Prometheus, 2015
Vol. 33, No. 3, 215–233, http://dx.doi.org/10.1080/08109028.2015.1114745

RESEARCH PAPER

Dishonest conformity in peer review

Sotaro Shibayama[a,b]* and Yasunori Baba[a]

[a]*Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan;* [b]*Department of Technology Management for Innovation, University of Tokyo, Tokyo, Japan*

Honesty in scientific publication is critical for scientific advancement, but dishonesty is commonly and increasingly observed in misconduct and other questionable practices. Focusing on dishonest conformity in peer review, in which authors unwillingly obey referees' instructions in order to have their papers accepted even if the instructions contradict the authors' scientific belief, the current study aims to investigate the determinants of dishonesty. Drawing on survey data of Japanese life scientists, this study shows that the conflict between authors and referees in peer review is common. A majority of scientists follow referees' instructions rather than refute them. The results suggest that conformity occurs more frequently (1) in biology than in medicine and agriculture, (2) when authors are in strong scientific competition, (3) if authors are associate professors rather than full professors, (4) if authors have no foreign research experience, and (5) in low-impact journals rather than in medium-impact journals.

Introduction

The progress of science relies on intellectual honesty (Barber, 1952; Zuckerman, 1977; Shamoo and Resnik, 2003). Scientific discoveries are disseminated primarily by means of publication, and the stock of publications is the basis of the cumulative advancement of science (Merton, 1973). Without the premise of honesty, science could not have been developed as efficiently as it has been. Dishonest publication compromises the reliability of the knowledge base; subsequent studies can be rendered baseless, which is a serious waste of resources for the scientific community. So, honesty is at the core of research integrity. The *Singapore Statement on Research Integrity*, for example, maintains 'honesty in all aspects of research' as one of its four principles,[1] and the grant policy of the National Institute of Health (NIH) refers to honesty as one of its four shared values.[2]

Nevertheless, breaches of honesty are not uncommon, ranging from obvious misconduct, such as fabrication, to other questionable practices, such as authorship abuse (Martinson *et al.*, 2005; Martin, 2013). Fanelli (2009) estimates that approximately 2% of scientists have fabricated or falsified data or results at least once. Azoulay *et al.* (2012) evaluate the negative impact of dishonesty by showing that after a paper is retracted for misconduct, other papers citing the retracted paper suffer a 5–10% decline in the rate of citations they receive. To avoid such grave consequences, preventive measures, such as journal ethics policies, have been implemented (Resnik and Master, 2013), though their effectiveness has been questioned (Anderson *et al.,* 2007a).

*Corresponding author. Email: shibayama@00.alumni.u-tokyo.ac.jp
Accepting editor: Stuart Macdonald

For greater research integrity, we need a better understanding of why and how scientists might deviate from the norm of honesty. Although intense competition and pressure for publication are likely causes (De Vries *et al.*, 2006; Anderson *et al.*, 2007b, 2007c), empirical evidence is limited. Obvious dishonesty is observed only infrequently, although increasingly, and is difficult to measure since it tends to be concealed (Van Noorden, 2011). These are particularly troublesome for the 'big three' in misconduct (fabrication, falsification, and plagiarism), on which prior literature has primarily focused. This study attempts to address these issues and offers an empirical account of the mechanism behind dishonesty. To this end, we draw on so-called 'questionable research practices', practices which are considered only minor deviations from honesty and are more commonly observed than the big three (Broad, 1981; NAS, 1992; Martinson et al., 2005; De Vries et al., 2006; John et al., 2012; Martin, 2013). Among many forms of questionable practices, this study focuses on dishonest conformity in peer review, in which authors unwillingly obey referees' revision instructions in order to have their papers accepted even if obeying is in conflict with their scientific belief (Frey, 2003; Tsang and Frey, 2007). This practice can bias and impair the published knowledge base. Though dishonest conformity has long been recognized, empirical evidence is scant. This study illustrates the process of peer review in depth, drawing on a unique questionnaire survey of Japanese life scientists, and investigates the determinants of this specific form of dishonesty. The remainder of this paper summarizes the literature on research integrity in general, and discusses dishonesty in the peer review process specifically. It then explains the empirical setting, describes our data on the peer review process, and analyzes the determinants of conformity in peer review.

Research integrity and dishonesty

Honesty plays a crucial role in the science system because the key mechanisms sustaining science, such as journal peer review, rely on the assumption that scientists are intellectually honest (Zuckerman, 1977). Scientists do not usually bother confirming that other scientists are truthful. Although the latest technologies, such as software for detecting image manipulation, may help detect traces of dishonesty (Van Noorden, 2011), carefully worked-out manipulations cannot be easily detected, and proving someone is lying is prohibitively costly and perhaps impossible. Instead of implementing an elaborate policing system, the science system has been deterring wrongdoing by heavy sanctions. If caught, wrongdoers can be ostracized from the scientific community and possibly prosecuted. These sanctions are formalized in some countries, but are informal in other countries. Some of the former countries have special investigation organizations, such as the Office of Research Integrity (ORI) in the US, the UK Research Integrity Office, and the Danish Committee on Scientific Dishonesty (Resnik and Master, 2013). Preventive measures have been introduced, such as research ethics policies articulated by professional associations, academic journals, and funding agencies (Macrina, 2007; Bosch et al., 2012), and responsible conduct of research (RCR) education programs provided by research organizations (Resnik and Master, 2013). Even so, uncovered incidents of misconduct have been increasing. Using bibliometric data, Fang *et al.* (2012) show that about two-thirds of retractions are attributable to misconduct and that the frequency of misconduct has increased by a factor of 10 since 1975 in the fields of biomedicine and life sciences. Similarly, Grieneisen and Zhang (2012) suggest that 47% of retractions are caused by publishing misconduct (e.g. plagiarism, authorship issues), 20% by research misconduct (e.g. falsification, fabrication), and 42% by questionable data or interpretation, and that the number of retractions increased by a factor of 20 from 2001 to 2010.

Prior literature and policy actions regarding research integrity have focused primarily on the big-three misconduct. For example, the definitions of misconduct of the NIH and the National Science Foundation include only fabrication, falsification, and plagiarism.[3] However, other forms of questionable practices, such as redundant publication and authorship abuse, have long been recognized (Broad, 1981; Marusic et al., 2011; Martin, 2013). As early as 1992, the National Academy of Sciences (NAS) called attention to questionable practices that may not meet the definition of misconduct, but that can erode confidence in research integrity, violate scientific traditions, affect scientific conclusions, waste time and resources, and weaken the education of scientists (NAS, 1992). Compared with the big three, these practices are common (Martinson et al., 2005) and probably detectable with limited bias even by self-reporting surveys because they are not really considered misconduct. Taking advantage of these features, Martinson et al. (2005) identify several questionable practices and estimate their frequencies. For example, 27.5% of NIH-funded scientists self-reportedly engaged in 'inadequate record keeping related to research projects', 15.3% 'dropped observations or data points from analyses based on a gut feeling that they were inaccurate', and 7.6% 'circumvented certain minor aspects of human-subject requirements' at least once within three years. One stream of literature attempts to understand the mechanisms behind these practices. For example, competition (De Vries et al., 2006), seniority (Martinson et al., 2005), and perceived inequality in resource allocation (Martinson et al., 2006) are seen as responsible for some questionable practices. Still, most of the literature on dishonesty remains mostly conceptual or descriptive.

