Teams prevent misconduct: A study of retracted articles from the Web of Science

Rathmann, Justus M K ; Rauhut, Heiko

Originally published at:
Rathmann, Justus M K; Rauhut, Heiko (2019). Teams prevent misconduct: A study of retracted articles from the Web of Science. In: 17th International conference on Scientometrics Infometrics ISSI 2019, Rome, Italy, 2 September 2019 - 5 September 2019. Editioni Efesto, 1032-1037.

Teams Prevent Misconduct: A Study of Retracted Articles from the Web of Science

Justus M. K. Rathmann[1] and Heiko Rauhut[2]

[1] *rathmann@soziologie.uzh.ch*
Institute of Sociology, University of Zurich, Andreasstr. 15, CH-8050 Zurich, Switzerland

[2] *rauhut@soziolgie.uzh.ch*
Institute of Sociology, University of Zurich, Andreasstr. 15, CH-8050 Zurich, Switzerland

Abstract
Collaborations become increasingly important among almost all scientific disciplines. Teams can be more productive and achieve more attention to their work. It is, however, unclear, whether teams also lead to higher research integrity. On the one hand, there may be a volunteer's dilemma in larger groups such that the responsibility diffuses who is controlling whom. The 'volunteer hypothesis' predicts that the more co-authors, the more scientific misconduct. On the other hand, larger groups may also achieve a higher level of social control. The 'control hypothesis' predicts that the more co-authors, the less scientific misconduct. Retractions are used as an operationalization of scientific misconduct. The data collection comprises of retracted articles from the Web of Science data set. In addition, control groups of non-retracted articles are constructed, using methods known from causal inference and bibliometrics. The analyses demonstrate that larger author groups have a lower retraction probability compared to smaller author groups. This suggests that teams prevent misconduct, most likely by their higher ability to social control. The results indicate that the development towards more and larger research collaborations may have positive macro-level consequences for the system of science.

Introduction
Collaboration is an emerging topic in science studies (Hall et al., 2018). The number and share of research in teams is steadily increasing. At the same time, teams become larger. These findings are robust across the scientific disciplines. There are good reasons for such developments; collaborating scientists can share their work and concentrate on their strengths, additionally, collaborating with others can generate new ideas. In addition, scientific articles with a larger number of co-authors are cited more often than articles with fewer co-authors (Wuchty, Jones & Uzzi, 2007). However, it remains an open question, whether teams also lead to higher research integrity.

The Office of Research Integrity (ORI) identifies research misconduct as consisting of data fabrication, data falsification and plagiarism and underlines that honest errors or differences of opinions are not misconduct (The Office of Research Integrity, 2011). One way of uncovering scientific misconduct and inherently flawed research are so-called retractions (Hesselmann et al., 2016). This is the case because only a small fraction of retractions is due to honest mistakes; the vast majority of retractions are due to various forms of scientific misconduct (Fang, Steen & Casadevall, 2012; Gewinner, Diekmann & Rathmann, u.r.).

Hesselmann et al. (2016) summarize the state of the art on retractions; compared to other publication types, retractions are very uncommon, representing only 0.02% of the PubMed database. However, the share of retractions has been increasing steadily since the end of the last century, this growth has accelerated since the early 2000s. In absolute terms, most retractions come from the United States, whereas the emerging science nations generate the largest share of retractions on total scientific output. Retractions are most prevalent in medicine, chemistry, life sciences and multidisciplinary studies. Articles in high-impact journals and highly cited articles are retracted comparatively more frequently.

Studies on the relationship between the number of co-authors and retractions have yielded mixed results so far. In addition, these investigations have mostly referred to studies in bio-medicine, medicine and life sciences. Foo (2010) found that only 6.6\% of all retractions in biomedicine and life sciences fall on individual authors. In these disciplines, however, single authorship is uncommon. Furman, Jensen & Murray (2012) have found no statistical evidence of the number of co-authors affecting the likelihood of retraction. However, other research concluded that fraudulent retractions had significantly more co-authors than erroneous retractions (Steen, 2010) and that randomized clinical trials with fewer authors had a higher likelihood of retraction (Steen & Hamer, 2014). This leads to the following research question: Does the size of the research team influence the probability of an article being retracted?

This research question will be addressed using publication data from the Web of Science (WoS). Using a matching procedure, the so-called bibliographic coupling (Kessler, 1963), retracted and ordinary research articles will be sampled into a control and a treatment group, resulting in a quasi-experimental design. The treatment is the retraction process. As a consequence, differences in the number of authors in the two groups can be attributed to the behaviour underlying the treatment.

