



# Ethics for a new scientific millennium

SVEN KOCHMANN

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:  
[sven.kochmann@gmx.de](mailto:sven.kochmann@gmx.de)

DATE RECEIVED:  
September 24, 2015

© Kochmann This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



## INTRODUCTION

For quite a long time, people discussed about terms and concepts like *Impact factor*, *'negative results'*, and *'reproducibility'* in all kinds of analog and digital media. Here, I want to make a little contribution to this discussion on my own, mainly because I feel that some important points (at least important for me) often get lost in the shuffle.

The title may be familiar to you. It is as some of the following concepts and ideas based on a book I read a while ago. The German version, which I read, is *Dalai Lama (2013). Das Buch der Menschlichkeit: Eine neue Ethik für unsere Zeit*, Bastei Entertainment, ISBN: 9783838749174 (eng. *Bstan-'dzin-rgya-mtsho XIV. (2001). Ethics for the New Millennium*, Riverhead Books, ISBN: 9781573228831).

How does a (buddhistic) book about ethics fit in here? Well, let us start this discussion with something, which has been discussed so many times that probably everyone is tired of it. Me included. However, I think it is important for explaining the later concepts.

## IMPACT FACTOR

Hence, let us talk about **impact**, first. What is actually impact? According to the [Oxford Dictionary](#) it is:

1. The action of one object coming forcibly into contact with another.
2. The effect or influence of one person, thing, or action, on another.

Do these definitions really tell us what impact is? I find them kind of vague. 'Influence' or 'effect' are exactly like 'impact': everyone has an imagination of the concept behind the word, but there is no math-like, exact definition. You cannot really **grab** it, can you?

If I throw a stone (a pack of journal-volumes) into a peaceful lake (a crowd of scientists), it will have an impact for sure. However, measuring all the effects will be impossible. First, the stone will displace the water. This will cause waves. Animals will try to dodge the falling stone in the water. It will eventually crush into the bottom of the lake and whirl up some sand or other stones. Maybe destroy something. Also, do not forget about secondary impacts such as the waves reaching the border of the lake and having their own impact there. Of course, this impact was caused by me throwing the stone.

Now, we try to measure this impact – despite the fact we do not even really understand what it is actually – by number of citations. For the metaphor from above, this would be like counting the waves the stone has caused and dividing it by the number of all waves (during a certain time period). On top, we do not even count it ourselves, we assign this task to a company with financial interests.

So, what about the impact of the stone under the surface? What about long-term impacts? What about

the stone itself? With no word, I described the stone. Not the size nor color nor material. It could be toxic for the lake or just stay there for hundreds of years and be a home for organisms. But wouldn't it be important to know all these things to get a more complete picture of the stone and the nature of its impact? To know what actually IS the stone and not (only) what it causes?

Analogously, wouldn't it be necessary to know the content of scientific work to assess it? Sure. However, the impact factor will not tell us that at all.

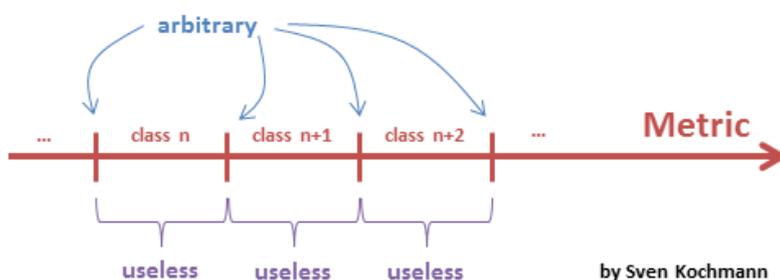
### SORITES PARADOX

Every publishing metric system so far including the impact factor is subject to the principle of the **sorites paradox**. If you have a heap of sand, it does not matter if you take a sand corn away or add one. However, we know that two, three, ... sand corns are not a heap and therefore it would matter. So, there must be an arbitrary boundary where a couple of sand corns become a heap, right? For metrics of scientific work this is – roughly – the arbitrary boundary between *very important* and *not so important*. Hypothetically spoken, if a metric system is subject to the principle of the sorites paradox, its nature is arbitrary and its value useless.

Does it matter that Cell has an actual impact factor of 32 (2014) instead of 30? Does it matter if one of my paper has 4000 Downloads instead of 3800? But it does matter if 100 people read this post instead of just 2! So, where is the boundary? 50 people? 25 people? For any metric, can you ever tell where exactly this boundary is? It seems arbitrary.

