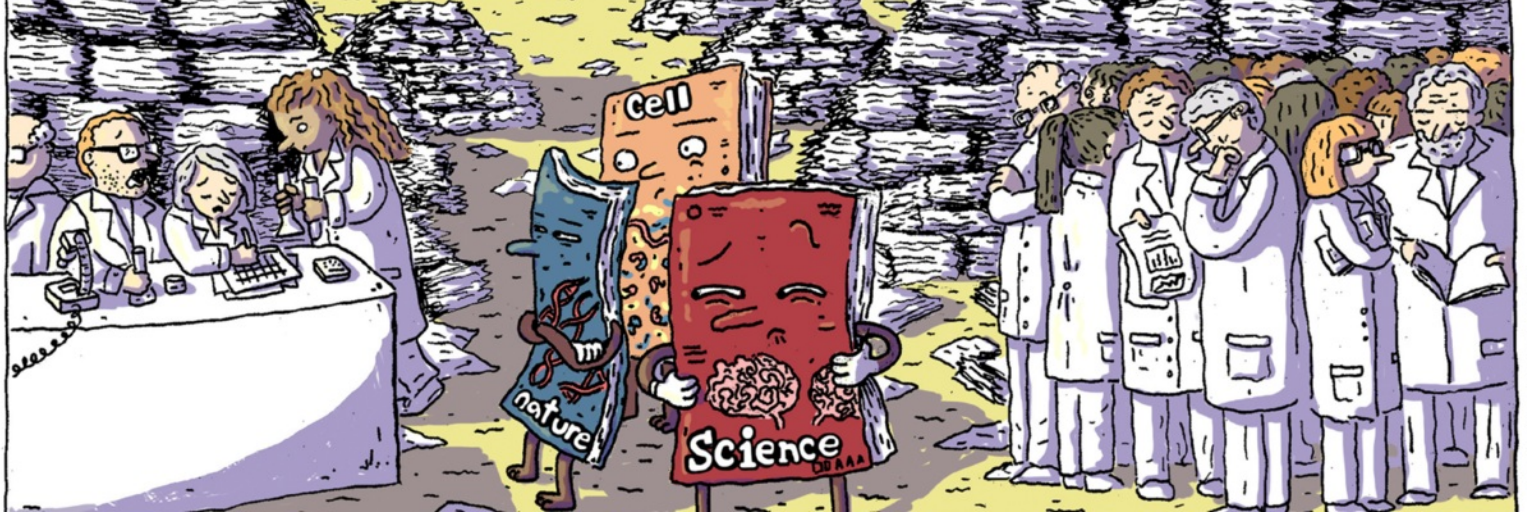


Reproducibility Crisis in the Sciences: The "Agitation"

Authors: Julieta Alcain
Submitted: 26. September 2019
Published: 30. September 2019
Volume: 6
Issue: 9
Affiliation: EGLC Journal, Argentina
Languages: English
Keywords: medicine, biology, research, sciences, experiments, reproducibility, Dr. Jekyll and Mr. Hyde
Categories: Life Sciences, Medicine
DOI: 10.17160/josha.6.9.603

Abstract:

The following article, previously published in the magazine El Gato y La Caja, presents and retracts a problem in the world of today's scientific laboratories. What about experiments? Why don't scientists get the same result twice? The strange case of Dr. Jekyll and Mr. Hyde was born in 1886 and quickly became a classic to the point that, as with most classics, you don't need to have read it to know the story: a renowned scientist discovers a potion that turns him into a perverse and violent subject, dissociating his personality and changing his physical appearance. Between both personalities, there is a tragic battle that knew how to have many reversions in literature and cinema. In the last pages, Dr. Jekyll, with new ingredients, wants to repeat his successful formula. But he does not succeed, because the effect of the potion was due to an impurity in the old inputs, already consumed. This impurity is impossible to trace, which naturally prevents Dr. Jekyll from getting the same result again. The illustrations of the cover by Chica Estelar.

new JOSHA
Journal of Science, Humanities and Arts

September 2019 Volume 6, Issue 9

THE AGITATION

JULIETA ALCAIN

El GATO Y LA CAJA

"I sent out for a fresh supply and mixed the draught; the ebullition followed, and the first change of colour, not the second; I drank it and it was without efficiency. You will learn from Poole how I have had London ransacked; it was in vain; and I am now persuaded that my first supply was impure, and that it was that unknown impurity which lent efficacy to the draught."

