC wrote:

The Publisher and Editor regretfully retract this article because the peer-review process was inappropriately influenced and compromised. As a result, the scientific integrity of the article cannot be guaranteed. A systematic and detailed investigation suggest[s] that a third party was involved in supplying fabricated details of potential peer reviewers for a large number of manuscripts submitted to different journals. In accordance with recommendations from COPE, we have retracted all affected published articles, including this one. It was not possible to determine beyond doubt that the authors of this particular article were aware of any third party attempts to manipulate peer review of their manuscript.

COPE, short for the Committee on Publication Ethics, is a forum for editors and publishers of peer-reviewed journals to discuss all aspects of publication ethics.

Since the incident had a significant impact on the reputation of the Chinese academic community, related authorities such as the China Association for Science and Technology (CAST) and the National Natural Science Foundation of China (NSFC) jointly launched an investigation. CAST is an association of Chinese science and technology workers, while NSFC is the organization that manages the national natural science funds of China, whose role is similar to that of the NSF in the United States. It is unusual for CAST and NSFC to investigate academic misconduct incidents themselves, since these incidents are conventionally handled by the institutions with which the researchers involved are affiliated. The investigation confirmed that most of the retracted Chinese articles were submitted through agencies, rather than the institutions themselves.

In fact, when the investigation was conducted, some institutions had already rendered their settlement suggestions, many of which including severe punishments for the authors. For example, the settlement suggestions of the medical school of a university included suggesting the university address the authors' title and tutor qualification and suspend their ability to teach and enroll students for five years, and recommending their current students change tutors.³ If these punishments had all been implemented, the academic careers of the authors could actually have ended. Based on the joint investigation and their own investigation, the NSFC concluded that six of the forty-one retracted articles were fabricated to various extents. Therefore, they recovered relevant grants and turned down grant applications based on these articles.

Immediately after the incident, BMC eliminated the requirement for authors to recommend potential reviewers. However, they did not punish the authors, for example, by not accepting their manuscripts for one or more years. This stance confirmed that BMC, as it said in its statement, could not determine whether the authors were aware of the agencies' attempts to manipulate the peer-review process, nor would it hold them at fault.

At the time of the incident, there were no definitive rules regulating the role of agencies in the production of academic articles in China, which meant that none of the agencies were punished. These agencies were, and remain, a common presence in other countries, mainly providing language services. Some publishers also recommend agencies to non-English authors to help with their language. Nevertheless, the services provided by the agencies in the BMC retraction incident were not limited to language but also included article production services, such as writing, modification, and submission. After the incident, China introduced "*Five Not Allowed*" in the *Publishing of Scientific Papers* (发表学术论文“五不准”). According to this document, entrusting article writing, modification, and/or submission to agencies was forbidden.

We conducted an anonymous survey of the authors via e-mail, which is introduced in detail later in the method section. In this survey, several authors expressed their views on the incident and the way relevant organizations dealt with it. For instance, an author voiced his anger at the way the incident was handled:

The punishments for authors by relevant organizations (hospitals, universities, and NSFC) are excessively severe and indiscriminate. They hit everyone using the same stick. It seems strict, but, in fact, it is irresponsible. (From questionnaire of responder no. 4)

Another author thought the responsibility for the incident rested with agencies, and that the punishments for authors were actually the relevant organizations' answer to their superiors:

It is apparently the fault of agencies, and it is also a problem of the whole society. However, it is us, the innocent clinicians, that are being spanked. It is nothing more than a statement abroad, and the organizations at different levels have

³ This settlement suggestion was contained in the statement of this medicine school on the incident and was provided by the chancellor of the university in an e-mail discussion about the incident organized by the Morality and Rights Committee of Science and Technology Workers under CAST, which we will introduce later in the method section.

been busy distancing themselves from it. What they are thinking about is not how to safeguard the legitimate rights and interests of clinicians but how to give an answer to superiors, and they punish the innocent clinicians quickly and severely to do so. This is the sorrow of the society. (From questionnaire of responder no. 1)

Some other authors thought that primary responsibility for the incident rested with the publishers and agencies, and that the punishments for the authors were unreasonable:

It is unreasonable to hold accountable the authors who were originally victims instead of the publisher in this incident. I painstakingly searched the literature, analyzed the data, and completed the article. Due to problems with my English, I found an agency. However, it is the authors who have been severely punished. I feel wronged. As an author, the retraction of the article is already a great loss. It is unreasonable. They should ask BMC how they will compensate for the authors' reputation loss when they made such a big mistake in the review process. Additionally, they should hold agencies accountable instead of looking at authors. (From questionnaire of responder no. 6)

The BMC retraction incident exhibited several typical characteristics of postproduction misconduct. For example, the articles were retracted because of fabricating peer reviews, which is a common display of postproduction misconduct. More important, some of the authors who were punished in this incident insisted that they were innocent. This is quite different from the traditional misconduct in which right is indisputably distinguished from wrong. Below, we explore how this happened.