Let us for a moment assume we have a metric scale with only one boundary and we knew exactly where this boundary is (our scientific cat told us!). Then it is obvious that this boundary divides our metric scale into two areas, a low one (i.e. *not so important*) and a high one (*very important*). Within these areas (other people would say 'classes') the metric value becomes completely useless. As for high impact journals it does not matter if a low impact journal has an factor of 1.0 or 1.5 or 2.0... It does not matter if two persons read my article or three did.

There seems to be a straightforward solution to this problem. You can simply create more classes or areas by introducing more boundaries. For instance, you could try to distinguish between few readers (around 5), many readers (around 100), and a lot of readers (around 1000). However, the problem remains the same: the boundaries between classes are arbitrary if set and the metric values itself become useless within a certain class. The following picture visualizes this thought.



In turn and in my honest opinion, this renders these metrics useless and arbitrary. Yet, people like and create this kind of metrics for a simple reason. In contrast to the definition of 'impact' from before, a simple number is graspable. People can work with simple numbers. Counting, comparing, sharing, simple math. The higher the number, the more of 'it' – whatever 'it' is. People feel better with 'more' instead of 'less'. People like having more IQ points (than others). People like having more impact points (than others). People like having more working hours (than others). The last one seems wrong at first glance, but since the impact of work generally is quantified by counting the number of working hours, more hours imply that a person is more valuable (for society). Analogously, the quality of peer-review is often **quantified by counting the number of days it takes**.

There is a reason why this kind of **'simple'** metric does behave as described. These metric systems count things, which may or may not be linked to the state of a work (e.g. important, of high quality, ...), instead of measuring a quantity, which is linked (directly or indirectly) to the state. The impact factor tries to describe the quality of scientific work in journals by counting citations. It **should be** obvious that the impact factor fails in doing this. If a journal just publishes (not obvious) frauds and people just reference to these works, because they investigate every single paper to expose these frauds then the journal possesses high impact but no quality. Hence, the state of quality is not given by counting citations. It might be an indication (in both directions), though.

Let me try to make this even clearer by taking a converse example from physics, viz. temperature scale. Any temperature scale (metric) is linked to the physical concept of temperature. Temperature gives information about the state of matter. You can draw exact boundaries (e.g. liquid water between 0°C and 100°C, ice below 0°C, etc.) and it matters anywhere on the scale (in thermodynamic systems!) if something has x degree or x+1 degree. Tungsten melts at 3422°C not at 3421°C. Oxygen freezes at -219°C not at -218°C. This is exactly the opposite behavior of the metric systems described above.

Another important characteristic of the latter metric system is the existence of outer boundaries, i.e. it seems to have natural limits on both ends. For temperature this might not be obvious. There is an absolute zero temperature on one end. Since temperature describes a thermodynamic equilibrium the absence of this equilibrium renders the specification of temperature completely useless although it is done for describing plasma states for example. In this case, 'temperature' becomes a countable metric – it just counts the thermal kinetic energy of particles. In consequence, there is an upper boundary for the temperature scale (i.e. when there is no thermodynamic equilibrium anymore).

Conclusively, if we want to assess scientific work and its quality with a metric system, we cannot just count things. We have to find or create a system, which is linked to the qualitative state of the work, which allows us to set exact boundaries, and which possesses outer boundaries (i.e. is restricted on both ends). The question is just: Does such a system exist?

#### **CRITERIA FOR ASSESSING SCIENTIFIC WORK**

Before we even think about a new metric system, we should find or define criteria by which we want to assess scientific work. No matter which we (or others) chose, in the end, the aim is always the same: we only seek a tool – no – THE tool, the one ring, the holy grail, the... to assess **them** all! To assess all scientific work!

Now the time has come to reference to these books I mentioned in the introduction. As said I read the German version and will therefore cite this here (did not find the original text). The following quote is about a principle to assess a moral action. One of the principles, I try to put into practice everyday.

Daraus können wir ableiten, daß ein Kriterium zur Beurteilung einer moralischen Handlung darin besteht, wie ihre Auswirkung auf die Erfahrungen oder Glückserwartungen anderer ist. Eine Handlung, die diese verletzt oder ihnen Gewalt antut, ist potentiell unmoralisch. Ich sage »potentiell«, weil die Folgen unserer Handlungen zwar wichtig sind, es aber noch andere Aspekte zu bedenken gilt, etwa die Frage nach der Absicht sowie die nach dem Wesen der Handlung selbst. Uns allen fallen Dinge ein, die wir getan und mit denen wir andere verletzt haben, obwohl das keineswegs in unserer Absicht lag. (Seite 36ff)

This is the English version (by Google translator and some modifications from my side – if someone has the original text, I would be more than happy to cite it!):

From this we can deduce that a criterion for assessing a moral action is its impact on the experiences or expectations of happiness of other people. An act, which hurts them, is potentially immoral. I say *potentially* because **the consequences of our actions are important, but there are other aspects that are important as well such as the intention and the nature of the action itself.** We all

remember things we did that hurt others, even though that was not our intention. (Page 36 et seqq.)