Dishonesty in peer review

Dishonesty can occur in various phases of scientific research. For example, falsification and fabrication occur in producing and analyzing data, plagiarism in writing, and authorship abuse in publication. Our focal practice, dishonest conformity, occurs in the phase of peer review. Peer review is a critical device to establish the scientific knowledge base. It is a process in which submitted papers are carefully scrutinized, necessary revisions are suggested, and acceptance or rejection for publication is recommended by expert referees. Though peer review has been used for a long time, many scientists believe that the current peer review system needs improving (Ellison, 2002a, 2002b). Peer review is based on volunteer referees, but their incentives are often misaligned (Pitsoulis and Schnellenbach, 2012). Referees have little reason to provide accurate and timely reports, which delays the review process (Ellison, 2002a) and results in erroneous decisions (Coupe, 2004). Referees may even be tempted to steal the results of submitted papers (Hagstrom, 1974). Taking advantage of veto power, referees might provide dishonest review comments. For example, they could write an unfairly negative report on a rival's paper (Ellison, 2002a) and require revisions favorable to their own research.

Editors' incentives are also somewhat misaligned. Since editors' reputations can be improved by raising the status of their journals, editors are unwilling to accept low-impact results, such as negative data and replication studies (Csada et al., 1996;

Fanelli, 2010), which can bias publishable findings. Editors also act dishonestly. For example, Wilhite and Fong (2012) find that editors of social science journals often coerce authors to cite the editors' papers to improve the editors' citation scores. In an attempt to address these problems, alternative systems have been proposed (Prufer and Zetland, 2010), and some of them have been actually implemented, such as post-publication open peer review (Ietto-Gillies, 2012). Nevertheless, the traditional peer review remains the norm.

In the midst of such an incentive structure, authors may also resort to questionable practices. For example, they might omit inconvenient data from publication because journals tend to reject imperfectly consistent results. Dishonest conformity, which is the focus of this study, is another questionable practice observed in the peer review process. We define dishonest conformity as authors unwillingly following referees' instructions and revising their papers in order to have the papers accepted even if doing so is in conflict with the authors' scientific belief. Describing this practice, Frey (2003) coined the term 'academic prostitution'. Given the veto power of referees and editors, Frey (2003) contends that 'authors only get their papers accepted if they intellectually prostitute themselves by slavishly following the demands made by anonymous referees'. This is not to say that referees and editors are always exploitative, but authors often encounter the dilemma of choosing whether to obey referees or to risk rejection by not following referees (Frey, 2003; Tsang and Frey, 2007). In fact, coercive citation (if coercively cited papers are scientifically irrelevant) is a form of dishonest conformity.

Of course, the scientific belief of authors can turn out to be false, which is the very reason why peer review is needed. Scientific consensus is usually not established by peer review alone, but is reached through scrutiny by the scientific community after publication (David, 1998). Thus, reporting what happens to be false is not a norm violation, but reporting what authors do not believe is. Ideally, when authors disagree with referees, they should make every effort to convince the referees. If authors spare this effort only to avoid rejection or to expedite publication, they essentially send out a message that they do not really believe. When this occurs, we regard it as a breach of the honesty norm. In addition, conformity in peer review is not necessarily attributed to dishonesty, because authors conform to referees' instructions for many reasons. In the following empirical analyses, we attempt to control carefully for relevant peer review conditions to highlight authors' dishonesty.

Methods and data

Context of the Japanese science system

This study draws on empirical data from the field of life sciences in Japan. Research integrity has only recently become an issue in Japan (Matsuzawa, 2013), although Japan is ranked third after the US and Germany in the frequency of retracted publications (Fang et al., 2012). National surveys reveal that the Japanese science system used to have very poor preventive measures against scientific misconduct. For example, only 97 of 838 research associations (12%) had an ethics code before 2004, and only 11% of all universities had any written misconduct policies before 2006 (Ishibashi and Ohtake, 2009). Responding to this situation, the ministry of science and education published a guideline for research misconduct in 2006 (MEXT, 2006), following which universities set institutional guidelines. As of 2008, 71% of all universities had some guidelines (Ishibashi and Ohtake, 2009). Thus far, prevention and enforcement measures have been left up to individual universities. Although the Science Council of Japan (SCJ) has advised that independent organizations responsible for research integrity be established (SCJ, 2013), the scientific community has not yet followed the recommendation.

A few characteristics of the Japanese science system are worth mentioning. The Japanese university system consists of three types of universities offering four-year undergraduate programs. As of 2014, it has 86 national, 92 regional (of prefecture or city), and 603 private universities.[4] Among the three categories, national universities are the main player of scientific research, whereas most private universities are education-oriented. Among the national universities, the top seven (Tokyo, Kyoto, Osaka, Tohoku, Nagoya, Kyushu, and Hokkaido) are designated 'pre-imperial colleges' and have been enjoying exceptionally prestigious status in both research and education (Kneller, 2007). Japanese universities have a three-level promotion system with full professors at the top, followed by associate professors, and then by assistant professors or lecturers. Before becoming an assistant professor or lecturer, Ph.D. graduates usually have a few years of postdoctoral experience. Life science research is usually conducted in a laboratory with a full or associate professor as the lab head in charge of junior faculty members (assistant professors and lecturers), postdocs, and students (Shibayama et al., 2015). Unlike American universities, junior faculty members are often under the supervision of lab heads. International mobility is rather limited, except for temporary visits (mostly to the US) (Lawson and Shibayama, 2014). The population of foreign-born scientists working in Japan is extremely small, as is that of Japanese scientists working abroad (5% and 3%, respectively, as of 2011) (Franzoni et al., 2012). The evaluation system used to be fairly free from extrinsic incentive mechanisms, but policymakers have introduced reforms and these have emphasized competition and merit-based evaluation since the 1990s (Kneller, 2007). This transition has affected younger scientists most strongly (Shibayama and Baba, 2015).

Data

For empirical analyses, this study draws mainly on data from Japanese life science professors. We conducted a survey as part of a research project regarding scientific publication, which covers not only peer review, but also various other aspects of publication. The project began with interviews of 21 professors. Each interview took between one and two hours. We investigated the interviewees' publication records in advance and asked a series of questions about their publication strategies, the impact of recent policies on their publication practices, and so forth. Based on the interview results, a questionnaire instrument was designed.