Theoretical Considerations

Two contradictory hypotheses can be derived from theories of actor models and group processes; 'social control' on the one hand and the 'diffusion of responsibility' on the other hand. In group contexts, people inform themselves about context-specific group norms by observing identity-consistent behaviour of core group members in order to adapt and integrate (Hogg et al., 2004). Good scientific practice is an important norm in science (The Office of Research Integrity, 2011). The more co-authors, the higher the probability that one of them has a sceptical attitude and demands receiving in-depth information on the production process of critical matters. Auspurg & Hinz (2011) show, for instance, that articles by single authors are more prone to publication bias than articles by multiple co-authors. This leads to the first hypothesis:

H_a: The larger the number of co-authors of an article, the lower is the likelihood of retraction.

The concept of 'diffusion of responsibility' predicts that the probability of volunteering to do something negatively correlates with group size (Darley & Latane, 1968). Based on this, Diekmann (1985) has developed the so-called 'volunteer's dilemma'. In this game one volunteer is sufficient to provide the public good, which is, however, costly to produce. Both the free-riders and the volunteer profit equally from the public good. The game has the interesting macro-level implication that volunteering becomes less likely the larger the group is. Translated into the scientific context, this means that only one volunteer is needed to review the data and analysis. Hence, the more authors, the higher the probability that all involved think that the others closely checked data collection or processed results, while in fact, nobody may have done it. This leads to the second hypothesis:

H_b: The larger the number of co-authors of an article, the higher is the likelihood of retraction.

Methodological Considerations

In a first step, all retracted articles from the WoS are sampled. There is no 'retraction'-category in WoS, therefore the retractions were identified by querying for articles with the document type 'correction' or 'correction, addition' and the word 'retract' in the title. These serve in the later analysis as a treatment group. Subsequently, four control groups are formed on different levels of similarity.

The control groups and their matching mechanism are presented in Table 1. The first control group (CG Field) is based on a random discipline sampling. The retractions are matched with a random article from the same WoS subject category. It is important to note that an article in the WoS can have several subject categories, which are assigned based on the subject categories of the journal. The second control group (CG Journal) is based on random journal sampling. Articles from the same journal and year are matched with the retracted articles.

The third and fourth control groups are determined on the basis of the bibliographic couplings. In this matching mechanism, the similarity of articles is determined on the basis of shared references; the more shared references between two articles, the more similar the articles are (Kessler, 1963). Articles by an author with the same name as the author of a retracted article are rejected as potential matches in order to avoid to match the author with himself. In case of a tie, one article is randomly selected from the best matches. The third control group (CG Coupling) is, thus, based on regular bibliographic coupling. Since the number of references of an article positively correlates with the number of co-authors (Glänzel & Schubert, 2004), an improved bibliographic coupling algorithm is introduced. The number of shared references, is in this algorithm divided by the number of co-authors to control for this correlation. This matching procedure is used for the fourth control group (CG Imp. Coupling).

Table 1. Control groups and measures of similarity

Control Group	*Matching Partner*
CG Field	Random article from the same field and year
CG Journal	Random article from the same journal and year
CG Coupling	Article with the highest number of shared references
CG Imp. Coupling	Article with the highest number of shared references divided by the number of co-authors

Figure 1 shows a schematic plot of the bibliographic coupling mechanism. The retracted article by Miller would be matched with the article by Johnson. The articles have two references in common which is more than the shared number of references by the articles of Perry and Miller. The other article by Miller is excluded.
This methodological structure results in a quasi-experimental design with pairwise-matching and no prior measurement.

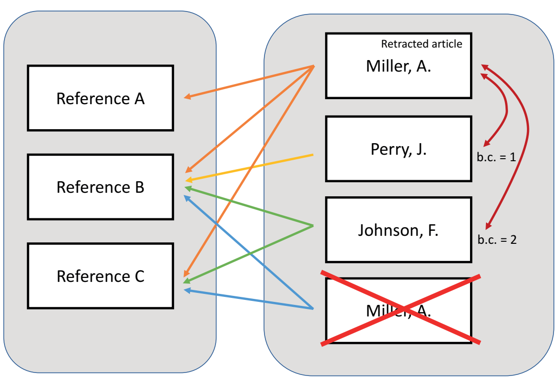

Figure 1. Schematic plot of bibliographic coupling

Preliminary Results

The data set consists of 3,549 retractions which make up the treatment group. CG Field and CG Journal each consist of 3,549 matched articles. The matching algorithms for the bibliographic coupling are not finished to date, therefore there is no data for CG Coupling and CG Imp. Coupling yet.