The strange case of Dr. Jekyll & Mr. Hyde - Robert Louis Stevenson

The strange case of Dr. Jekyll and Mr. Hyde was born in 1886 and quickly became a classic to the point that, as with most classics, you don't need to have read it to know the story: a renowned scientist discovers a potion that turns him into a perverse and violent subject, dissociating his personality and changing his physical appearance. Between both personalities, there is a tragic battle that inspired literature and cinema. In the last pages, Dr. Jekyll wants to repeat his successful formula with new ingredients. But he does not succeed, because the effect of the potion was due to an impurity in the old inputs, already consumed. This impurity is impossible to trace, which naturally prevents Dr. Jekyll from getting the same result again. Who would have thought that in this fascinating metaphor of the psychiatric disorder of multiple personalities, in this philosophical approach to human duality, its author Robert Louis Stevenson was also going to introduce a fundamental concept of science: reproducibility in scientific experiments.

Another brick in the wall

Knowledge (understood as a collective and collaborative good) is built like a wall: with bricks. Each new discovery is a brick, on which others are placed. So that the wall does not collapse, it is important that these discoveries are solid. Those who read these data will take them as reliable and as a starting point from which to obtain new data. If this were not so, all the laboratories in the world would be constantly reinventing the wheel. In order to make sure inside the laboratory that the results are reliable, a basic and quite intuitive rule (among many) is that the same experiment must be done several times, and every time - or most of the time - the result must be similar. Thus, the researcher can assume, with a high degree of confidence, that the phenomenon discovered is not due to chance, or to the fact that it just rained that day, or to the fact that a drop of substance X just fell on the test tube with sugar, flowers and many colors, and without wanting it, the girls' three favorite feminist heroines of the '90s were created. The degree of 'similarity' between these results is never going to be 100%, but with statistical tools we can estimate it, put a limit on it and rest assured that the experiments are working.

But for the rest of the scientific community, who does not see the experiments, how can we communicate that the results obtained are real and that their interpretation is correct? In order to make the whole science communication system work, the so-called peer review was invented. This basically consists of proposing a discovery in the form of a written article, with a predetermined format, detailing why, how and for what purpose the experiments were carried out, showing and interpreting the results. This proposal, called a manuscript, is sent to a journal that has a staff of scientists (the 'peers') who are experts in the subject or subjects related to the article, who read it and criticize it. Depending on the quality of the results and the manuscript, the reviewers may accept it, or they may decide that the manuscript does not meet the standards of the journal, in which case it is not published, and they may even suggest new approaches or more experiments to make the findings robust. Thanks to this correction by the experts, the manuscripts are put to the test, the data become reliable and the scientists recognized.

As humans love to measure themselves, 'rankings' of the best and worst journals were quickly built based on how expert the scientists who review them are, how important the discoveries are that are published in them, and how many people read and quote them. Thus, the 'giants' are the magazines we have all more or less heard of: Nature, Science, Cell; while smaller or more specialized magazines are a little (or quite) less known. Nowadays, moreover, when a researcher competes for a subsidy, a scholarship, or a position in an institute or in a university, normally one of the first points of his curriculum that are evaluated is how many articles he published and in which position of the ranking the journals in which he did it are.

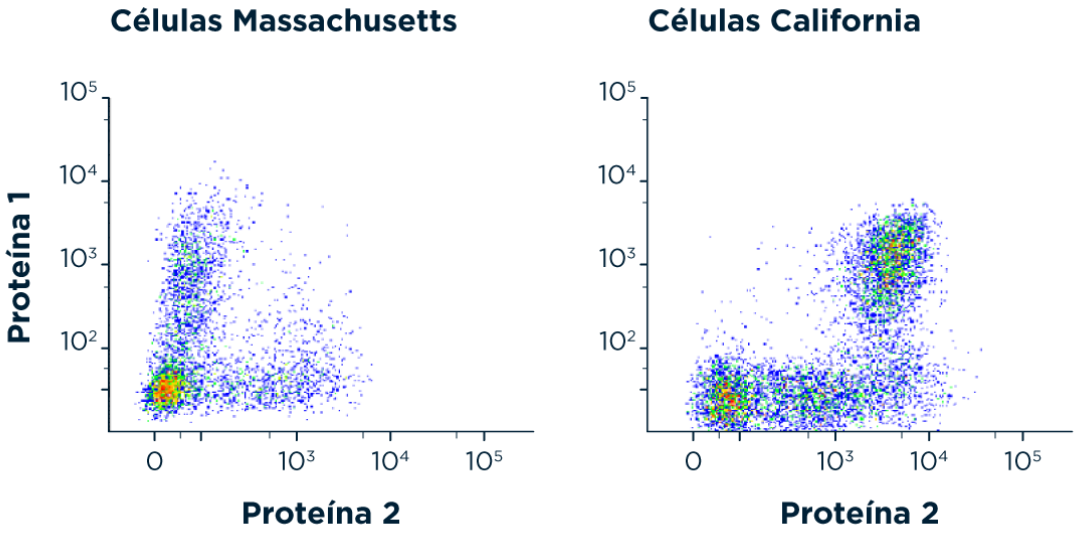
In the east is the Agitation

In 2013, two U.S. laboratories, one in California (west coast) and one in Massachusetts (east of the country), collaboratively worked on a breast cancer project. In the first instance, they obtained healthy breast cells and proceeded to study them with two different approaches: in one of the laboratories they were going to break the cells to see what was inside while in the other

laboratory they needed them whole to see how they behaved while they were alive. The laboratory in California had years of experience in one method, while in Massachusetts they were specialists in the other. So, the idea was that each laboratory would obtain the cells independently, produce data with the technique they best handled and share them among themselves to move forward with the joint project.

In order to study the cells, they had to extract them from the patients using biopsies. A biopsy is a piece of tissue removed from a patient, with different types of cells, a certain organization, a structure. Then, in both laboratories, the first step was to obtain isolated cells, which implied breaking this structure in order to release them.

They were taking that first step when they realized that the cells isolated in California were very different from those obtained on the other side of the country. Every time they processed biopsies in the west, they got the same type of cells, and every time they were isolated in the east the cells obtained were the same, but the shape, the size, the characteristics of their surface (what we would call cell profile) in the west were totally different from those in the east. In fact, one of the most important differences, which was evidenced by a routine test, was the type of proteins that the cells had on their surface. These proteins are involved in processes of cellular communication, migration, response to external stimuli, etc. The fact that the protein repertoire was not the same in both laboratories was worrying because it could affect the behavior of the cells in future experiments.



In these cell profile graphs, each point is a cell. The further to the right or higher up they are, the more of a given protein they present in their membrane (i.e. on the surface). It is clear that the difference was too big to play dumb.

The fact that a result cannot be repeated in other laboratories is normally a source of anxiety, panic and professional crisis among researchers. And it is not uncommon to hear: it happens more often than scientists, investors, pharmaceutical industries and society in general would like. The whole system is based on reproducibility, i.e. that a result can be achieved again

regardless of who runs the experiment or where it is performed, if the same steps are followed. It is one of the strongest proofs that a discovery is robust, and not a coincidence or a fraud. Today, however, the monster of reproducibility lurks in every corner, in every laboratory of every discipline. Therefore, when describing the discoveries of a research group, it is very important to explain in detail how the experiments were made. It is not enough to comment (for example) what techniques were used and to declare with which animal or which type of cell one is working, but it is also necessary to record data such as where they were acquired, which country they come from, from which brand and origin all the drugs and reagents used in the study are, which genetic modifications are used, etc. The more details are given, the more information there is for other research groups to repeat the experiment in the same conditions and check that the result is the same, thus contributing to its reproducibility and preventing the kind of problems that brought so many headaches to Dr. Jekyll, or at least detecting them in time.