The survey sample was selected with four criteria: scientists who (1) published as corresponding author at least one paper in 2012–13 in life science journals registered in the Web of Science (WoS);[5] (2) are affiliated with life science departments in Japanese national universities;[6] (3) are full or associate professors and so have authority in decision-making in publication; and (4) are Japanese.[7] As we did not have a comprehensive list of Japanese professors, we first collected the information of WoS publication authors. We found 26,886 papers and identified 13,877 unique corresponding authors satisfying conditions (1) and (2). From this sampling frame, we selected 1,700 authors by employing a stratified sampling strategy with the journal impact factor (JIF) of papers and the rank of the institutions to which corresponding authors were affiliated. JIFs were split into three groups: top 10%, middle 70%, and bottom 20%. Affiliations were stratified into three tiers: Tier 1 is the seven pre-imperial universities; Tier 3 is the bottom half of all universities in publication performance; and Tier 2 is the rest.[8] After the sampling, we examined the authors' employment status as of 2013 (using public information) and removed those who did not satisfy criteria (3) and (4). The final sample consisted of 777 professors. We mailed the questionnaire in November 2013 and collected 358 responses (response rate = 46%) after two waves of requests.

The respondents were in the fields of biology (53%), medicine (26%), and agricultural sciences (21%); in the university strata of Tier 1 (33%), Tier 2 (32%), and Tier 3 (35%); and in the publication JIF strata of the top (29%), middle (43%), and bottom (27%). Fifty-five percent were full professors, and 45% were associate professors. Six percent were female and 94% were male.[9]

Measures

As potential determinants of conformity, we prepared several variables for individual and institutional factors as well as peer review conditions. The individual and institutional factors include the following 13 variables. According to the sampling frame, dummy variables were prepared for three scientific fields (agriculture, biology, and medicine) and three university tiers (1, 2, and 3). *MD* is a dummy variable assigned one if a respondent's first degree was from a medical school. To be precise, the field of medicine includes both clinical and non-clinical research; the latter can be conducted by non-MD scientists. *AP* is a dummy variable assigned one for associate professors and zero for full professors. *Age* is the age of respondents at the time of the survey. *Female* is a dummy variable assigned one for women and zero for men. We measured the number of competitors in the respondents' research field with a five-point scale: (1) none, (2) 1–2, (3) 3–5, (4) 6–10, and (5) 11 or more (*#competitors*). We measured length of research experience outside Japan with a six-point scale: (1) none, (2) half a year or less, (3) one year, (4) two years, (5) three years, and (6) four years or more (*foreign experience*). To measure the scientific performance of the respondents, we summed up the citation counts of each respondent's papers published in 2009–13 drawing on WoS and took its logarithm [*ln(#Times cited)*]. Measures for peer review conditions are explained in detail in the next section.

Description of conformity in peer review

In the questionnaire, we needed to identify the respondents' peer review experience and whether they had unwillingly followed referees' instructions. To this end, we focused on one paper that each respondent had submitted as either a leading or corresponding author over the previous year. On average, our respondents had submitted 4.8 papers, of which 3.2 were to be revised following peer review. Twenty-one percent of our respondents had no paper that needed revision and were dropped from the following analyses. We asked the rest of the 284 respondents whether they had received a revision instruction that contradicted their scientific belief, and 136 respondents (48%) answered that they had had such an experience at least once during the year. For these 136 respondents, we surveyed specific instances of the peer review process.[10] In what follows, we describe the key variables. Table 1 presents the descriptive statistics and correlation matrix of all variables.

Reaction to revision instructions. We first asked how the respondents reacted to the revision instructions [Figure 1(A)]. We found that 63% of the respondents followed inconsistent instructions. The remaining 37% did not follow the referees' instructions; 19% resubmitted the manuscript to the same journal, 16% submitted the manuscript to a different journal, and 2% gave up publishing the paper at all. We coded a dummy variable one for the first reaction and zero for the other three reactions (*conformity*). To be precise, observed conformity may or may not be attributed to dishonesty.

Scientific value. Regardless of the respondents' reaction, we asked how they perceived the value of the overall review comments by asking 'How did you think it would affect the scientific value of your paper to follow the instructions completely?' with a five-point scale ranging from 1: negatively to 5: positively (*perceived value*). While 45% of the respondents positively appreciated the referees' instructions, 15% perceived a negative impact, and 40% thought that revision would make no scientific difference [Figure 1(B)]. While nearly half thought that the peer review improved the value of publication, the other half regarded it as a waste of time with either no meaningful impact or negative impact. Although negative perception is associated with a lower likelihood of conformity, 47% of negatively perceived instructions still met with conformity.

Cause of inconsistent instruction. To examine possible causes of inconsistency, we asked 'Why do you think the referees gave such an instruction [that is inconsistent with your belief]?' with three options [Figure 1(C)]. While 49% of the respondents admitted that the authors themselves were responsible, for example, because they had not been sufficiently clear, 46% blamed referees for their lack of understanding or relevant knowledge. In addition, 15% answered that a competitive relationship between referees and authors was the explanation for inconsistency. For each reason, we prepared a dummy variable (*author's fault*, *referee's fault*, and *competition between author and referee*), coded one, if applicable. Not surprisingly, authors' own fault is positively correlated with conformity, while blaming referees is negatively correlated.

Type of instruction. We inquired into types of revisions required and found that 68% of respondents were instructed to add data or experiments,[11] 60% to change the interpretation of results, and 12% to modify hypotheses or research questions [Figure 1(D)]. Among the respondents who were advised to add data or experiments, 36% considered completely following the instruction technically infeasible. Thus, we prepared a dummy variable coded one for such cases (*technical infeasibility*), which is negatively correlated with conformity, as expected.

Prestige of journals. We asked the name of the specific journal which gave inconsistent instructions. Only 58% of our respondents answered this question, possibly