For the following analyses, the number of authors of an article is field-normalized. The field-normalization is similar to the *z*-transformation. The field-normalized number of authors A_{norm} is obtained by subtracting the average number of authors in the field μ_{field} from the number of authors A and dividing the result by the standard deviation of the number of authors in the field σ_{field} (see equation below). The field is determined by the subject category in the WoS database.

$$A_{\text{norm}} = \frac{A - \mu_{\text{field}}}{\sigma_{\text{field}}}$$

T-tests are used to compare the field-normalized average number of authors in the treatment group with the number of authors in CG field and CG journal. The mean value of the treatment group is -0.33. This means that retracted articles have 0.33 standard deviations fewer authors than it is usual in the respective field. The mean value of CG field is -0.02, the t-test yields a *t*-value of 14.8. The mean value of CG journal is 0.09, the t-test yields a *t*-value of 15.2. The difference for both control groups is therefore significant at $p < 0.01$. Retracted articles, thus, have significantly fewer authors than randomly matched articles from the same field and the same journal.

Additionally, logistic regression models with the variable *retraction* as the dependent variable are conducted. This variable is dichotomous, 1 for retracted articles and 0 for regular articles. The detailed regression tables are presented in Table 2 in the appendix. The number of authors is modelled by a second order polynomial function to model a non-linear relationship. The predicted retraction probabilities are depicted in Figure 2 and figure 3. Since each sample consists of 50% retracted and 50% regular articles the baseline retraction probability is 50%.

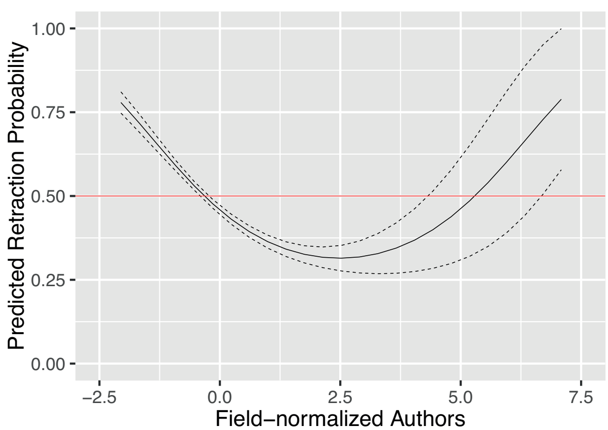

Figure 2. Predicted retraction probability given field-normalized authors (CG Field)

Figure 2 plots the predicted regression retraction probabilities in the random field sample. The predicted retraction probability decreases steadily as the number of authors increases until reaching 2.5 standard deviations. Therefore, the function increases steadily resulting in a u-shaped function. Thus, articles from author groups smaller than usual in the field are more likely to be retracted. The same effect can be found for very large author groups with more than 5 standard deviation from the field mean.

Figure 3 plots the predicted regression retraction probabilities in the random journal sample. As in the field sample, the predicted retraction probability continuously decreases as the number of authors increases up to 2.5 standard deviations. Unlike in the field sample, the predicted retraction probability in the journal sample decreases before reaching a plateau at 5 standard deviations. This sample also shows that articles of groups whose size is below average have a higher probability to become retracted. This pattern holds for a wide range of author group sizes.

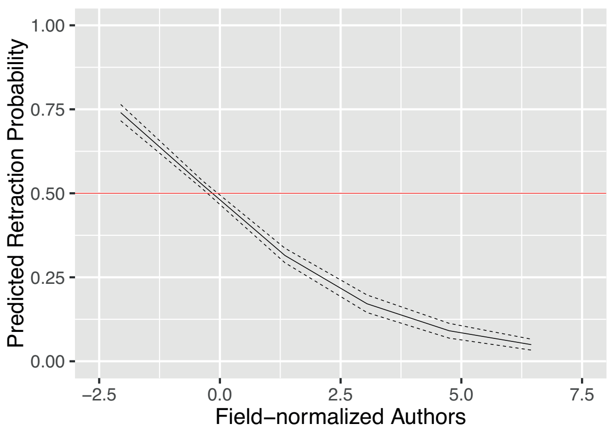

Figure 3. Predicted retraction probability given field-normalized authors (CG Journal)

Preliminary Conclusion

Retracted articles are more likely to originate from author groups which are smaller than usual in the respective field. This result can be shown in both t-tests and logistic regressions in different samples. These preliminary results therefore discard the idea of a diffusion of responsibility developed in hypothesis H_b and indicate a case for social control as stated in hypothesis H_a. Albeit, the diffusion of responsibility might play a role for very large author groups. However, one can also argue for a selection effect; if a researcher knows he wants to perpetrate scientific misconduct, he might choose to work alone or in a small team in order to have as few accessories as possible. These findings are backed up by (preliminary) findings in other areas of misconduct; Auspurg & Hinz (2011) find publication bias more likely to be present in articles by single authors, Horbach & Halffman (2019) present evidence that articles containing problematic text recycling have less authors than articles not containing problematic recycling. Although the results so far are robust, it is not yet possible to draw a final conclusion. The bibliographic coupling allows a significant step forward in the determination of similarity and could lead to new results.

