

Science is best when the data is an open book

October 23 2015, by Alex O. Holcombe



Data needs to be an open book if science is to be made more reliable. Credit: Quinn Dombrowski/Flickr, CC BY-SA

It was 1986, and the American space agency, NASA, was reeling from the loss of seven lives. The space shuttle Challenger had broken apart about one minute after its launch.

A Congressional commission was formed to report on the tragedy. The physicist Richard Feynman was one of its members.

NASA officials had testified to Congress that the chance of a shuttle

failure was around 1 in 100,000. Feynman wanted to look beyond the official testimony to the numbers and data that backed it up.

After completing his investigation, Feynman summed up his findings in an appendix to the Commission's official report, in which he [declared](#) that NASA officials had "fooled themselves" into thinking that the shuttle was safe.

After a launch, shuttle parts sometimes came back damaged or behaved in unexpected ways. In many of those cases, NASA came up with convenient explanations that minimised the importance of these red flags. The people at NASA badly wanted the shuttle to be safe, and this coloured their reasoning.

To Feynman, this sort of behaviour was not surprising. In his career as a physicist, Feynman had observed that not just engineers and managers, but also basic scientists have biases that can lead to self-deception.

Feynman believed that scientists should constantly remind themselves of their biases. "The first principle" of being a good researcher, according to Feynman, "is that you must not fool yourself, and you are the easiest person to fool".

Many eyes

A scientist can build a career out of a theory, and then find she has a lot riding on that theory being true. And even those of us who are less theory-bound still hope that each new data point will support our current theory, even if we only thought of that theory yesterday.

In the official report to Congress, Feynman and his colleagues recommended an independent oversight group be established to provide a continuing analysis of risk that was less biased than could be provided

by NASA itself. The agency needed input from people who didn't have a stake in the shuttle being safe.

Individual scientists also need that kind of input. The system of science ought to be set up in such a way that researchers subscribing to different theories can give independent interpretations of the same data set.

This would help protect the scientific community from the tendency for individuals to fool themselves into seeing support for their theory that isn't there.

To me it's clear: researchers should routinely examine others' raw data. But in many fields today there is no opportunity to do so.

Scientists communicate their findings to each other via journal articles. These articles provide summaries of the data, often with a good deal of detail, but in many fields the raw numbers aren't shared. And the summaries can be artfully arranged to conceal contradictions and maximise the apparent support for the author's theory.

Occasionally, an article is true to the data behind it, showing the warts and all. But we shouldn't count on it. As the chemist Matthew Todd has said to me, that would be like expecting a real estate agent's brochure for a property to show the property's flaws. You wouldn't buy a house without seeing it with your own eyes. It can be unwise to buy into a theory without seeing the unfiltered data.

Many scientific societies recognise this. For many years now, some of the journals they oversee have had a policy of requiring authors to provide the [raw data](#) when other researchers request it.

Unfortunately, this policy has failed spectacularly, at least in some areas of science. Studies have found that when one researcher requests the

data behind an article, that article's authors respond with the data [in fewer than half of cases](#). This is a major deficiency in the system of science, an embarrassment really.

The well-intentioned policy of requiring that data be provided *upon request* has turned out to be a formula for unanswered emails, for excuses, and for delays. A data *before request* policy, however, can be effective.

A few journals have implemented this, [requiring](#) that data be posted online upon publication of the article.

Open Data Week?

Adoption of this new data-posting policy has been slow, held back by a second defect in the system of science. Currently, researchers are rewarded – in the form of job promotions, and grants – for their articles announcing their findings, but not for the data behind the articles.

As a result, some scientists hoard data. With each data set, they publish as many articles as they can, but resist publishing the data itself.

To fix science, we need to change these incentives: sharing data should be rewarded; providing a critical re-analysis of data should be rewarded; poking holes in others' claims about a data set should be rewarded.

If the returns of professional scepticism can be increased, science will waste less time pursuing false theories.

As I write this, we are nearing the end of the eighth International Open Access Week. This is a week to celebrate that increasing numbers of scientific articles are available for free rather than being published behind paywalls, and a time to advocate for more.

Open access to articles is important, but we need to open up the data too. Do we need to start an international Open Data Week? In a better system of [science](#), [data](#) sharing would be de rigueur.

This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).

Source: The Conversation

Citation: Science is best when the data is an open book (2015, October 23) retrieved 19 April 2024 from <https://phys.org/news/2015-10-science.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.