Table 1. Descriptive statistics and correlation matrix

Variable	Mean	S.D.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Conformity	0.62	0.49	0.00	1.00																					
2. Perceived value	3.28	0.89	1.00	5.00	***0.23***																				
3. Author's fault	0.48	0.50	0.00	1.00	***0.39***	***0.35***																			
4. Referee's fault	0.46	0.50	0.00	1.00	***−0.20***	***−0.37***	***−0.32***																		
5. Competing of author & referee	0.15	0.36	0.00	1.00	−0.10	−0.11	***−0.31***	−0.11																	
6. Technical infeasibility	0.27	0.44	0.00	1.00	***−0.45***	0.02	−0.17	0.01	0.12																
7. Low JIF	0.24	0.43	0.00	1.00	0.16	0.01	0.19	−0.01	***−0.32***	0.03															
8. Middle JIF	0.60	0.49	0.00	1.00	−0.05	−0.09	−0.12	0.09	0.12	−0.22	***−0.68***														
9. High JIF	0.16	0.37	0.00	1.00	−0.11	0.11	−0.06	−0.11	0.21	***0.26***	***−0.24***	***−0.54***													
10. Tier 1	0.30	0.46	0.00	1.00	−0.08	−0.11	−0.16	0.09	***0.18***	0.06	−0.22	0.09	0.14												
11. Tier 2	0.33	0.47	0.00	1.00	0.12	0.17	0.09	−0.08	−0.10	0.01	0.04	−0.10	0.08	***−0.46***											
12. Tier 3	0.37	0.48	0.00	1.00	−0.04	−0.07	0.06	−0.01	−0.08	−0.07	0.17	0.01	−0.21	***−0.50***	***−0.54***										
13. Biology	0.51	0.50	0.00	1.00	0.12	−0.08	0.02	−0.13	0.13	−0.12	−0.14	0.24	−0.15	***0.24***	−0.12	−0.12									
14. Agriculture	0.16	0.37	0.00	1.00	−0.17	0.07	−0.05	0.15	−0.12	0.15	0.20	−0.10	−0.10	−0.13	−0.11	***0.24***	***−0.44***								
15. Medicine	0.33	0.47	0.00	1.00	0.01	0.03	0.02	0.02	−0.05	0.01	−0.01	−0.17	0.23	−0.15	***0.21***	−0.06	***−0.72***	***−0.31***							
16. MD	0.28	0.45	0.00	1.00	0.14	−0.05	0.00	−0.04	0.11	−0.03	−0.24	0.05	0.21	−0.16	0.08	0.07	***−0.44***	***−0.27***	***0.67***						
17. AP	0.41	0.49	0.00	1.00	***0.24***	−0.09	***0.21***	0.02	0.08	−0.12	0.10	−0.05	−0.04	−0.06	0.17	−0.10	−0.06	−0.13	0.17	0.06					
18. Age	52.19	7.25	38.00	66.00	−0.02	0.11	−0.01	−0.08	−0.04	0.02	−0.05	−0.06	0.14	0.14	***−0.24***	0.10	−0.08	0.16	−0.03	0.02	***−0.51***				
19. Female	0.07	0.25	0.00	1.00	−0.06	−0.12	0.01	−0.04	−0.02	0.07	−0.02	0.00	0.03	−0.10	0.17	−0.06	0.06	−0.12	0.02	−0.02	***0.19***	−0.12			
20. ln(#Times cited)	4.13	1.47	0.00	7.27	−0.06	0.08	−0.11	0.00	0.06	−0.07	***−0.34***	0.14	0.21	***0.31***	−0.01	***−0.28***	0.05	***−0.41***	***0.27***	***0.31***	−0.17	−0.04	−0.06		
21. #Competitors	2.90	1.21	1.00	5.00	0.12	−0.05	0.02	0.01	0.15	−0.03	−0.22	0.11	0.11	−0.10	***0.22***	−0.12	−0.04	0.05	0.00	***0.19***	0.00	−0.03	0.16	0.16	
22. Foreign experience	2.25	1.52	0.00	5.00	−0.14	0.16	−0.04	−0.04	0.12	0.10	***−0.26***	0.06	0.21	−0.08	0.11	−0.02	−0.02	−0.18	0.16	***0.26***	−0.10	0.07	−0.02	0.16	***0.20***

Note: Bold italic: p<0.05. N=120 (except for variables 7–9). Due to missing variables, 16 are dropped from 136 respondents.

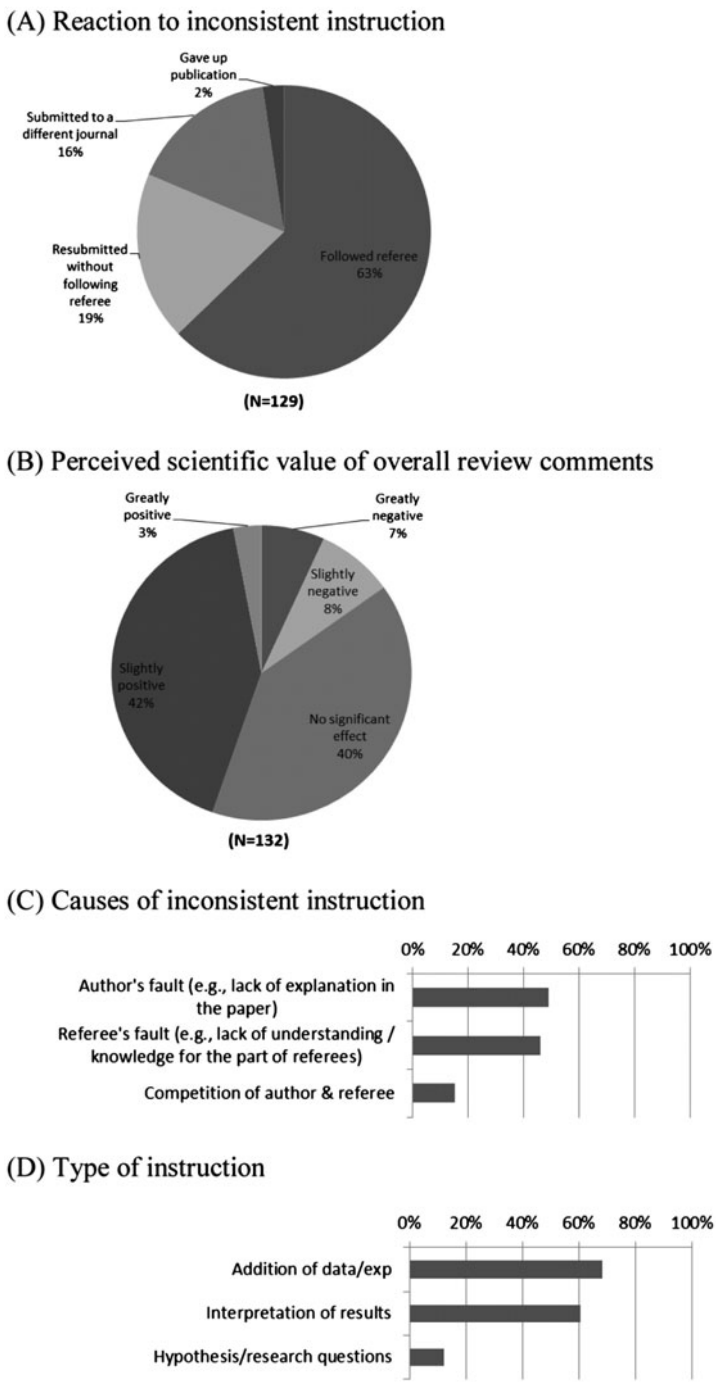


Figure 1. Description of conformity

because writing down a journal name is cumbersome. With the available data, we analyzed how journal prestige affects the decision of authors. To this end, we drew on JIF since it is a popular journal ranking metric in the life sciences (McAllister et al., 1980).[12] On the basis of our interviews and the literature (Shibayama and Baba, 2015), we transformed JIFs into three categories: low (JIF<2), high (JIF>8), and middle (JIF between 2 and 8) and prepared corresponding dummy variables (*low JIF*, *high JIF*, and *middle JIF*).

Result of peer review. We inquired into the result of peer review. We examined the result of publication for conformers and non-conformers respectively. While 93% of conformers succeeded in publishing in the journal giving inconsistent instructions, 48% of non-conformers eventually had to change journals [Figure 2(A)]. Interestingly, even when the respondents did not follow the inconsistent instructions and resubmitted the manuscript to the same journal, 70% had their resubmission accepted whereas 12% were instructed to revise once again and only 18% had their papers rejected [Figure 2(B)].[13] Thus, not following referees' instructions may seem risky, but immediate rejection does not often result.