In all matching processes, there is a chance to match a retracted article with an article that will be retracted in future. It can take as long as 35 months for an article to be retracted (Decullier, Huot & Maisonneuve, 2014). However, this chance is very small due to the low prevalence of retractions. Furthermore, relatively few authors are responsible for a large part of the retractions (Gewinner, Diekmann & Rathmann, u.r.) and this bibliographic coupling algorithm excludes matching an author with itself.

Acknowledgements

The Authors thank Antonia Velicu for research assistance. Justus Rathmann is funded by the Swiss National Science Foundation (SNSF) under the SNSF Starting Grant "CONCISE" (S-64408-01-01).

References

Auspurg, K & Hinz, T. (2001) What fuels publication bias? *Jahrbücher für Nationalökonomie und Statistik*, 231(5-6).

Darley, J.M. & Latane, B (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology* 8(4), 377-383.

Diekmann, A. (1985). Volunteer's dilemma. *Journal of Conflict Resolution*, 29(4), 605-610.

Decullier, E., Huot, L. & Maisonneuve, H. (2014). What time-lag for a retraction search of PubMed? *BMC Resarch Notes*, 25(7), 395.

Fang, F.C., Steen, R.G. & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*, 109(42), 17028-17033.

Foo, J.Y.A. (2010). A retrospective analysis of the trend of retracted publications in the field of biomedical and life sciences. *Science and Engineering Ethics*, 17(3), 459-468.

Furman, J.L., Jensen, K. & Murray, F. (2012). Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine. *Research Policy*, 41(2), 276-290.

Gewinner, J., Diekmann, A., & Rathmann, J.M.K. (under review). Why are Articles Retracted?

A Systematic Analysis of the Reasons Behind Retraction from the 'Retraction Watch' Blog. *Science, Technology, & Human Values*.

Glänzel, W. & Schubert, A. (2004). Analysing scientific networks through co-authorship. In H.F. Moed, W. Glänzel & U. Schmoch (Eds.), Ha*ndbook of Quantitative Science and Technology Research* (pp. 257-276). Kluwer Academic Publishers.

Hall, K.L., Vogel, A.L., Huang, G.C., Serrano, K.J., Rice, E.L., Tsakraklides, S.P. & Fiore, S.M. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist*, 73(4), 532-548.

Hesselmann, F., Graf, V., Schmidt, M. & Reinhart, M. (2016). The visibility of scientific misconduct: A review of the literature on retracted journal articles. *Current Sociology*, 65(6). 814-845.

Hogg, M.A., Abrams, D., Otten, S. & Hinkle, S. (2004). The social identity perspective. *Small Group Research*, 35(3), 246-276.

Horbach, S.P.J.M. & Halffman, W. (2019). The extent and cause of academic text recycling or self-plagiarism. *Research Policy*, 48(2), 492-502.

Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10-25.

Steen, R.G. & Hamer, R. (2014). A case-control comparison of retracted and non-retracted clinical trials: Can retraction be predicted? *Publications*, 2(1), 27-37.

Steen, R.G. (2010). Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics*, 37(4), 249-253.

The Office of Research Integrity. (2011). *Definition of research misconduct*. Retrieved 25 January 2019 from: https://ori.hhs.gov/definition-misconduct.

Wuchty, S., Jones, F.B. & Uzzi, B. (2007). The increasing dominance of teams in the production of knowledge. *Science*, 316(5827), 1036-1039.

Appendix

Table 2. Retractions compared to the average number of authors in the field and journal

	Field sample	Field sample	Journal sample	Journal sample
Authors	-0.414***	-0.492***	-0.511***	-0.527***
	(0.029)	*(0.031)*	*(0.030)*	*(0.030)*
Authors^2		0.099***		0.013***
		(0.014)		*(0.001)*
Intercept	-0.077***	-0.168***	-0.078***	-0.091***
	(0.024)	*(0.028)*	*(0.025)*	*(0.025)*
N	7,098	7,098	7,098	7,098
AIC	9,623	9,578	9,496	9,486

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; S.E. in parenthesis