A PROPOSITION TO GIVE VALUE TO THE WORK OF REFEREES

This text has been written by Vincent Lafforgue, with the help and suggestions of Jean-Benoît Bost, Thierry Bouche, Jean-Pierre Demailly, Alain Genestier, Gérard Laumon, Assaf Naor, Ricardo Perez-Marco, Boris Pioline, Claude Sabbah, Flaminio Squazzoni, and Vanessa Vitse (but they have no responsability in this text).

We start from the following two observations :

a) the work of the referees is fundamental and deserves to be valued,

b) the difficulty that the editorial boards of the journals may have in finding referees who check the articles with the greatest care is a serious problem for the sustainability of a flawless scientific research.

Moreover b) has the following two causes (among others) :

(i) the time spent by researchers in many tasks, for example administrative tasks, and especially the time wasted in the search for funding,

(ii) the excessive pressure to publish, compared to the lack of reputational rewards for referees, so that too many articles are written and they are not enough checked and improved in the refereeing process.

We suggest to create a system enabling researchers to prove, if they wish, the referee work they did, without compromizing their anonymity as referees. We hope it would help for a) and (ii).

Our motivation is only scientific : to have fewer articles, but with a better refereeing process. The problem of the cost of commercial journals is very serious but it is not discussed here (however it is not completely orthogonal since institutions would, by the system we propose, encourage refereeing in journals with good practices, and they could include a low price in these good practices).

We require very demanding specifications for such a system. We then show by an example that it is possible to construct a system satisfying these specifications.

This text is just a suggestion to be studied, I do not claim it is a good idea. In particular it should not open the door to automated evaluation of research which is a very bad idea. In fact what is proposed is rather a crypto-currency system, where journals give crypto-coins to referees (both for accepted and rejected articles, and taking into account the quality of the reports). The total number of crypto-coins a journal can give is equal to the number of pages published each year, and the value of these crypto-coins would be set by research institutions, depending on the scientific level, virtue and good practices of the journals. For example the institution could make a list of the serious journals for which it supports refereeing and classify them in three categories of different qualities. For

Date: 7 mai 2018.

each researcher, the institution would add these crypto-coins given by the journals of the three categories, and these three numbers would indicate how many pages have been published thanks to his/her referee works in each of the three categories of journals (these numbers would be less than the numbers of pages of the articles on which he/she wrote reports, because of rejected articles and multiple reports). Then researchers could use these numbers for reputation purpose in their careers. According to some social studies [SBT13], it is not a good idea to give material incentives to referees because this would have a demoralizing effect on disinterested and altruist referees, and decrease the quality of reports. We hope that the system we propose would not have such a negative effect, but this is a serious question.

Cryptography is used to ensure the anonymity of referees is not compromized. But I am not expert in cryptography. If the system was considered useful, it should be evaluated and improved by experts in cryptography.

1. Specifications for such a system

The system must enable each researcher to prove the referee work he/she carried out, while respecting the following conditions :

(C1) the system must absolutely preserve the anonymity of the referees,

(C2) the system must be resistant to cheating attempts by any of the actors,

(C3) almost no extra work should be required from referees and editorial boards, and only limited and automatable work from journals and institutions,

(C4) the functioning of journals should not be disturbed.

The system we propose does not require any central authority. It involves, in addition to researchers and journals, institutions. In fact the role of an institution is simply to create a platform for computing total scores. The total score can be a number, or better a set of numbers indicating the number of pages published thanks to the reports of the researcher in three categories of journals (see the end of the second section for a discussion). Everybody could create such a platform, but in practice we think that the platforms chosen by the researchers would be created by big research institutions, national science fundations, or national scientific societies, which inspire confidence (for example in Maths, CNRS, Mathdoc or SMF in France, NSF or AMS in the USA). The qualities an institution must possess to inspire confidence are the following :

(I1) it reliably attaches a total score to the proofs of referee work provided by a researcher,

(I2) it does not disclose the names of the journals that provided such proofs of referee work (and thus preserves the anonymity of referees).

The main point is that the total score will no longer contain any information which could compromize the anonymity of the referees.

There are two possible options : the total score will be

— either published by the institution,

— or certified by the institution in encrypted form, but not publicly disclosed.

In both options the researchers will be able to use their total scores as they wish, for example for their applications, promotions, requests for funding or activity reports. I don't know which option is the best one. A total score like the number of pages published thanks to the referee works of a researcher in journals of different categories has a tangible value and there is no need to compare it with the total scores of other researchers. A reasonable possibility is that each researcher could choose

— either to publish his/her total score
— or to keep it secret (but certified by the institution in encrypted form) and still use it for his/her promotions,...
— or never to ask for it.

The total score is not only a proof of an essential work done for the community and the progress of science, but also a proof of competence since the referees are chosen by the editorial boards for their expertise. It would also be a way, for researchers who have a slowdown in their publications, to show that they are still active and in touch with current research. Therefore even a narrow-minded employer (which would care only about a country or a university, and not about the progress of science) would have interest to take it into account.

Conditions (C1) et (C2) are detailled by the following more technical conditions :

(C1a) the researchers will only provide the institution with the names of the journals for which they have written reports and the scores that the journals have assigned to them for these reports, but not the titles of the articles on which they wrote reports,

(C1b) using these scores the institution will calculate a total score by taking into account the quality of the journals so that the total score no longer contains the names of the journals for which the researcher has written reports,

(C1c) it will be impossible ever to find the names of the referees from the encrypted data published by the journals to certify the scores they have attributed, whatever the future progress of computers (even quantum ones) and cryptanalysis are,

(C2a) the journals will not be able to artificially inflate the scores in order to attract referees to the detriment of other journals,

(C2b) a researcher will not be able (with the current strength of computers in cryptanalysis) to produce a false proof of referee work.

2. Concrete description of a simple system

We want to show that it is easy to build such a system. For this we will describe precisely a simple system respecting the specifications. To describe this system we explain successively the role of editorial boards, of journals, of researchers, and of institutions.

The technical details, which are standard in cryptography and easy to implement, can be skipped.

2.1. **Role of the editorial board.** The editorial board assigns a score to each report by its prefered method. We may suggest to the editorial board

— to assign the same scores to positive or negative reports, resulting in publication or rejection, in order not to create incentives in one direction,
— to fix once and for all an automated algorithm assigning a score to each report according to the number of pages of the reviewed article and a box checked by an editor, to distinguish short, detailed or very detailed reports, and add a bonus depending on the quality of the report, the difficulty of the article, and especially the errors found and improvements suggested by the referee (which should be rewarded by a high bonus).

The additional workload for the editorial board is therefore low, once the automated algorithm is put in place.

2.2. **Role of the journal.** For each report, the journal creates an alphanumeric text YJSN by concatenating

— the year Y (determined by the journal, for example, the year in which the decision is made to publish or reject the article),
— the journal J, more precisely J will be the name or acronym of the journal (as the journal prefers) followed by its ISSN code (to avoid any ambiguity),
— the score S,
— the name of the referee N, in the form first name-(initial)-last name.

For example, if Athanase B. Clifford writes in 2018 a report for the journal Annales de l'Institut Fourier with ISSN code 1777-5310 and obtains the score 40, YJSN will be

2018+AIF-1777-5310+Score40+Athanase B. Clifford

We put here + signs to separate the parts Y, J, S, N whose length can vary, and it is assumed that none contains the + sign (but it is a technical detail to be adjusted later).

Then

— the journal concatenates to YJSN a random text r, for example YJSNr will be

2018+AIF-1777-5310+Score40+Athanase B. Clifford+9d39f68f8afc23f80ee8884e89ffd44b55c1997f822ce1bd72333fd834b5cfee

— the journal computes with a hash function the hash h=Hash(Nr) and it deduces the concatenated text YJSh, in our example YJSh will be

2018+AIF-1777-5310+Score40+ea7a269899b049920f71d8f372433399fd4ae805e56a3f2500736ab29097255d

— every year the journal publishes the list of all YJSh (for all positive or negative reports), and also publishes the sum of the scores S and the coefficient

$$\beta_{J,Y} = \frac{\text{number of pages published in the year}}{\text{sum of the scores}},$$

— the journal sends by email to the referee the alphanumeric text YJSNr.

The journal can sign the list of all YJSh to certify it.

A hash function is a function which is easy to compute but for which it is extremely difficult (and impossible with current computers) to find an inverse

image, or two strings with the same image. So the idea (standard in cryptography) is that h certifies N if we know r.

The above operations are very simple. For example under Linux or by opening a Terminal with a Mac, we can obtain a pseudo-random hexa chain of 256 bits r by the command openssl rand 32 -hex, and h can be computed with SHA-256 by the command echo -n Nr | shasum -a 256 (the above examples were obtained in this way). We can choose other random generators and hash functions of course. To limit the work of the journal, all these operations should be automated.

This system respects the specifications :

— if YJSh has been published to certify the referee work of a researcher N, a researcher N' (different from N) cannot, with current computers, find r' such that the hash of N'r' is h, which guarantees (C2b), nor can the journal find Nr and N'r' having the same hash (which would have allowed it to use the same YJSh for several researchers and would have contradicted (C2a)),

— we take r at least as long as the hash h so that, when r varies, the hash h of Nr takes almost all the possible values of h, and therefore, even if in the future supercomputers are able to compute all the inverse images for the hash function, they will never be able to deduce N from h, so this guarantees (C1c),

— the interest of the coefficient $\beta_{J,Y}$ is that $\beta_{J,Y}S$ is a normalized score such that the sum of the normalized scores is equal to the number of pages published in the year by the journal : everybody can verify the computation of $\beta_{J,Y}$ (as the quotient of the number of pages published by the journal J in the year Y by the sum of the scores S for all YJSh in the list published by the journal), and this prevents the journal from inflating artificially the scores and therefore guarantees (C2a).

It would be more pleasant for referees that the journal normalizes itself the scores at the end of the year, when it sends the emails to the referees, so that $\beta_{J,Y} = 1$. Indeed the normalized score has a clear meaning : it indicates the number of pages published by the journal using the report. Of course this number will be less than the total number of pages of the article because of rejected articles and multiple reports on the same article.

2.3. **Role of the researcher.** If the researcher does not wish to prove a referee work he/she just destroys the email by which the journal communicated YJSNr to him/her, and the hash h in YJSh published by the journal will remain undeciphered forever.

If the researcher wishes to prove his/her referee works to an institution (so that it gives him/her a total score that he/she can use for his/her applications, promotions ...), he shall communicate to it, for each report, the alphanumeric text YJSNr which the journal has sent to him/her by email. For example, it could copy and paste YJSNr from this email to a box on a web page of the institution.

Then the researcher could use this web page to obtain his/her total score corresponding to some interval of years and, in the option where these total scores are not publicly disclosed, a way to certify it.

So this system requires almost no additional work for researchers.

2.4. **Role of the institution.** The institution creates a file containing all YJSh for all journals (participating to the system) from the lists they have published (since the file is the same for all institutions, it only needs to be created once). With 100000 mathematical articles per year, and 100 octets for each YJSh, this makes a maximum of 10Mo each year. The institution checks that each journal has correctly normalized the scores so that $\beta_{J,Y} = 1$, or otherwise it computes $\beta_{J,Y}$ (as the quotient of the number of pages published by the journal J during year Y by the sum of the scores S for all the YJSh of the file).

When a researcher communicates YJSNr to it, the institution can then compute h = Hash(Nr) since the hash functions are public (for example, by the command echo -n Nr | shasum -a 256). It checks that YJSh appears in the file, and that N is the name of the researcher.

All journals are not of the same quality, but the names of the journals for which a researcher made reports must remain secret. To value the referee works I see two solutions.

First solution. The institution gives to the referee work the value $\alpha_{J,Y,I}\beta_{J,Y}S$, where $\alpha_{J,Y,I}$ is a quality coefficient that the institution I assigns to each journal J (and that it can re-evaluate each year Y). By taking the sum of the $\alpha_{J,Y,I}\beta_{J,Y}S$ for a given researcher the institution obtains a total score for all the referee works that he/she wishes to be valued. The total score does no more contain the names of the journals for which the researcher has written reports, and thus preserves the anonymity of the referees. The system therefore satisfies conditions (C1a) and (C1b). To limit the work of the institution, the computation of the total score as the sum of the $\alpha_{J,Y,I}\beta_{J,Y}S$ should be automated. The institution chooses the coefficients $\alpha_{J,Y,I}$ as it wishes. Reasonably the coefficient $\alpha_{J,Y,I}$ should depend on the scientific level of the journal, an estimate of its rejection rate (the higher the rejection rate is, the more the journal needs referees for the same number of published pages), and also on its virtue and good practices.

Second solution. The institution regroups serious journals in three (or more) categories of journals of comparable quality (with the same criteria as above). Then it computes the sum of the $\beta_{J,Y}S$ for each category (if the categories are big enough this will not compromize the anonymity of the referees). In that case the total score is a set of numbers, namely for each category, the number $\sum_{J\in\text{category}} \beta_{J,Y}S$ which indicates the number of pages published in the year Y by this category of journals using the researcher's reports. This number will be less than the total number of pages of the articles (published by this category of journals) on which he/she has written reports because of rejected articles and multiple reports on the same article. Note that, in order to protect the anonymity of referees, these numbers would have to be rounded to integers, and even to multiple of 5 if they are big.

The second solution seems better because
— it does not look like an automated notation system that we want to avoid

— the number of pages published in each category of journals thanks to the referee works of a researcher has a clear meaning.

3. Conclusion

The system proposed above thus satisfies all the specifications. It goes without saying that, if the idea is applied, the system should be examined by experts in cryptography and the technical details will have to be revised depending on the practical constraints of its implementation.

Typically the system could be started in the following way. First, an institution and a core of journals adopt it. After consultation of their computer experts, they standardize

— the format of the alphanumeric texts YJSNr and YJSh (in order to enable the researcher to copy and paste YJSNr from the email sent by the journal to the institution's web page),
— the hash function,
— the format of the files published by the journals containing the list of the YJSh.

This is the only standardization necessary to start the system. Otherwise the system is completely decentralized. All the journals and institutions can freely join it. Given the difficulty in finding referees, journals will have interest to adopt the system (which requires little effort from them). One institution in the world is enough to compute total scores, but if the system spreads there will probably be many.

In fact free journals (which everybody wants to encourage) could find it difficult to do the necessary programming to give the scores to referee reports, apply the hash function, and send the emails to the referees. Therefore it would be better if a free software to do this was available (in other words it would be better if the first journals adopting this system create a free software which other journals can use if they wish).

To avoid any misunderstanding, it is important to clarify the following three points.

This text is in no way intended to promote automated evaluation systems. It is only to preserve the anonymity of the referees, according to the draconian conditions (C1a), (C1b) and (C1c), that we proposed an automated system. The other activities of the researchers, which are not anonymous, are much better evaluated in a non-automated way.

This text is not a plea for the anonymity of referees. A referee may sign his/her report if he/she wishes. The proposed system only avoids anonymity being compromized against the referee's will.

Finally, the proposed system would not disturb the current functioning of journals, nor would it hinder the progress that many people wish (notably towards free open access). On the contrary, by the choice of the coefficients $\alpha_{J,Y,I}$ or of the categories to which the journals belong, institutions would encourage journals to raise their quality and to adopt good practices.

4. Another use of the proposed system

The following idea was suggested by Jean-Pierre Demailly. Often the reports received by the editorial boards contain comments of great interest for the public, and are more detailed and relevant than the reports published on MathSciNet after the publication of the article. One could imagine that journals encourage voluntary referees to extract from their report a public report, which the journal would post on the web together with the published article. This public report could be anonymous or not. It could then be used by databases such as MathSciNet or zbMATH. This is independent of the system proposed in this text, except that the journals could reward such an extra work of voluntary referees by a higher score.

5. Complements

5.1. **Possible drawbacks of this system.** As we already said, according to some social studies [SBT13], it is not a good idea to give material incentives to referees because this would have a demoralizing effect on disinterested and altruist referees, and decrease the quality of reports. We could fear this also for the system we propose. In particular we could fear that researchers would be demotivated to referee for low level journals, because it would bring them only a small reward, or no reward at all. However some authors may submit good articles to low level journals for specific reasons (for example to please an editor, or because of the country, the region or the university where the journal is published). Another unintended consequence could be that journals which institutions ranked in a low category start competing to attract reviewers with material incentives.

Another problem comes from the fact that the institution has to give to each journal a quality coefficient $\alpha_{J,Y,I}$, or at least classify serious journals in three categories, where in each category the normalization is given by the number of pages published each year. The choice of these quality coefficients may be considered as a delicate question by the institution. On the other hand, the normalization by the number of pages has the drawback that pages may be larger or smaller according to the journal. Moreover online journals could propose their articles in a form such that they have no clear number of pages. The number of characters published each year by the journal would be a better normalization than the number of pages.

Another question is whether scores should be made public or kept secret (and used by researchers only for their applications, promotions or requests for fundings). What we suggest in this text is that each reseacher would choose for himself/herself. An argument in favor of public scores is that they would have greater emulating effect. But their list could be misused by people or institutions.

5.2. **Problems of change of name and homonymy.** The name of a researcher is not a perfect identifier because

a) persons may change their name (especially after a marriage),

b) there are perfect homonyms, with the same name, (the same initial) and the same first name.

In case a), the institution must be informed by the researcher of his/her change of name.

The rest of this subsection is about case b). The problem is that perfect homonyms could cheat by pooling their proofs of referee work but this seems extremely unlikely : it would be necessary that two perfect homonyms are both dishonest and come into contact to pool their proofs of referee work, even though the fraud would remain detectable (in case of doubt about homonyms an institution can always send an email to a journal to check an identity).

If we still want to distinguish perfect homonyms, we can add in N an additional identifier, for example

— the birth date (provided by the referee to the journal),
— for mathematicians, the MR Author ID number (that the journal can find on MathSciNet).

Given the diversity of practices between countries and disciplines we propose that N could be any alphanumeric string containing the name of the referee and one or more identifiers, separated by ; signs. For example N could be

Athanase B. Clifford or

Athanase B. Clifford ;BIRTHDATE 15 June 1987 or

Athanase B. Clifford ;MR Author ID 123456.

In conclusion we suggest to standardize the system to be able to add other identifiers in N (e.g. ORCID numbers if they become universal), but to add none for the moment.

5.3. **A possibility to check the correctness of the computation of the total scores.** The system described above was devised with the idea that the institution is absolutely reliable, so that nobody would question the correctness of the computation of the total scores. The cryptography used in this text is very elementary. Experts in cryptography could probably create a system where everybody can check the computation of the total scores by the institutions.

However I am not sure that such a modification would be a good idea because

— anyway, researchers need reliable institutions in which they can trust,
— the institutions which rule the platforms will play an important role in evaluating the scientific level of journals and encouraging good practices, and only reliable institutions will do that in a honest way.

Références

[SBT13] F. Squazzoni, G. Bravo, K. Takács. Does incentive provision increase the quality of peer review ? An experimental study. *Research Policy 42* (1) 287–294 (2013)