Opinion about honesty. To examine whether the observed conformity in peer review is attributable to dishonesty, we additionally inquired into subjective opinions about honesty in publication (Figure 3). Based on the interviews, we designed two questionnaire items asking the extent of agreement on statements: (A) 'Explaining the complex nature perfectly consistently is difficult, so not reporting results inconsistent to author's claim is allowed to some extent', and (B) 'When scientists find results that are inconsistent with their previous paper, they should be reported' with a five-point scale from 1: disagree to 5: agree. Rather low agreement on the former (14%) and high agreement on the latter (74%) are interpreted as a high level of honesty in the respondents. As expected, measure (A) is found to be positively correlated with conformity, though insignificantly (r=0.14, p=0.13), and measure (B) significantly negatively correlated (r=–0.20, p<0.05), suggesting that conformity in peer review is associated with authors' honesty.

Prediction of dishonest conformity in peer review

Conformity in peer review can occur for various reasons, some of which might be irrelevant to honesty. For example, as shown in the previous section, technical feasibility should affect whether authors can or cannot conform, which has nothing to do with honesty. To investigate the determinants of dishonesty, therefore, we need to exclude the influence of such factors. To this end, we use regression analyses to control for peer-review conditions in order to predict the likelihood of conformity by individual and institutional factors. Table 2 presents four models employing the whole sample (Model 1) and subsamples (Models 2–4). Since all models show qualitatively similar results, we concentrate our explanation on Model 1.

Peer review condition

Technical feasibility is obviously a strong deterrent (b=–2.63, p<0.001). Since technical infeasibility is an inevitable reason for not conforming, Model 2 drops

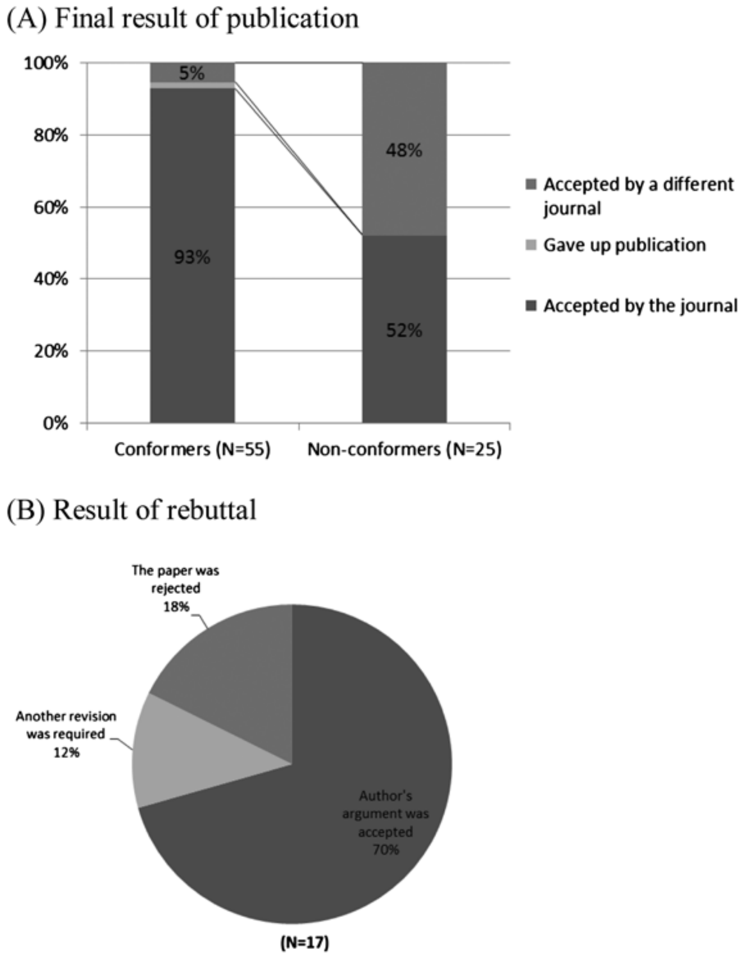
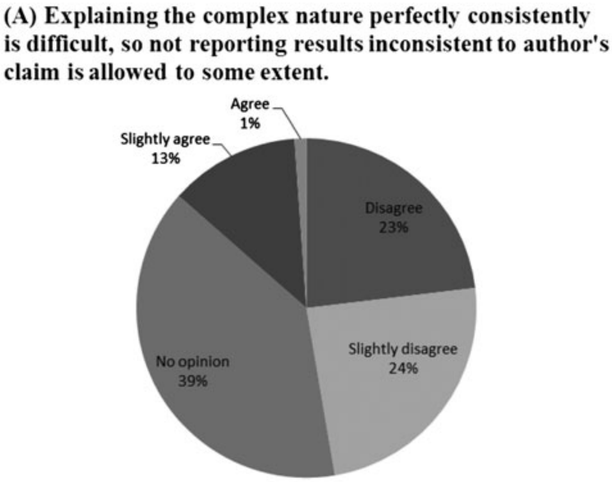


Figure 2. Result of peer review
Note: Non-respondents to the focal questions and respondents whose peer review result had not been known at the time of the survey are dropped.

respondents who thought that completely following the referees' instructions was technically infeasible. Furthermore, because unwillingly adding experiments or data might not violate the honesty norm, Model 3 removes cases where addition of data or experiments was the only instruction. The results are qualitatively similar to Model 1. When authors positively appreciated the overall scientific value of referee comments, the likelihood of conformity is high (b=0.101, p<0.05). After controlling for other variables, three perceived causes of inconsistency become insignificant, though the signs of the coefficients are as expected.

Scientific field

Comparing the three scientific fields, scientists in biology research are more susceptible to referees' instructions than those in agriculture and medicine. The difference between biology and agriculture is statistically significant (b=–2.77, p<0.05). This result might be related to whether research is basic or applied, considering that



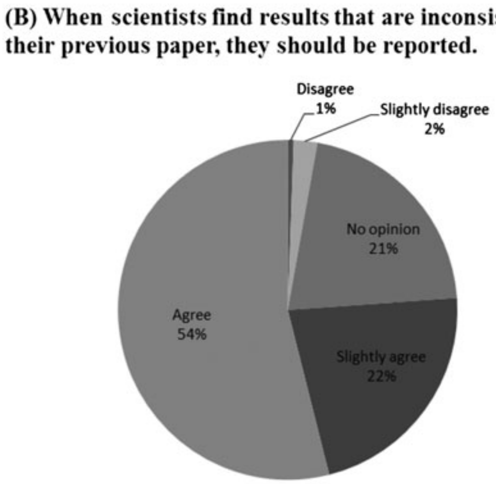


Figure 3. Opinions about honesty

agriculture and medicine are applied fields while biology is more basic. For example, priority might be more important in basic fields, which pressures biology scientists to secure their right to priority.[14] Alternatively, basic research, which tends to be exploratory, might allow various interpretations and give greater room for authors to conform with the recommendations of referees. Indeed, respondents in biology were more often instructed to change interpretations than those in other fields. Comparing the fields of degrees, we also find MDs are more likely to conform (b=2.14, p<0.05). Thus, while scientists engaging in medical research are not prone to conformity, those who received their education in medical schools are.[15] These results may suggest that propensity to conformity is influenced by early-career training.