3 Method

The first step in our research involved searching all the retracted Chinese articles in the BMC website and gathering some basic information from them, such as their titles, their abstracts, the journals in which they were published, and the names, e-mail addresses, and affiliations of the lead and corresponding authors. This information gave us an overview of the retracted articles in order to assess for commonalities in regard to the types of manuscripts that were retracted. For our reference, we also retrieved from the Internet such things as news reports, opinions, and analyses related to the incident. Some of these materials are of high value, which is discussed later in this article.

As previously mentioned, organizations including CAST and NSFC jointly launched an investigation after the incident. The investigation team visited the institutions with which the authors were affiliated and interviewed the authors and other personnel associated with twenty-five retracted articles to learn about their opinions on the incident and the publication details of these articles. They also interviewed the authors of another six articles by telephone. By means of the cooperative relationship between CAST and Tsinghua University on the research of scientific ethics, we were able to join the team to participate in several interviews as well as obtain the reports based on all interviews, which were not accessible to the public. We also got some important information through conversations with other members of the team during and after

the investigation. In addition, the Morality and Rights Committee of Science and Technology Workers under CAST organized an e-mail discussion about this incident, and most of the e-mails were provided to us. The committee is composed of senior experts with extensive knowledge of the research system. Many of the members are from the China Academy of Sciences and China Academy of Engineering.

We also investigated the agencies involved in the incident. While conducting interviews with the authors, five separate agencies were mentioned. Among them, two had well-established websites and provided consulting services. Thus, we collected information through online consultation. The contact number of another agency was provided by an author, which enabled us to contact this agency via phone. The remaining two agencies did not have websites, and the authors did not provide any valid contact information. This resulted in us being unable to contact these agencies for consultation. During the investigation of these agencies, another three agencies were mentioned, and we then investigated them as well. Furthermore, we retrieved the Administration of Industry and Commerce data for these agencies through the National Enterprise Credit Information Publicity System.

Finally, we also conducted an anonymous survey via e-mail. The survey focused on the authors' direct incentive to publish these articles and the role of agencies in the publication process. To enable the authors to express their thoughts fully, our questionnaire contains some open-ended questions about their opinions on the incident and the way in which it was handled as well. We sent the anonymous questionnaire to almost all of the lead and corresponding authors as well as some coauthors, with a total of 104 questionnaires being distributed. Out of the 104 questionnaires distributed, only 7 valid questionnaires were returned, resulting in a relatively low response rate. We have hypothesized that the chief reason for the low response rate, despite the questionnaire being anonymous, was that some authors still did not trust us because the incident was incredibly sensitive and could impact the careers of the individuals surveyed. Despite the low response rate, the authors who returned the questionnaires provided us with some key information for our analysis.

4 An Overview of the Retracted Articles

According to the retraction notes, forty of the forty-one Chinese articles were retracted because of fake peer reviews, while the remaining one was retracted due to the misuse of a patient's images.⁴ Analysis of the basic information shows that the forty retracted articles were published in eleven journals. Regarding impact factors, twenty-eight papers were published in journals with impact factors between one and two, while eleven articles were between two and three and only one article over three. We learned in the interviews that in the biomedical research field a journal is considered of low quality if its impact factor is below two, while it is considered of high quality if its impact factor is over three. Many authors also admitted that the quality of their articles was not very high. One even stated, "I am also curious about how it was accepted."⁵

⁴ All forty-one articles were retracted on the grounds of fake peer reviews at first, but BMC corrected the retraction note on one article later.

⁵ Stated by a clinician in one of the interviews that we conducted, at Jinan Military General Hospital on 2 June 2015.

The authors' information indicated that most of the authors were clinicians. Thirty-eight of their affiliations were hospitals, and thirty-four of those were AAA hospitals, that is, the top-level hospitals in China. As to the hospital type, twenty-four were university-affiliated hospitals, while nine were military hospitals, and the rest were regular hospitals. In terms of location, these hospitals were located in seventeen cities of fourteen provinces, and most of the cities were provincial capitals. Among them, twelve hospitals were located in Shanghai.