And now I'm going to test in my body this experiment that I did with supplies of dubious origin.

This level of detail that journals require for the publication of manuscripts has increased over the years, because a problem that was quite uncomfortable for the scientific community began to go public. Here is what happened during the collaboration between the laboratories of California and Massachusetts.

The project was very ambitious and there was quite a lot of money invested in it (and that meant it could be an important brick in the wall of knowledge). So, they decided to stop the experiments that were going on to determine why the cells obtained at different places were so different, and to try to solve that issue before doing the experiments that really interested them.

Because, in addition, a problem without a solution, a question without an answer, is the gasoline of the machine called 'scientific vocation'.

What could be the reason for such a big difference? The possibility that the differences were in the biopsies was discarded from the beginning, because the design of the experiment required a certain level of similarity between the people from whom samples could be extracted, as well as in the way of obtaining them. They then imagined a list of possible factors that could differ between laboratories:

- The equipment used for analysis.

- Calibration of such equipment.

- How to obtain the sample.

- The origin of inputs and reagents.

- The brand of these inputs.

- The way to process the tissues to obtain isolated cells.

- Methods of cell characterization.

Both groups did extra work to check that instruments of the same brands were used, they recalibrated the equipment using the same criteria, they changed the reagents, culture media and inputs, so that they were all the same. Even when all this failed, they exchanged reagents (to be sure that the manufacture of the reagents in the east and in the west was not different, although they were of the same brand). None of these measures managed to bring them any closer to an explanation of what was happening.

The clue that helped them to get out of the mess of problems, that was already beginning to make them waste time, money and motivation, appeared a year later, when they decided to process the samples in California and send them to Massachusetts for analysis and discovered that the cells were much more similar to those always obtained on the west coast. Therefore, they concluded that the difference was in the preparation of the cells. The next step was to obtain a sample, divide it into two equal parts, and have one person from each group in the same laboratory, elbow to elbow, process one half of the sample as usual. There they discovered that, although the methods seemed, in theory, exactly the same, there were details (which they believed to be insignificant at first) that changed everything: with exactly the same reagents, the 'California method' generated 'California cells', while the 'Massachusetts method' generated (at this point it sounds logical, but involved many headaches and pocket money) 'Massachusetts cells'.

To make it short, after sitting down to review in detail each step of the methods used to isolate and characterize the cells, they found that the problem laid in the way the test tubes were shaken. If they were lovingly shaken, California cells were generated, and when they were shaken more

vigorously, they obtained Massachusetts cells. The reason for such a difference is not very clear, but the scientists responsible for this study try to explain it from a possible "fragility" of one of those surface proteins, called CD44, which does not withstand the shake and is easily released from the cells. The point is that the matter was finally resolved and the methods were unified, but the time and money that were lost teaches us a beautiful lesson.

Lose the game

What if these two laboratories hadn't taken the time to think about why the measurements were so different? Or if they hadn't been collaborating and generated each their own results? Maybe each laboratory would have continued its experiments without knowing that the differences between the cells were due to the way the biopsies were processed. Perhaps those differences between the cells would have impacted their behavior in front of a drug, and it would not be unreasonable to state that the California group could have concluded that this compound kills the cells whereas in Massachusetts they would have discovered that the same drug does not affect them. Most likely, each laboratory would have published its results independently and would have contributed to the ocean of contradictory information that exists today.