Seniority

Associate professors are more prone to conformity than full professors (b=1.36, p<0.1). A few interpretations are plausible. First, associate professors are under

Table 2. Prediction of conformity

	Model 1		Model 2		Model 3		Model 4	
	Whole sample		Technically feasible		Instructions for other than data/exp addition		JIF available	
Technical infeasibility	–2.63***	(0.69)			–3.11*	(1.21)	–5.55**	(2.14)
Perceived value	1.01*	(0.40)	1.57**	(0.58)	2.17**	(0.78)	1.01	(0.80)
Author's fault	1.13	(0.70)	0.99	(0.90)	1.16	(0.95)	1.61	(1.41)
Referee's fault	–0.16	(0.69)	–0.38	(1.00)	0.11	(0.98)	–2.26	(1.50)
Competing of author & referee	–0.52	(0.90)	–2.17†	(1.26)	–3.85*	(1.60)	–0.93	(1.38)
Tier 1 (base group)								
Tier 2	0.67	(0.80)	–0.47	(1.17)	–0.60	(1.18)	2.31	(1.53)
Tier 3	0.27	(0.84)	–1.26	(1.37)	–1.79	(1.48)	–0.81	(1.45)
Biology (base group)								
Agriculture	–2.77*	(1.11)	–4.20*	(1.95)	–6.91*	(2.72)	–5.87*	(2.87)
Medicine	–1.51	(0.92)	–1.48	(1.14)	–1.66	(1.23)	–6.70**	(2.55)
MD	2.14*	(1.00)	0.96	(1.19)	1.08	(1.25)	4.96*	(1.97)
AP	1.36†	(0.77)	3.53**	(1.31)	4.02**	(1.55)	0.85	(1.89)
Age	0.06	(0.05)	0.14†	(0.08)	0.20*	(0.10)	0.17	(0.12)
Female	–1.84	(1.20)	–2.37	(2.13)	–1.31	(1.89)	–3.91	(2.54)
ln(#Times cited)	–0.62*	(0.27)	–0.45	(0.45)	–0.72	(0.46)	–1.01	(0.65)
#Competitors	0.67*	(0.30)	0.96*	(0.43)	1.50**	(0.57)	1.35*	(0.69)
Foreign experience	–0.68**	(0.25)	–0.98*	(0.42)	–1.33*	(0.53)	–1.19*	(0.59)
Low JIF							4.03*	(1.93)
Middle JIF (base group)								
High JIF							1.90	(1.69)
χ^2	73.26***		47.53***		67.89***		56.35***	
Pseudo R^2	0.459		0.480		0.578		0.598	
N	120		88		89		68	

Note: Logit regressions. Unstandardized coefficients and standard errors (parentheses). Two-tailed test. †p<0.10; *p<0.05; **p<0.01; ***p<0.001.

stronger pressure to publish and thus cannot afford to forgo the opportunity of publication by not following the instructions of referees. Second, associate professors may be less established and less confident in their scientific claims and thus more likely to yield to referees. Alternatively, full professors might be more skilled in preparing manuscripts dishonestly before submission by anticipating referees' comments. In consequence, they might not need to behave dishonestly after peer review.

Performance

Performance measured by citation count shows a significantly negative effect (b=–0.62, p<0.05), suggesting that high performers are less likely to follow referees than low performers. This can be understood in a similar way to the effect of seniority; high performers may not be desperate to publish many papers, or may be confident enough not to adhere to the referees' instructions. Though university tiers are significantly correlated with scientific performance, no significant difference is found among the three tiers.

Competition

The number of competitors has a positive effect (b=0.67, p<0.05), straightforwardly suggesting that competitive pressure causes dishonesty. This is consistent with the literature suggesting that intense competition is associated with misconduct and questionable practices (De Vries et al., 2006; Anderson et al., 2007c).

Foreign experience

Foreign experience shows a strongly significant effect (b=–0.68, p<0.01). Thus, returnees from foreign experience are less likely to conform. Examining this in detail, we find that propensity to conform is significantly higher for those who stayed abroad for less than one year (34% of our sample) than for those who stayed abroad for one year or longer (66%). A few interpretations are plausible. First, scientists who stayed abroad might have received better RCR training than is offered in Japan, though this interpretation may be unconvincing given that the training effect has been questioned (Anderson et al., 2007a). Second, returnees may be more confident in their scientific capability or skilled in debating with referees. Third, their expanded international network might expose them to stronger monitoring, discouraging dishonest behavior.[16]

Journal impact

Model 4 introduces the dummy variables of JIFs using a subsample whose JIF data are available. We find that peer review in low-JIF and high-JIF journals is more prone to conformity than in middle-JIF journals (Model 3: b=4.03, p<0.05; b=1.90, p>0.1, respectively). The positive coefficient of high-JIF, though insignificant, is straightforward because the temptation for publication should be great. This is consistent with Van Noorden (2011), who suggests that retraction tends to occur in high-impact journals. Our interviewees also suggested that increasing emphasis on high-impact journals in evaluations had invited deceitful publications in top journals. Interestingly, our result also suggests that conformity is significantly more likely to occur in low-JIF journals as well. In fact, Van Noorden (2011) finds that the recent increase in retractions is attributed in large part to low-impact journals. A few interpretations are plausible. For example, since low-JIF journals do not attract as many readers, authors may think that dishonesty is unlikely to be uncovered or that possible damage to the scientific community by being dishonest will be limited. Authors' primary motivation to publish in low-JIF journals may be to gain a publication metric rather than to inform the scientific community, so they do not really care about what their papers claim.

Discussion and conclusion

Although scientific publication and intellectual honesty therein are critical for the progress of science (Merton, 1973; Zuckerman, 1977), dishonesty has been commonly and increasingly observed in the forms of misconduct and other questionable practices (Martinson et al., 2005; Fanelli, 2009; Fang et al., 2012; Martin, 2013). Given that existing measures to prevent this conduct do not seem to be effective, we need a better understanding of why scientists deviate from the honesty norm. Among the various forms of dishonesty, this study focuses on dishonest conformity in peer review. Although dishonest conformity has long been known (Frey, 2003) and its negative impact on scientific advancement can be serious, empirical evidence has been scant. This study adds to the literature on research integrity and the peer review system by using survey data of Japanese life science professors to investigate dishonest conformity.

Our results suggest that conflict between authors and referees is quite common; approximately half of our respondents who received revision instructions felt that the instructions were inconsistent with their own scientific belief. While half the respondents appreciated the overall value of referee comments, the other half thought that referee instructions had either no or negative impact. This casts some doubt on the effectiveness of the current peer review system. Presented with instructions inconsistent with their own beliefs, the majority of the respondents followed the instructions. The payoff from conformity seems to be great in that most conformers successfully published in intended journals while half of non-conformers had to change journals. However, rebutting referees' arguments is not entirely futile; when authors resubmitted to the same journal without following inconsistent referee recommendations, only one in five papers was immediately rejected. This may imply that editorial monitoring is functioning reasonably well.

Further, regression analyses indicate some determinants of dishonesty. First, the propensity to conform differs by field of research and education; in particular, scientists engaging in biology research are more likely to conform than those in medicine and agriculture, as are those trained in medical schools. This result is different from Fanelli (2009), who finds that misconduct is more frequently committed by medical and pharmacological researchers than others. Perhaps the determinants of dishonesty depend on specific types of misconduct and questionable practices. In addition, the trait of dishonesty may be developed in early-career training.

Second, associate professors are more likely to conform than full professors. Martinson *et al.* (2005), comparing mid- and early-career scientists, find that the former are more prone to questionable practices. Taken together, propensity to dishonesty may be determined by the balance between pressure for publication, which is higher in early careers, and capabilities to circumvent honesty, which may be higher in later careers. Third, consistent with De Vries *et al.* (2006), scientists in competition are more likely to conform, suggesting that excessive pressure for publication can compromise research integrity. Fourth, scientists with foreign research experience are less likely to conform. Although this result needs further examination, international network or advanced RCR education may be possible deterrents. Finally, conformity tends to occur in low-impact journals compared with middle-impact journals. The existing literature suggests that emphasis on impact invites dishonesty in high-impact journals (Van Noorden, 2011), but this study suggests a different mechanism for dishonesty is at work in low-impact journals.

This study has several limitations, and future research is needed to further the understanding of honesty in science. The measure of conformity needs careful interpretation since it is self-reported, and desirability bias is possible. Though we assume that dishonest conformity occurs after referees provide comments, authors might dishonestly prepare the manuscript before submission, anticipating the likely instructions of referees (Frey, 2003). While this study focuses on dishonesty on the authors' side, future research needs to investigate the dishonesty of referees as well as editors. Our results invite several plausible interpretations. More qualitative approaches should be helpful to elucidate in-depth mechanisms behind dishonesty. Generalizability is another limitation. In the light of the peculiar incentive and career systems in Japan, other countries need to be investigated. Comparison among more diverse disciplines should also be helpful. This study draws on a sample of scientists in research-intensive universities, but those in education-oriented institutions may behave differently. This study focuses on dishonest conformity in peer review, but different types of questionable practices also need to be investigated in depth since background mechanisms could differ. For example, conformity in peer review is an interactive process between authors and referees (and editors), but dishonesty could occur differently in more unilateral processes or in processes involving other stakeholders.[17] The extent of perceived deviation from the honesty norm depends on each questionable practice, which might well change how the practice occurs.

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.org/10.1080/08109028.2015.1114745

Acknowledgements

This research was funded by the Japan Society for the Promotion of Science under the Grant-in-Aid for Encouragement of Young Scientists (B) Program (#24710160); the Ministry of Education, Culture, Sports, Science and Technology of Japan under Grant-in-Aid for Scientific Research (B) Program (#23330121); the Hoansha Foundation; and the Inamori Foundation. We thank the general editor of *Prometheus*, Stuart Macdonald, and two anonymous referees for critical and constructive review comments. We also acknowledge Seongwuk Moon and Scott Stern for reviewing the earlier version of this paper presented at the AIEA-NBER conference in 2014.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. http://www.singaporestatement.org/statement.html. The other three principles are accountability, professional courtesy and fairness, and good stewardship.
2. http://grants1.nih.gov/grants/research_integrity/whatis.htm. The other shared values are accuracy, efficiency, and objectivity.
3. http://www.nsf.gov/oig/resmisreg.pdf, http://grants.nih.gov/grants/research_integrity/research_misconduct.htm.
4. School Basic Survey (http://www.mext.go.jp/b_menu/toukei/chousa01/kihon/1267995.htm).
5. Fields are determined by WoS subject categories. A complete list of subject categories is available in supplementary data (Table S1 available online at http://dx.doi.org/10.1080/08109028.2015.1114745).
6. As of 2014, among 86 national universities, 53 have departments related to life sciences (e.g. medicine, science, agriculture).
7. We set this condition to spare the effort of preparing questionnaires in foreign languages. Comparing foreign-borns and Japanese scientists is of interest, but because the proportion of foreign-born scientists in Japan is extremely small (Franzoni et al., 2012), meaningful comparison would have been difficult.
8. We first grouped the authors into three scientific fields (medicine, agriculture, and biology) based on WoS subject categories. Then we employed stratified random sampling in

each field. For the JIFs, since we anticipated that authors of high-JIF papers might be relatively unwilling to respond because of being busy, etc., we oversampled high-JIF authors. As we were also interested in the behavior of low-JIF authors, they, too, were oversampled.Sampling weights of 5.0, 1.0, and 2.5 are given to top, middle, and bottom JIFs, respectively. For university ranks, sampling weights of 1.0, 1.5, and 2.0 are given to Tiers 1, 2, and 3, respectively. We oversampled lower-ranked universities because their population is smaller.

9. Japanese academia is highly gender-biased. Geuna and Shibayama (2015) find that the proportion of female professors in STEM fields is between 2% and 8%.
10. If they had multiple instances, we asked them to choose the specific case which revealed the most serious inconsistency.
11. Unwillingly adding experiments/data may not violate the honesty norm, although we still believe that adding unnecessary experiments/data is considered questionable in that it is a waste of resources.
12. We used the *Journal Citation Reports* of 2012, published by Thomson Reuters, owners of Web of Science.
13. The interpretation requires caution because of small sample size and selection bias.
14. When two papers report similar results, the first published paper is given greater reputational reward than the second published paper (Merton, 1973). We suppose that the difference of reward given to the two papers is larger in basic fields than in applied fields.
15. The field of medicine includes both clinical and non-clinical research. We tested whether propensity to conformity differs between them, but did not find a significant difference.
16. However, when we examined the effect of the size of network (e.g. the number of coauthors, etc.), we did not find a significant effect.
17. For example, interaction among multiple authors is relevant to authorship abuse; interaction between the principal investigator and students may be relevant to data fabrication.

ORCID

Sotaro Shibayama http://orcid.org/0000-0002-6701-9828

References

Anderson, M., Horn, A., Risbey, K., Ronning, E., De Vries, R. and Martinson, B. (2007a) ‘What do mentoring and training in the responsible conduct of research have to do with scientists’ misbehavior? Findings from a national survey of NIH-funded scientists’, *Academic Medicine*, 82, pp.853–60.

Anderson, M., Martinson, B. and De Vries, R. (2007b) ‘Normative dissonance in science: results from a national survey of US scientists’, *Journal of Empirical Research on Human Research Ethics*, 2, pp.3–14.

Anderson, M., Ronning, E., De Vries, R. and Martinson, B. (2007c) ‘The perverse effects of competition on scientists’ work and relationships’, *Science and Engineering Ethics*, 13, pp.437–61.

Azoulay, P., Furman, J., Krieger, J. and Murray, F. (2012) *The Effects of Scientific Paper Retractions*, NBER Working Paper, National Bureau of Economic Research, Washington DC.

Barber, B. (1952) *Science and the Social Order*, Collier, New York.

Bosch, X., Hernandez, C., Pericas, J., Doti, P. and Marusic, A. (2012) ‘Misconduct policies in high-impact biomedical journals’, *PLoS One*, 7, 12, e51928.

Broad, W. (1981) ‘The publishing game – getting more for less’, *Science*, 211, pp.1137–39.

Coupe, T. (2004) ‘What do we know about ourselves? On the economics of economics’, *Kyklos*, 57, pp.197–215.

Csada, R., James, P. and Espie, R. (1996) ‘The “file drawer problem” of non-significant results: does it apply to biological research?’, *Oikos*, 76, pp.591–93.

David, P. (1998) ‘Communication norms and the collective cognitive performance of “invisible colleges”’ in Navaretti, G., Dasgupta, P., Mäler, K.-G. and Siniscalco, D. (eds) *Crea-*

tion and Transfer of Knowledge: Institutions and Incentives, Springer, Heidelberg, pp.115–63.
De Vries, R., Anderson, M. and Martinson, B. (2006) 'Normal misbehavior: scientists talk about the ethics of research', *Journal of Empirical Research on Human Research Ethics*, 1, pp.43–50.
Ellison, G. (2002a) 'The slowdown of the economics publishing process', *Journal of Political Economy*, 110, pp.947–93.
Ellison, G. (2002b) 'Evolving standards for academic publishing: a q–r theory', *Journal of Political Economy*, 110, pp.994–1034.
Fanelli, D. (2009) 'How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data', *PLoS One*, 4, 5, e5738.
Fanelli, D. (2010) 'Do pressures to publish increase scientists' bias? An empirical support from US states data', *PLoS One*, 5, 4, e10271.
Fang, F., Steen, R. and Casadevall, A. (2012) 'Misconduct accounts for the majority of retracted scientific publications', *Proceedings of the National Academy of Sciences of the United States of America*, 109, pp.17028–33.
Franzoni, C., Scellato, G. and Stephan, P. (2012) 'Foreign-born scientists: mobility patterns for 16 countries', *Nature Biotechnology*, 30, pp.1250–53.
Frey, B. (2003) 'Publishing as prostitution? – choosing between one's own ideas and academic success', *Public Choice*, 116, pp.205–23.
Geuna, A. and Shibayama, S. (2015) 'Moving out of academic research: why do scientists stop doing research?' in Geuna, A. (ed.) *Global Mobility of Research Scientists*, Elsevier, Amsterdam, pp.271–303.
Grieneisen, M. and Zhang, M. (2012) 'A comprehensive survey of retracted articles from the scholarly literature', *PLoS One*, 7, 10, e44118.
Hagstrom, W. (1974) 'Competition in science', *American Sociological Review*, 39, pp.1–18.
Ietto-Gillies, G. (2012) 'The evaluation of research papers in the XXI century. The open peer discussion system of the world economics association', *Frontiers in Computational Neuroscience*, 6, article 54.
Ishibashi, E. and Ohtake, Y. (2009) 'Report of the survey of scientific, technological and academic activities in the universities', *NISTEP Report*, 167.
John, L., Loewenstein, G. and Prelec, D. (2012) 'Measuring the prevalence of questionable research practices with incentives for truth telling', *Psychological Science*, 23, pp.524–32.
Kneller, R. (2007) 'Prospective and retrospective evaluation systems in context: insights from Japan' in Whitley, R. and Glaser, J. (eds) *Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Springer, Dordrecht, pp.51–73.
Lawson, C. and Shibayama, S. (2014) 'International research visits and careers: an analysis of bioscience academics in Japan', *Science and Public Policy*, 42, 5, pp.690–710.
Macrina, F. (2007) 'Scientific societies and promotion of the responsible conduct of research: codes, policies, and education', *Academic Medicine*, 82, pp.865–69.
Martin, B. (2013) 'Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment', *Research Policy*, 42, pp.1005–14.
Martinson, B., Anderson, M., Crain, A. and De Vries, R. (2006) 'Scientists' perceptions of organizational justice and self-reported misbehaviors', *Journal of Empirical Research on Human Research Ethics*, 1, pp.51–66.
Martinson, B., Anderson, S. and De Vries, R. (2005) 'Scientists behaving badly', *Nature*, 435, pp.737–38.
Marusic, A., Bosnjak, L. and Jeroncic, A. (2011) 'A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines', *PLoS One*, 6, 9, e23477.
Matsuzawa, T. (2013) 'Research misconduct in Japan: macro-analysis based on open information', *Journal of Information Processing and Management*, 56, pp.156–65.
McAllister, P., Anderson, R. and Narin, F. (1980) 'Comparison of peer and citation assessment of the influence of scientific journals', *Journal of the American Society for Information Science*, 31, pp.147–52.
Merton, R. (1973) *Sociology of Science*, University of Chicago Press, Chicago.
MEXT (2006) *Guideline for Measures against Research Misconduct*, Tokyo.

NAS (1992) *Responsible Science, Volume 1: Ensuring the Integrity of the Research Process*, National Academies Press, Washington DC.
Pitsoulis, A. and Schnellenbach, J. (2012) 'On property rights and incentives in academic publishing', *Research Policy*, 41, pp.1440–47.
Prufer, J. and Zetland, D. (2010) 'An auction market for journal articles', *Public Choice*, 145, pp.379–403.
Resnik, D. and Master, Z. (2013) 'Policies and initiatives aimed at addressing research misconduct in high-income countries', *PLoS Medicine*, 10, 3, e1001406.
SCJ (2013) *Prevention and Enforcement Measures against Research Misconduct*, Science Council of Japan, Tokyo.
Shamoo, A. and Resnik, D. (2003) *Responsible Conduct of Research*, Oxford University Press, Oxford.
Shibayama, S. and Baba, Y. (2015) 'Impact-oriented science policies and scientific publication practices: the case of life sciences in Japan', *Research Policy*, 44, pp.936–50.
Shibayama, S., Baba, Y. and Walsh, J. (2015) 'Organizational design of university laboratories: task allocation and lab performance in Japanese bioscience laboratories', *Research Policy*, 44, pp.610–22.
Tsang, E. and Frey, B. (2007) 'The as-is journal review process: let authors own their ideas', *Academy of Management Learning & Education*, 6, pp.128–36.
Van Noorden, R. (2011) 'The trouble with retractions', *Nature*, 478, pp.26–8.
Wilhite, A. and Fong, E. (2012) 'Coercive citation in academic publishing', *Science*, 335, pp.542–43.
Zuckerman, H. (1977) 'Deviant behavior and social control in science' in Sagarin, E. (ed.) *Deviance and Social Change*, Sage, Beverly Hills, pp.87–138.