# Why an open mind on open science could reshape human knowledge

In the year 1610, Galileo observed a ring-like shape around a distant planet (Saturn). After realising the significance of his discovery, Galileo wanted to record it to be able to claim it as his own contribution once it was announced. To do that, he wrote a letter to a colleague stating the following: "*smaismrmilmepoetaleumibunenugttauiras*"*.* That meaningless text was Galileo's **encoding** of what he really wanted to say: "*Altissimum planetam tergeminum observavi*", which translates to: "*I have observed the most distant planet to have a triple form*". At the time, encoding was the only way scientists could guarantee no one stole their findings.
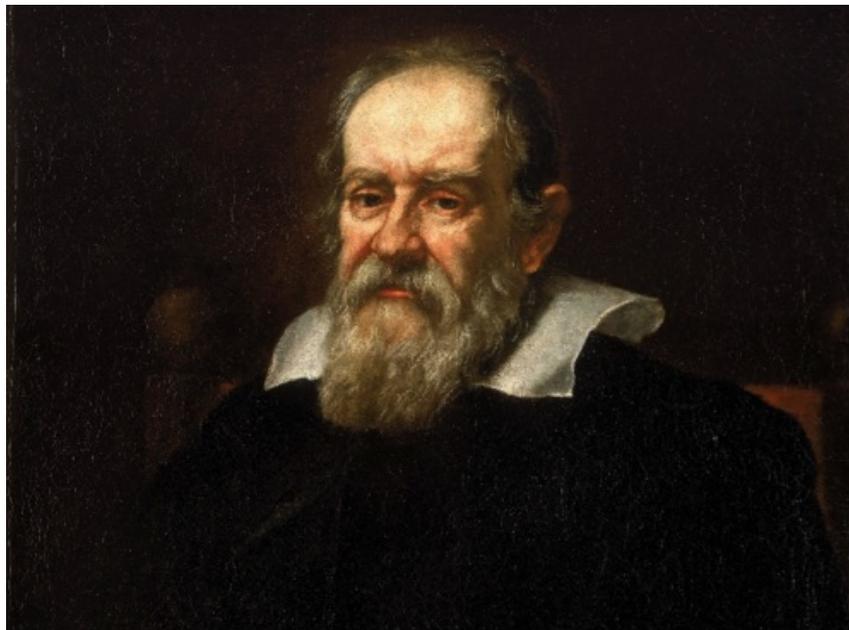

Portrait of Galileo Galilei

Luckily, these days, scientists have many venues where they can claim ownership and declare victory on solved problems. Publication in journals, conference proceedings, surveys, studies, reports, white papers, books, and other forms of documentation are official outlets for scientific discoveries. Such publications constitute the *body of knowledge* of a certain *scientific domain*; contributing to this 'body' is not arbitrary, it is a rather well-defined process. Peer-reviewing for example is one method of evaluating research submissions before they are accepted (for a journal publication or a presentation at a conference).

The importance of the mentioned publications stems from the notion that they form the basis of reference for many aspects of our lives. Most of the time, we tend to point to studies and use them as part of a professional or friendly discussion, in policy making, on TV, radio, media debates, and in decision-making. An experiment is usually required for the publication to be adopted and/or deemed *scientific*. The results of an experiment indicate a success or a failure of a certain proposed hypothesis.

It is common to see published studies presented in other "*cooler*" forms (such as in websites, magazines, or newspapers). We often see titles such as: "*a glass of red wine equals one hour at the gym*", "*artichokes reduce the risk of cancer*", or the famous (but inaccurate) "*green jelly beans cause acne*"; among many other surprising **correlations** that we sometimes accept, share, and even use for life-changing choices.