I think, we can adapt this concept for assessing scientific work. Let us just replace *moral actioni* by *'experiment'* or *'scientific work'*, *'consequences'* by *'results'*, and *'nature of actioni'* by *'performance'*. Then the marked section becomes **'the results of our experiments are important, but there are other aspects that are important as well such as the intention and the performance of the experiment itself'**. Thereby, we can in general assess any experiment and in consequence any series of experiments (i.e. publication) by individually evaluating its three parts, viz. intention, performance, and results.

The **intention** covers the initial idea, the hypothesis, and/or the overall goal. It usually contains an essay about the state of research and how the work fits in there (introduction). **Performance** accurately describes how the experiments were designed and performed. What tools, instruments, and materials/chemicals were used. What was the raw data and how was the raw data processed and evaluated. **Results** covers interpretation, discussion, conclusion, etc.

Think of classical treasure hunting for a simple metaphor: **Intention** is the idea to get rich by digging up a pirate treasure in the Caribbean and outline of the plan to do so. **Performance** describes the tools (ship, shovel, crew members, parrots, ...) you used including a picture of the map and how you got from the base harbor to the treasure island. Also, how you found the way from the beach to the X, what you lost on the way, and what the treasure was like. **Results** will discuss what you gained (gold, experience, illness), and if it was worth the effort. In the end, you can refer to your initial overall goal (getting rich) and judge because of your experiment (or maybe you tried several times?) if treasure hunting is a feasible way to reach this goal.

At the current scientific state, every publication describing treasure hunting will lead to the same conclusion, i.e. that it is in principle a feasible way to get rich. Only if the authors successfully found a treasure of gold, their research will be published, though. Also, there are some problems and issues people have to solve before 'in principle' becomes 'in fact' (reducing the number of dying crew members, buried treasures are a limited resource, parrots eat all the crackers, ...).

This means we need to get away from looking at the **Results**, only. Immediately. So, let us look at the other parts, too, in order to actually quantify the quality of the overall work!

#### ASSESSING SCIENTIFIC WORK

For now, all three parts described in the previous section shall be equivalently weighted, i.e. 1/3. Thus, this reduces the importance of results by 2/3 (from almost 100%) while the importance of the other two parts strongly increases (from almost 0%). Then, let us analyze each section and discuss how we could assess it.

We start with a tough one. The content of **intention** is hard to judge – almost impossible – and trying to do so (like reviewers do for example) causes the “we cure cancer”-phenomenon. Authors feel the urge that their contribution to science **has to** be something bigger, something that scratches at the doors to Stockholm. Also, by every journal policy it has to be a novel piece. In consequence, authors justify all of their intentions with “a new way” to “eventually cure cancer”. Maybe one of my readers thinks now “is this really so important?”.

As the title of this blog posts suggests, the content is about ethics. Is the ethical and moral OK to (indirectly) force people to hide their true intentions? I think not. Allow people to be honest! Accept that people were “just” interested in this because no one else did it! Accept that people have cool ideas by doing nerdy stuff because for the sake of (“Why?” “Because we can!”)! Accept that people tried to reproduce things from another scientific work, but were unable to do, and want to publish their efforts including discussion and conclusion! **Let them tell their story in a simple language!**

Second section to assess is the **performance** section, which reminds you probably of the “Materials

and methods” section of most papers. However, in which journal did you see that this is 33% of the work? Usually, it consists of many short subsections roughly describing the experimental methods by omitting all kinds of important details (humditiy, anyone?). If it is too long you can look up the remaining parts in the supplementary files, a wild jungle of text and pictures no one cared about to write, format, or review. Let me pull up a quote something from the Hitchhiker’s guide to the galaxy:

The plans were on display. [...] even if you had popped in on the off chance that some raving bureaucrat wanted to knock your house down, the plans weren’t immediately obvious to the eye [...] I eventually had to go down to the cellar! [...] With a torch! [...] It was on display in the bottom of a locked filing cabinet, stuck in a disused lavatory with a sign on the door saying “Beware of the Leopard”.

So, the experimental details may be all there. Some in the manuscript, some in the supplementaries. You just have to find them! And be able to open it, because they might be submitted in a unknown or rare data format.

However, the performance of an experiment is the central point of any scientific work. We cannot talk about the lack of reproducibility, if we force authors to reduce the fraction of this central point to 5%, maybe 10%, of the whole work. Some journals even print this section in a smaller font visually reducing further the importance.

Here, we again force authors to do something, they may not want, i.e. shorten and part their description of the performance. An then we accuse their work of not being reproducible. This reminds me of Nelson telling people “stop hitting yourself” while constantly hitting them in the face. It should be clear that if you shorten something to a certain degree, you have to omit things. If the description of an experiment requires 10 pages, you cannot reduce this to one paragraph and on the same time keep the same grade of detail. You cannot tell people to write the “most important things” in the manuscript and put the rest in the supplementaries. I do not even see the point.

In our digital era, there is absolutely no need to shorten something for the sake of shortening it or because it looks “nicer” and fits the journals policies. However, journals and publisher dictate what and how things will be scientifically published. Academic ethics have to defer to journal policies. For the sake of ethics and science, we have to change publishing.

These policies cause another issue. Publishers practically implemented something I will call “reproducibility by obscurity” (referring to “security by obscurity”). As described above, finding the details of methods and experiments is sometime a real hassle. Thus, we withdraw responsibilities from the authors and publisher and push it to the reader. If you cannot reproduce something from a paper, you never know if you just couldn’t get all of the details together (so the problem is on your side) or something is missing in or really wrong with the description. When in doubt, you just try to find another work. Thus, nobody (or hardly anyone) contest the reproducibility of the work, i.e. it has some kind of existence – even it may be only virtual (Schrödinger’s cat may be dead by now!). Just because of obscurity. There is absolutely no ethics in this developement. It obfuscates the state of reproducibility. This is a problem. A big problem. (Personally, I noticed that the higher the impact factor the less details you find in the papers. This is probably just my own impression, though.)

Last but not least, let us talk about the most controversial section on the list, i.e. **results**. So far, results have been the only quantify to assess scientific work. Well, to be more precise, arbitrary concepts of *positive* and *negative results* have been. However, how can results be positive or negative at all? I’ll make an action, there is an equal reaction. I’ll do an experiment, there is a result. Always. Case closed.

Unless you apply your expectations on the results. Then you derive a new concept of “results” I will call here “expectults” but everyone else will keep calling “results”. Obviously, expectults change with expectations. If my expectation on any publication is that its results will cause world peace and solve starvation, well, then practically every publication published negative expectults. So, they should be

retracted. All of them!

While writing the last paragraph I realized how stupid the word “expectults” is. I think it fits very well, because this is what we do. We literally expect that every publication cures cancer. On top, we only publish things, which fit in the big picture of curing cancer. Things, which are politically (i.e. by policy) not acceptable, will not considered for publication. This fits the definition of **copyright**.

The odd thing is that editors and reviewers are enforcing this. Both groups are scientists. Working in the same system. Maybe dooming and cursing this system. They could change it. Still, they are doing the same process over and over again. In the end, scientist are just censoring themselves.

Do you know what is even worse than negative expectults? Let us assume you have a collaboration with another group and they should do something for you (measure, synthesize, etc). After a while, they tell you there were “no results”. “No results?” you ask. “Yeah, no results. Did not work.” they say. This is a direct result of above policy and I am sooo tired of this. I spend (i.e. waste) so many hours of tearing answers out and getting the actual results from them, so we can track down issues and work on them. Why not just say “This is the spectra. It does not show the expected product lines, but here you can see the lines for educt 1, 2, 3, and an unknown line. What do you think?”. Does this only happen to me?

How do you fight censorship? By transparency, i.e. open science. However, you have not only to publish reviews of manuscript (versions), which were accepted, but on all those which were **not accepted**. Transparency does not work in one direction, only. Of course, this renders a distinction between *accepted* and *not accepted* publications completely useless. So, simply let the people publish whatever results they have. Let everyone review and assess the work. In public. This is the only way.

For assessing the results we want to exclude any (personal) expectations out of it. Otherwise, it will only cause discussions on virtual experiments to perform, which may or may not show some point the reviewer wants to make – most likely to prove somehow that the authors did not do the experiments properly or at all. Well, but if you want to do more experiments, then do it. Write it up and publish it yourself!

Also, is it even ethical to always assume the worst? Why not just assume the authors did everything right (pro reo) and write supportive reviews instead of assuming they were all douchebags (contra reo) and try to destroy their work?

Ok, we roughly discussed now what we want to assess. How can we meter it? This is what the next section is about.

### **METERING THE QUALITY OF SCIENTIFIC WORK**

If you read the previous section, then you can already presume the conclusion of this section. There is no specific metric, which can do what we want. Each work is individual and, therefore, should be treated individually. Read individually. Reviewed individually. Discussed individually. Assessed individually, by words and language not by some arbitrary metric system.

Even formal things are impossible to meter. Scoring or measuring the quality of explanation and presentation, for example. Text styles are as many in number as there are people. Some might come up with ideas such as “giving the text to 10 random-selected, undergraduate students – every student who understood it is a point”. This is random, arbitrary, and elaborate.

Many things could be considered binary maybe trinary, such as logic of an explanation or reproducibility of an experiment. The explanation is or is not logic (or partly). The experiment can fully, can partly, or cannot at all be reproduced. Integers could be assigned to these states (e.g. 0, 1, 2). Then, you could try to count these things and divide them by the total amount, e.g. count reproducible experiments. I explained before why counting things to quantify quality is a bad idea. For experiments, where does one experiment start and another one ends? Is setting a solution to a certain pH one or two

experiments (adding acid/base as one, measuring the pH as the other)? Two simple experiments may be easier to reproduce than an experiment involving the Large Hadron Collider.

Let us assume that we could somehow meter the quality of the individual parts, viz. intention, performance, and results. How will the individual parts be weighted? For the last part of my post, I set them to 33% each, mainly to reduce the importance of and focus on the results section. Difficult. Also, all section will most likely have different metric systems unless we implement a simple 5-stars-option (this works good for hotels all over the world, right?). How to combine different metric systems to a final one?

Another option would be badges! Achievements! Everyone likes achievements, right? Let us give paper the 'Golden Reproducibility Badge' for works being replicated by at least three other research groups! Let us give people the 'Mo-Mo-Monster Review'-Achievement for writing ten good (respectful and helpful) reviews! At least **they are very popular among the open communities**. Why not in science?

As I wrote before, in the end we only seek THE one tool to assess all scientific work! Maybe someone in future finds a nice system. I do not know. No one can know this. However, I know that we become so focused on this quest that we already forgot about the thing, which matters most: the scientific content behind the number(s).

### ETHICS FOR A NEW SCIENTIFIC MILLENNIUM

In **another article I already described** the imbalance of so-called *peer*-reviewers and authors of a submitted manuscript. Anonymous *peer*-reviewers have been granted power. Enforcing their anonymous point of view on authors. Demanding more and more experiments and citations. **Being a sexist**. Being douchebags. However, being a douchebag or sexist is independent of anonymity. For instance, see several Twitter 'discussions'. I think, we need some guidelines for assessing and discussing in a scientific context.

Of course, there already a bunch of guidelines and rules for good scientific practice, which are enforced by research foundations like **DFG**. However, all of these guidelines and rules are only for creating scientific work (i.e. scientific misconduct, which is important!), but there are no real guidelines for assessing scientific work. So, I wrote up some ethical guidelines, which might help you and others. I do not and will not want to call them rules, since I am not seeking to enforce them on anyone (see # 1). They are in no specific order.

1. I do not enforce my opinions and views on others, no matter what. All my comments, remarks, and assessments are meant to be respectfully written recommendations and advices.
2. I will not assess any scientific work by its results only nor will I ever use or apply the arbitrary concepts of *negative* and *positive* expectults, or *no* results.
3. I am aware that I might be wrong on anything. If I am wrong, I will admit this. If I am not sure about something, I will admit this. If I have no clue, I will admit this.
4. I will not support or tolerate *reproducibility by obscurity*. Performances of experiments should be as accurate as possible. Accuracy is mandatory.
5. I will not demand more results, experiments, citations (esp. on my own work), or other things for the sake of. Additional things should offer added value. Also, they should not be mandatory.
6. I will not use any metric for judging or proving the quality of scientific work.
7. I will bear full responsibility for my assessment, comments, and remarks as authors of scientific work bear full responsibility for their work.
8. I will not tolerate, obscure, or turn a blind eye on any form of discrimination, harassment, or the like, at any point.
9. I do not directly or indirectly force people to act in an improper or unethical way.

Some of these points remind you probably of rules and guidelines in discussion forums. But what is assessing work other than a discussion? At least it should be one. Maybe you find the list redundant. Maybe you find some things be über-natural. Well, please go ahead and search for the lists of all these

'respectful' comments from reviewer 3.

This list and its points are not set in stone. They are mine. I try to follow them and maybe extend them in future. Everyone can make his/her own list. My point is just: We need ethical guides for assessments. They will not turn douchebags to upright people – I know this. However, they are a tool, which allows me to clarify **my position** and distinguish myself from douchebags. I simply want to do science.

**FINAL REMARK**

I hope, you liked my little contribution. Leave a comment with your own point of view and share it (and mine) with others!