There are so many cases similar to this one that we are talking about a real crisis of reproducibility, which is wasting our time (and money) as humanity. At the same time that we are developing more sensitive tools and more specific knowledge, we run the risk of losing an integral view of the problems addressed, giving rise to this type of inconvenience. If an experiment cannot be reproduced by other people, is that okay? Does it count? In a case like this, who would we believe? Those who would say that the drug kills cells or those who would say otherwise?

All this added to the current paradigm of 'publish or perish', which is related to that need to publish a lot in good journals in order to obtain funding, equipment or workspace, generates an enormous amount of contradictory information. In principle, this would not be such a serious problem (indeed, it could be a good push towards new experiments and more interesting discoveries) if there were a habit of finding out the reason for the contradiction. But that's not what usually happens. You don't always check the reasons for that contradiction, partly because there's no time in the paper production line to take the trouble to think why other people can't reproduce the results you get: 'their problem, after all, I've already published my paper'. And can't a difference as simple, as insignificant, and unnoticed as the way to shake a tube, be one of the possible reasons why there's so much contradictory information?

Does this mean that then everything is a lie and that nothing is good for anything? Not at all. Science, and the hypothetical deductive method as a scientific methodology, remain until now the most effective instruments in the search for 'The Truth', or something that at least resembles it enough. Scientific discoveries are constantly tested inside and outside laboratories: during new experiments, in technological developments, in medical treatments. Let us not forget that, if the bricks are not solid, the wall falls by its own weight. And, without going any further, it was thanks to science that it was possible to arrive at an explanation and with it at a solution to the problems of reproducibility in this case. Because although in science there is always 'room

for doubt', there is also always an experiment or an observation that helps to clear them with arguments. Researchers spend an enormous number of hours a day in a laboratory. For some people, publishing an original work with an interesting and new discovery is the end of the journey of a project of years. The truth is that it's just the beginning: each answer generates more questions and you never know when an innovative development can be turned into cutting-edge technology to solve real-life problems. Science is not a career whose ultimate goal is a discovery; in reality it is more like a career of posts, without end, where each goal is always a new starting point. To lose sight of this is to lose the game.

Recorded reports

Some dynamics are beginning to be applied in certain areas to address the issue of reproducibility. There are journals that admit what are called 'registered reports', and that arise as an alternative to solve problems such as the one mentioned here. These are publications that have two instances of peer review: first, the hypotheses to be tested and the methods and protocols to be followed for the experiments are published. This is reviewed and criticized by other professionals, and the results obtained with these methods remain to be reported, which will then also be subject to revision. This allows the steps to be tested before beginning to invest time and money in the experiments. Furthermore, the approval of the protocol implies a 'pre-acceptance' of the work, i.e. researchers have an additional incentive to publish the results of these trials, whatever they may be. Ah, because there's that problem too: many journals prefer not to publish what are called 'negative results', those that demonstrate the lack of effect of the factor under study. This implies tons of 'this doesn't even tickle this disease' conclusions hidden in a closet and which, if published, would also save the scientific community time and money. In the same vein, another advantage of recorded reports is that, if the project and methods are not good enough to be pre-approved, the scientist may decide to recalculate in time before getting down to work on generating useless or uninteresting results.

From the laboratory, the best thing to do is to be in permanent contact with people with other training, such as statisticians, bioinformaticians, pharmacists, biotherium specialists, doctors, biologists, epistemologists and even philosophers, so that from all possible disciplines constructive criticism is made and projects are strengthened before converting them into data, results, conclusions and publications with real force. If we aim for the results to be taken by future researchers, by States that design policies based on these evidences and by companies that generate products that directly impact people's lives, the responsibility of the scientific community is to constantly dialogue, exchange methods and protocols, generate true collaborations and produce reliable knowledge for the construction of this wall. In short, to make data bricks that are not only solid, but also make sense when intertwined with the results of the rest of the world, because no laboratory is an island and a brick is not a wall.

W. C. Hines, Y. Su, I. Kuhn, K. Polyak, and M. J. Bissell, "Sorting Out the FACS: A Devil in the Details," CellReports, 2014.