Figure 2: Green Jelly Beans Cause Acne
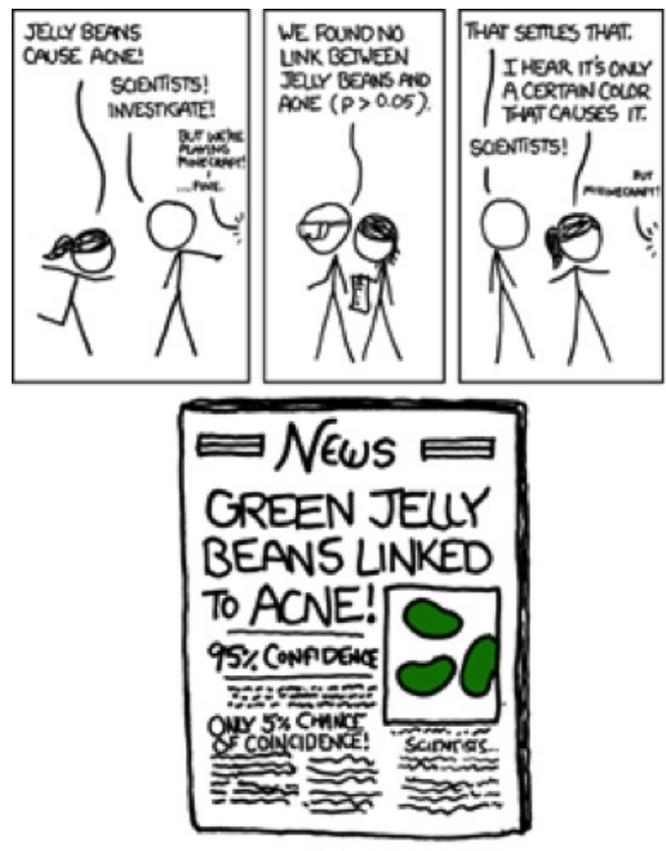
**By xkcd, under a CC-BY-NC-2.5 licence**

The sad reality is that some inaccurate studies are cherry-picked by profit-driven media organisations, and sometimes find their way into conventional wisdom and become, for many, facts. A recent poll showed that **65 per cent** of experimental results in journal articles cannot be reproduced if executed again. Furthermore, **75 per cent** of medical studies suffered from a false-positive outcome, but what does that mean? If someone runs enough tests, a false-positive will almost certainly emerge. For example, assume that you flipped a coin six times in January and got *four heads* and *two tails*; and then flipped the coin again in February, and got *four tails* and *two heads*; can you then make the following claim: *"In January,* heads *have double the chance of emerging when flipping a coin"*? The answer is no; because if you flip the coin one million times every month, all months will get almost 50 per cent *heads* and 50 per cent *tails*. The point here however is that flipping a coin one million times every month is very expensive! And so is running most scientific experiments.

Moreover, and in many cases, outcomes are only accurate when measured with a set of pre-existing conditions or biases (such as a recency bias or an anchoring effect). Other experiments might have data biases – such as: data sampling methods, model overfitting, or using a subset of a dataset that yields to the desired outcome of an experiment.

Some mathematical principles however, such as statistical confidence, significance, p-value, and correlations are used to evaluate experimental results. Those principles can help evaluate the correctness of research. The main problem is that there are many other *institutional* reasons why most reported scientific experiments tend to be inaccurate:

- The main career currency for academics is publishing: Academics are encouraged to publish as many papers as possible (i.e. publish or perish!). Hence, in some cases, *more papers* with less impact are better than fewer papers with noticeable impacts. The **impact factor** is one way of measuring the importance of a publication, but that is ignored most of the time.
- There is little to no value in reproducing or verifying published research. Although duplicating an experiment would greatly benefit a scientific field, replicating studies is not rewarded in the scientific community.
- Scientists don't get paid for peer-reviewing submitted works: There is no unified standard for reviews. Reviews, in many cases are not extensive, and can be limited to one or two superficial sentences (something most scientists have witnessed at some point of their careers).
- Science and technology policies: we take *Open Internet* (freedom to browse the web without explicit consequences) for granted. **Net-Neutrality** has protected us from the domination of internet providers, and any potential control of our seamless pursuit of everything that is online! The biggest and most recent danger to our liberties and freedoms would be taking that right away – this is a heavily debated topic these days.