According to the investigation reports that were based on all of the interviews, among the thirty-one authors of retracted articles interviewed in person or by telephone, twenty-nine admitted that their manuscripts were submitted through agencies, while the remaining two insisted they themselves submitted the articles. The authors who submitted their articles through agencies all claimed that they were not sure whether agencies manipulated the peer-review process. The authors who submitted the articles by themselves denied manipulating the peer-review process at all. In addition, all of the authors interviewed insisted that their articles were completed independently. The authors whose articles were submitted through agencies claimed that they only entrusted language polishing and submission to agencies.

However, although there are indeed articles whose production process was not in question, the investigation found that a large portion of retracted articles also contained traditional misconduct of various forms. In the interviews, some authors' answers to the questions about the article were contradictory and full of errors.⁶ For example, an author insisted that he completed his article entirely by himself at the start of the interview. But when we asked him to repeat the analysis process in his article on site, he said he was not very familiar with the statistical analysis software used in the article. Eventually, he stated that the data analysis in his article was performed with assistance from the agency. In a separate interview with another author, it was revealed that the author allowed the agency to help produce the article as well. This author also insisted at the start of the interview that he completed his article entirely on his own. But when asked which databases were referred to, he could only identify one, whereas, in fact, several databases were referred to in his article. Finally, he said he only wrote part of the article while the rest was completed by the agency. This kind of scenario occurred repeatedly in the interviews, indicating that the agencies were involved in the production process of some retracted articles.

Furthermore, in a few cases there were two versions of contracts between authors and agencies.⁷ While the formal version stated that agencies should provide publication-related services, such as language polishing, the supplementary versions clearly stipulated that articles should be completed entirely by agencies. In addition, we learned in our investigation of agencies that they did not provide submission services separately. Instead, submission was attached to other services, such as article writing, modification, and polishing.

Given this information, we can see that postproduction misconduct and traditional misconduct became intermingled in this incident, and the common denominator was

⁶ This happened in both the interviews we conducted and those we did not. We learned about the latter through conversations with other members of the investigation team.

⁷ We learned about this information through conversations with other members of the investigation team as well.

the agencies. It is almost certain that they manipulated the peer-review process by fabricating peer reviews. What's more, they also got involved in the production of a large portion of retracted articles by helping the authors produce, modify, or polish their articles. In a few cases, the articles were even completed entirely by agencies. With this knowledge, it is imperative that we illuminate the role these agencies played in the BMC retraction incident.

5 The Role of Agencies in the Incident

As we mentioned above, agencies can be institutions or individuals. According to the investigation reports provided by CAST, among the twenty-nine retracted articles submitted through agencies, eighteen were submitted through institutions, while the remaining eleven were submitted through individuals. The nature of individuals, compared with that of institutions, is more difficult to recognize. For example, some authors whose articles were submitted through individuals claimed they were not acquainted with the individuals and they paid those individuals for the submission, while others stated the individuals were their friends and they did not pay those friends. Based on the information obtained, we can hardly determine the real relationship between authors and the individuals, or whether the individuals were affiliated with particular institutions. Thus, we focus on the institutions in this article, and we refer to institutions when we talk about agencies hereafter.

Based on the investigation reports, the authors of fifteen articles that were submitted through institutions provided the names of these institutions, while the authors of the other three articles claimed they had forgotten the institutions' names. In total, five institutions were mentioned. It was surprising that two of the five institutions shared the same contact person. In our investigation of these five institutions, another three were mentioned, and they were also subjects of our investigation. Among the eight agencies above, four had well-established websites, so we contacted them through online consultation. Two more agencies did not have well-established websites, but their contact information was provided by an author or another agency. Thus, we made telephone calls to contact them. The other two did not have websites, and no valid contact information was provided. As a result, we were unable to contact them to gather evidence through interviews.

We learned through online consultation and telephone calls that all agencies available provided services related to SCI (Science Citation Index) article publication, from subject selection, experimentation, and data analysis to article writing, modification, polishing, and submission. By accessing the National Enterprise Credit Information Publicity System, we found that all of the firms were officially registered. Their registered business scope mainly included technology development, counseling, and services in the areas of biological technology, pharmaceutical technology, and information technology. A surprising result was that four of the five institutions involved in this incident were closely connected. They shared the same legal representative or shareholder or, as mentioned earlier, contact person, indicating that they were likely influenced by the same group. Moreover, another two of the three institutions mentioned in our investigation of the five institutions above were also included in this agency circle. The remaining institution of the three described the prominent agency of

the circle as their main competitor, which takes most of the market share in Shanghai and adjacent areas.

The business model of these agencies was strikingly similar, no matter whether they belonged to the agency circle or not. The agencies' approaches to developing a customer base were manifold, including door-to-door marketing, distributing leaflets, e-mail advertising, forum posting, and offering consultation to potential customers through websites and telephone calls. In the BMC retraction incident, most authors had face-to-face communication with salespeople of the agencies, and some even made friends with them. Other authors communicated with salespeople through e-mail or QQ (a Chinese instant-messaging software). After reaching an agreement on the service contents, such as journal level, expected publication date, and fees, agencies would generally sign an entrustment contract with the customer and collect a deposit before starting the task. After the task was completed and all fees were paid, agencies could provide an invoice of all fees to the customer for reimbursement, and the invoice items usually were technical support or article polishing.

In terms of research capabilities, some agencies claimed that they had laboratory and research teams of their own. In fact, SCI-publication-related services were just additional business for them. Beyond that, they mainly provided services such as experiment outsourcing and biological data analysis. Other agencies claimed they did not have their own research powers and provided publication-related services only. At the same time, they emphasized that if experimentation was needed, they could have their business partner conduct those services. Regarding the submission process, all agencies claimed they did not have a cooperative relationship with publishers. However, due to their rich submission experience, they were able to guarantee that the article would be published within the time stipulated in the entrustment contract. Meanwhile, some agencies indicated that they could reach potential reviewers, thereby improving the acceptance rate.

The service fees of the agencies were very similar as well. For authors who chose to entirely entrust the publication of an article, from writing to submission, in a journal with an impact factor below one, the cost was approximately 40,000 RMB. While it would cost about 60,000 RMB if the impact factor of the journal was between one and two, and about 90,000 RMB if the impact factor was between two and three. All the service fees above refer to the costs of publishing articles when experimentation was not needed. If experimentation was necessary, the fees could be higher. For instance, publishing an article that needed experimentation in a journal with an impact factor between one and two entirely through agencies cost approximately 80,000 RMB. Since publishing an article in a journal with an impact factor above three was relatively difficult and generally required experimentation, it would cost more than 200,000 RMB. If original materials were provided, agencies would determine the fees after evaluating the materials and the demand of the customer. If a Chinese or English draft was provided, agencies would determine the fees based on the quality of the draft and the demand of the customer as well. If only language polishing was needed, the fees were approximately one RMB per word, while the fees for translation and polishing were about 1.6 RMB per word.

It is now clear that there was a mature market of publication-related services behind the BMC retraction incident. Researchers could get almost any publication-related services from this market. To a certain extent, the production process of a large portion

of articles in the BMC retraction incident bear a resemblance to “ghost writing,” which has been frequently reported by the media in recent years. Some STS scholars have conducted impressive studies on this issue as well (Lemmens 2004; Ngai et al. 2005; Götzsche et al. 2007). But we should note that in “ghost writing,” an article is generally produced without the participation of the “author.” While in this incident most of the authors who sought help from agencies also made a contribution to their articles, such as doing some of the analysis, providing original materials or drafts, and so on. A few cases in which the articles were produced entirely by agencies, would be considered “ghost writing.”

Another phenomenon that is similar to this case is “ghost management,” which was studied by Sergio Sismondo and other STS scholars (see, e.g., Mirowski and Horn 2005; Moffatt and Elliott 2007; Moynihan 2008; Sismondo 2007, 2008, 2009, 2015). Simply speaking, “ghost management” refers to the phenomenon that pharmaceutical companies and their agents control or shape multiple steps in the medical research, from the analysis of the results to the publication of the articles. Through “ghost management,” pharmaceutical companies play a large role in determining the message that articles convey, often in an effort to promote particular drugs to doctors. In both “ghost management” and the BMC case, articles are not completed entirely by the “authors.” The difference is that in “ghost management,” the “authors” whose names appear at the top of the article are selected to add a sense of independence and credibility to the article. While in our case the “authors” are those who need the authorship to meet the requirements of title assessment.

The existence of the market of publication-related services seriously tarnished the academic ecology of China. But at the same time it could be very risky for the researchers to have agencies help them produce articles, given that China has been taking academic misconduct very seriously in recent years, which can be highlighted by reviewing the severe punishments for the authors involved in other incidents. Therefore, it is surprising that researchers continued to seek services from these agencies. Thus a question might be, why did these researchers seek publication-related services from the market?

6 The Emergence of the Market of Publication-Related Services

In the interviews we conducted, the authors, most of whom were clinicians, provided three reasons for their seeking publication-related services from agencies. First, many emphasized that their workload was very heavy, and thus they did not have enough time to do research. This situation was confirmed by some studies (Zhang 2014; Wen et al. 2015), which found that Chinese clinicians worked an average of up to ten hours a day, and the better the hospitals were, the longer the clinicians had to work. Second, their English was usually intermediate, and therefore they were faced with language problems when publishing articles in English. Third, most of the authors lacked experience with publishing SCI articles. In fact, the articles involved in the incident were often the first SCI articles for many of them.

In addition to the difficulties above, the authors also spent a large amount of money to publish their articles. In addition to the fees paid to agencies, authors needed to pay

publication fees to the publishers as well, which were usually more than 10,000 RMB. Some authors paid the service and publication fees with their research funds, while others paid them out of their own pockets. For the latter, it was a heavy burden. One author in an interview said the money “was half of my annual salary.”⁸ Besides, as mentioned earlier, the authors had to take huge risks when they decided to have agencies help them. Therefore, a more important question arises: why did the authors publish these articles, to be exact, English SCI articles?

According to the investigation, the direct incentives of the authors to publish these articles included such things as title assessment, position promotion, grant application, and degree application. Among them, the most frequently noted was title assessment. It might be difficult for people in other countries to imagine how important professional title is to Chinese clinicians. In short, a clinician’s income, status, and reputation are all directly associated with his or her professional title. The extreme importance of professional title means that the title assessment system plays a prominent role in guiding clinicians’ behaviors.

In recent decades, the requirements of publication in the title assessment system had been increasing and creating systematic pressures for authors. The earliest document of the current title assessment system for clinicians in China is the *Health Professional Post Proposed Bylaw* (卫生技术人员职务试行条例), introduced in 1986. In *Bylaw*, publication is included but is optional rather than mandatory. For example, the requirements for a chief physician (a high-level professional title) include that “he/she is able to solve complex and difficult major technical problems, or has published scientific monographs, articles or experience summaries of a high level.” However, in *Assessment Requirements for Medium and High Level Professional Qualifications in Clinical Medicine* (临床医学中、高级技术资格评审条件“试行”), which was introduced in 1999, article publication had become a mandatory requirement for clinicians in medium- and high-level title assessment. For example, the requirements for an associate chief clinician (a medium-level professional title) include that “during working as an attending clinician, he/she has published at least two articles as the lead author on professional journals or has made reports in academic conferences of provincial level or above.”

What’s more, in practice the publication requirements for clinicians are even higher than those described above. For example, in a certain hospital the requirements for an associate chief clinician mandated that “he/she has published four articles as the lead author (among them, at least one was published in SCI journals and two in domestic core journals).”⁹ Besides, the emphasis on SCI article publication reaches a pathological level in the actual title assessment process. As an author said in the interview, “In title assessment, ten ordinary articles are less than one article on SCI.”¹⁰

Without a doubt, the storyline of the BMC retraction incident has become clear. As the professional title is very important to clinicians in China, they have to manage to meet any requirement of the title assessment system. In practice, this system put a great emphasis on SCI article publication. Therefore, clinicians who were faced with title assessment had to try their best to publish SCI articles, often with the pressure to publish these in English. But some clinicians had various difficulties in publishing

⁸ Stated by the same clinician as in footnote 5.

⁹ The specific requirements of various hospitals differ, but they all put a huge emphasis on publication.

¹⁰ Stated by an author in one of the interviews we conducted, at Ludong University on 4 June 2015.

English SCI articles on their own, leading to the emergence of a market of publication-related services. As a result, agencies got involved in the production of academic articles by helping clinicians organize, modify, or polish their articles, and then submit them. To make sure that articles can be published successfully, agencies fabricated peer reviews, which finally resulted in this retraction incident.

7 Structural Factors behind the BMC Retraction Incident: An STS Interpretation

After an analysis of the retracted articles, it does not appear that the research data was indeed fabricated. According to Biagioli, in postproduction misconduct the science reported in the article is usually not in question (Biagioli 2016). Instead, researchers mainly want to make their work more influential, by managing to publish in a more prominent journal or obtain more citations, so that they can obtain an elevated and respectable score on the metrics of scientific evaluation. But at the same time, researchers involved in postproduction misconduct are generally not aiming for the really high-profile publications, since that's too risky. The aim is to produce publications that are nearly invisible but that can give them the curriculum vitae that matches the performance metrics used by their academic institutions. This appears to be the case in the BMC retraction incident. Though the retracted articles in this incident are of poor quality and contain traditional misconduct since they were produced with the help of agencies, they are authentic. The articles were published in SCI journals, which were considered by the authors and their academic institutions to be better than the domestic journals in China. Meanwhile, the impact factors of these SCI journals were relatively low, indicating that these articles would not be widely read. If the articles were not retracted, the authors would have added a new publication to their curriculum vitae, which would help them meet the publication requirements of the title assessment system.

Biagioli argues that postproduction misconduct is more common in emerging countries, partly because academic institutions there place the most emphasis on metrics (Biagioli 2016). These institutions are usually not the world's leading institutions, but they are trying to break into the top rank. This is roughly the case in the BMC retraction incident as well. The scientific enterprise of China has been growing rapidly in recent years, and it seems true that postproduction misconduct is more common in China than in other countries. One study even indicates that more than half of the articles retracted due to fake peer reviews were written by Chinese researchers (Qi et al. 2016). In addition, the hospitals and associated universities in this case are good but not the best ones in China, and they might want to break into the top rank.

Our case has some characteristics that advance Biagioli's argument as well. First, in other postproduction misconduct cases, peer reviews were fabricated by the authors themselves, while in our case peer reviews were fabricated by the agencies. Although BMC stated that they were not sure whether the authors were aware of agencies' attempts to manipulate the peer-review process of their articles, we are almost certain that the authors did not know. Because if they knew, they would not spend large amounts of money on the production of these articles anymore. The role that agencies played in the BMC retraction incident indicates that it was common for clinicians to have agencies

help produce and submit their articles. It can be verified by the subsequent retraction incidents regarding Springer, Elsevier, and Nature, in which the authors were also mainly clinicians. We also do not know the magnitude of the issue without knowing how many more articles like this have not been retracted. This leads to the second characteristic of this case, that is, the systematism.

The clinicians in these incidents worked at hospitals in different areas of China, but the direct incentive for their committing academic misconduct was roughly the same: the title assessment system. Therefore, it is no longer adequate to attribute the misconduct in these cases to the authors' lack of academic morality. Of course, the authors of these articles committed traditional misconduct when they had agencies help them produce their articles. Agencies that fabricated peer reviews should obviously be blamed. But the title assessment system played a more fundamental role in leading to this incident. The publication requirements raised by the title assessment system were too difficult for some clinicians to meet on their own, and thus they *had* to seek help from the agencies, giving them an opportunity to get involved in the production and manipulate the peer-review process. The BMC retraction incident provides an important perspective in understanding the incentive of academic misconduct over and above individual factors, that is, the systematic effect of metrics-based academic evaluation.

To prevent this type of widespread misconduct from happening again, systematic precautions need to be taken. Of course, researchers should be educated to have a stronger sense of academic morality. But more important, some adjustments to the title assessment system are needed. We think the most urgent adjustment should be the distinction between clinicians and medical researchers, which is also suggested by the senior experts in the Morality and Rights Committee of Science and Technology Workers under CAST. There is no distinction between clinicians and medical researchers currently in China, thus clinicians have to do both medical work and academic research. Through distinction, clinicians can focus on diagnosis and treatment, while academic research is the main task of medical researchers. Of course, it is not an absolute distinction, that is, clinicians can also do some research, and medical researchers may be scheduled to do some amount of diagnosis and treatment each week. But the aim of the distinction is to make them focus on their main tasks. Accompanying the distinction are adjustments to the metrics of evaluation. Specifically, clinicians should be evaluated more by their performance of diagnosis and treatment, though publication can also be an indicator. Similarly, medical researchers should be evaluated more by their publications, while diagnosis and treatment can play a less important role. Furthermore, associated systems, such as the training system and rewarding system, need to be adjusted accordingly as well.

As for the agencies, supervision should be strengthened, and their role should be clearly defined. They can be helpful to the scientific enterprise if their scope of business is limited to language services and other legitimate research services. In addition, publishers should share more information when this kind of postproduction misconduct occurs. In this incident, BMC did not share the details of how the peer reviews were fabricated, even with the authors. They might consider it as proprietary information, but it is not helpful in dealing with the incident. Instead, the exposure of more information by publishers can help prevent it from happening again.

References

- Biagioli, Mario (2016). "Watch Out for Cheats in Citation Game." *Nature* 535, no. 7611: 201 (<http://escholarship.org/uc/item/0b05plj6>).
- Götzsche, Peter C., Asbjørn Hróbjartsson, Helle Krogh Johansen, and Mette T. Haahr (2007). "Ghost Authorship in Industry-Initiated Randomized Trials." *PLoS Med* 4, no. 1: e19.
- Lemmens, Carl (2004). "Pharma Goes to the Laundry: Public Relations and the Business of Medical Education." *Hastings Center Report* 34, no. 5: 18–23.
- Li, Jingbo, and Ya Li 李靖波、厉亚 (2011). "Academic Misconduct: Connotation, Category, Origin, and Control" 学术不端:内涵、类别、根源与治理". *Science-Technology and Publication* 科技与出版 8: 80–82.
- Mirowski, Philip, and Robert Van Horn (2005). "The Contract Research Organization and the Commercialization of Scientific Research." *Social Studies of Science* 35, no. 4: 503–48.
- Moffatt, Barton, and Carl Elliott (2007). "Ghost Marketing: Pharmaceutical Companies and Ghostwritten Journal Articles." *Perspectives in Biology and Medicine* 50, no. 1: 18–31.
- Moynihan, Ray (2008). "Key Opinion Leaders: Independent Experts or Drug Representatives in Disguise." *British Medical Journal* 336: 1402–03 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2432185/>).
- National Academy of Sciences, National Academy of Engineering, and Institute of Medicine (1992). *Responsible Science, Volume I: Ensuring the Integrity of the Research Process*. Washington, DC: The National Academies Press (<https://doi.org/10.17226/1864>).
- Ngai, Stephanie, Jennifer L. Gold, Sudeep S. Gill, and Paula A. Rochon (2005). "Haunted Manuscripts: Ghost Authorship in the Medical Literature." *Accountability in Research* 12, no. 2: 103–14.
- Qi, Xingshun, Han Deng, and Xiaozhong Guo (2016). "Characteristics of Retractions Related to Faked Peer Reviews: An Overview." *Postgraduate Medical Journal* 93, no. 1102 (<http://dx.doi.org/10.1136/postgradmedj-2016-133969>).
- Sismondo, Sergio (2007). "Ghost Management: How Much of the Medical Literature Is Shaped behind the Scenes by the Pharmaceutical Industry?" *PLoS Med* 4, no. 3, 286.
- Sismondo, Sergio (2008). "Pharmaceutical Company Funding and Its Consequences: A Qualitative Systematic Review." *Contemporary Clinical Trials* 29, no. 2: 109–13.
- Sismondo, Sergio (2009). "Ghosts in the Machine: Publication Planning in the Medical Sciences." *Social Studies of Science* 39, no. 2: 171–98.
- Sismondo, Sergio (2015). "Key Opinion Leaders: Valuing Independence and Conflict of Interest in the Medical Sciences." *Oxford Scholarship Online*: 31–48 (DOI:10.1093/acprof:oso/9780199689583.003.0002).
- Wen, Jin, Tianyou Hao, and Xiuying Hu 文进、郝天佑、胡秀英 (2015). "Study on the Situation of Chinese Clinicians' Work Burden 中国医生工作负荷的现状研究." *Chinese Journal of Evidence-Based Medicine* 中国循证医学杂志 2: 133–36.
- Zhang, Xinqing 张新庆 (2014). "Deep Comparison of Clinicians' Career Pressures 医生职业压力纵深比较." *China Hospital CEO* 中国医院院长 7: 47–49.

Junhui Han is a PhD candidate of sociology of science at the Institute of Science, Technology, and Society, Tsinghua University, Beijing. His professional interests include sociology of science and computational social science.

Zhengfeng Li is professor of sociology of science and deputy dean of the School of Social Sciences, Tsinghua University, Beijing. He also chairs the Specialized Committee of Sociology of Science in the Chinese Sociological Association. His professional interests include sociology of science, philosophy of science and technology, and policy of science and technology.