Figure 3. Freedom of the internet status (87% of world population)
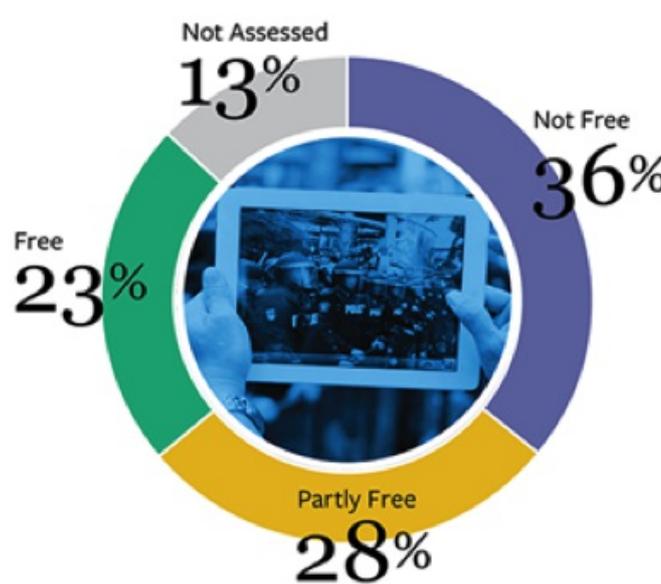
- Most importantly, **Open Science**: There is a serious lack in sharing material/resources that are used to creating the building-blocks of a research experiment. These blocks include: **1**. the raw **data** used in and experiment, **2**. the **pre-conditions** of an experiment, **3**. the **code**, **4**. **software** used, **5**. **experimental setup**, and **6**. **environmental information**.

Therefore, for science to be open (Open Science), those **six** pillars (and potentially others) need to be available (*open*). That is essential in addressing the increasing demand in re-evaluating the current scientific process. In the early 2000s, the internet's own boy, Aaron Swartz *(please see the author's note in the comment section)*, took a stab at creating a repository of scientific papers for free and public access; and established the **Open Access Manifesto** – however, Aaron's story didn't have a fruitful or happy ending. He was prosecuted with two counts of fraud, before committing suicide in 2013.

There are some silver linings in the story of *Open Science*. Most journals now present a form of free-to-the-public publishing through what is known as Open Access. Open Access is usually expensive for the authors (it could cost a couple of thousand dollars to publish a paper), albeit free to the public. Another 'open' initiative is the Open Data Initiative. Open Data demands federal agencies to share their public data in machine-readable formats with the public. Many journal publications have recently been asking the submitters to send their data sets along with the paper. As for code and software, initiatives such as Open Source, Open AI, GitHub, and Google Open Source have been advocating, publishing and hosting software code with the public (for free).

Figure 4. Silver linings in Open Science

If you believe that access to science is a human right (as I do!), *Open Science* is the direction that we all need to adopt. All scientists, researchers, and publishers need to participate in the process of 'opening' the sciences to the masses. The mentioned pillars (Open Data, Open Source, Open Access, and Open AI) are possible to accomplish, but the most difficult goal will be the seventh pillar: *an Open Mind on Open Science* by all parties involved.

<div align="center">♣♣♣</div>

Notes:

- *The post gives the views of its authors, not the position of LSE Business Review or the London School of Economics.*
- *Featured image credit:* Science is great, by Martin Clavey, under a CC-BY-SA-2.0 licence
- *When you leave a comment, you're agreeing to our Comment Policy.*

Feras A. Batarseh is a research assistant [professor](#) with the College of Science at George Mason University (GMU), in Fairfax, Virginia. His research spans the areas of data science, Artificial Intelligence, and context-aware software systems. Dr. Batarseh obtained his Ph.D. and M.Sc. in Computer Engineering from the University of Central Florida, and a Graduate Certificate in Project Leadership from Cornell University (2016). His research work has been published at various prestigious journals and international conferences. Additionally, he published and [edited](#) several book chapters. Dr. Batarseh has taught data science and software engineering courses at multiple universities including GMU, UCF as well as George Washington University (GWU). Prior to joining GMU, Dr. Batarseh was a Program Manager with the Data Mining and Advanced Analytics team at MicroStrategy, Inc., a global business intelligence corporation based in Tysons Corner, Virginia. During his tenure, he helped several clients make sense of their data and gain insights into improving their operations. For more information on his research, and contact details, please refer to this webpage: