

COMMENT

POLICY There must be more to the future of transport than the driverless car **p.29**

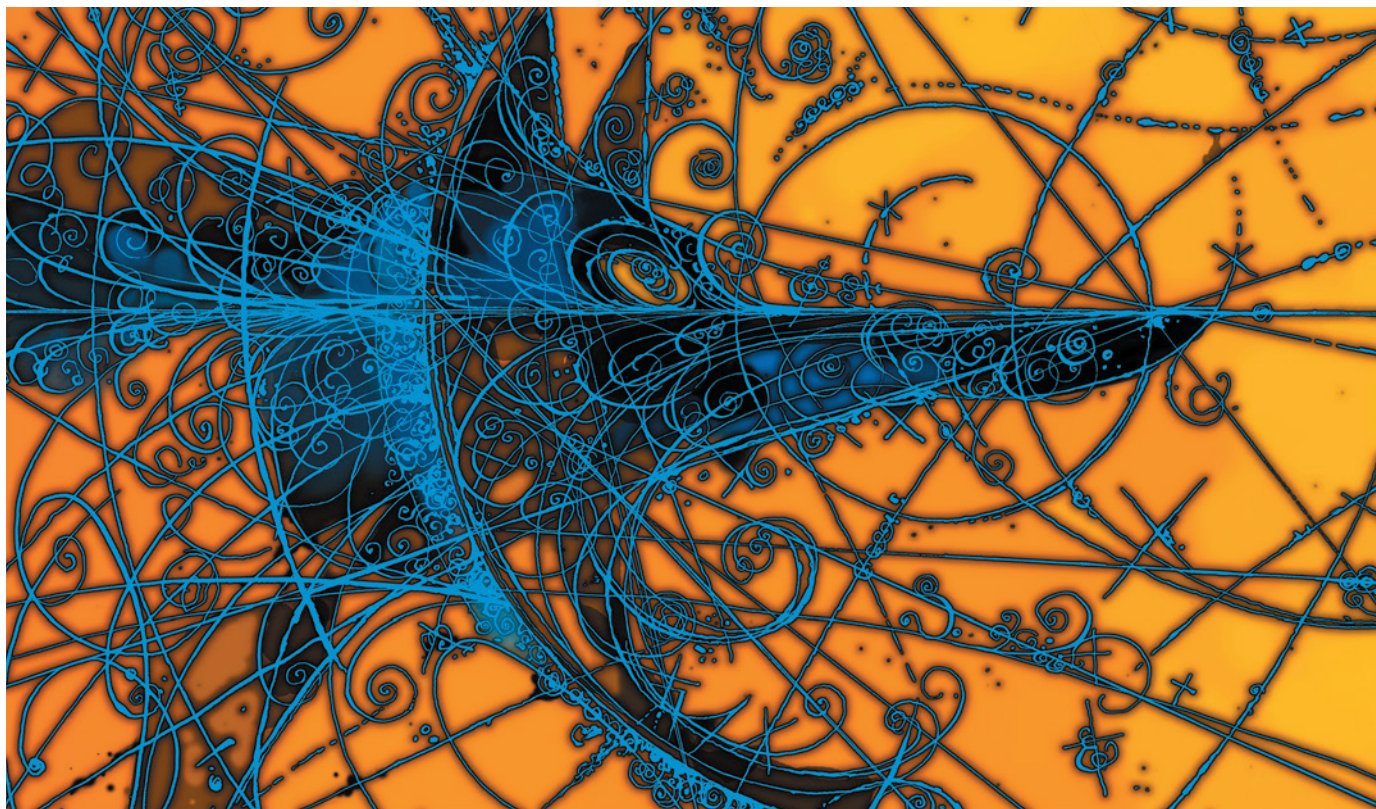
ENERGY Clarity needed about power output of renewables per unit area **p.32**

PALAEONTOLOGY Who is hominin discovery centre for? **p.33**



IMMUNOLOGY A humanized history of monoclonal antibodies **p.34**

ILLUSTRATION BY DAVID PARKINS



Don't cry wolf

Tighten the requirements for declaring physics breakthroughs, says **Jan Conrad**.

The past few years have seen a slew of announcements of major discoveries in particle astrophysics and cosmology. The list includes faster-than-light neutrinos; dark-matter particles producing γ -rays; X-rays scattering off nuclei underground; and even evidence in the cosmic microwave background for gravitational waves caused by the rapid inflation of the early Universe. Most of these turned out to be false alarms; and in my view, that is the probable fate of the rest.

There are consequences to broadcasting seemingly extraordinary results to peers and the public before they are reviewed, or despite knowing that better data are just around the corner. Colleagues who once got excited now shake their heads and joke about 'yet another dark-matter candidate'. The field

has cried wolf too many times and lost credibility. One colleague told me that granting panels are becoming wary of funding astrophysical searches for dark-matter particles.

I also worry that false discoveries are undermining public trust in science. As cosmic phenomena come and go — not to mention endless speculation about hypothetical concepts such as parallel and holographic universes — why should anyone believe that any scientific result will hold?¹

Several trends have brought us to this state of affairs. Intense competition, increased use of public data sets and online publishing of draft papers without proper refereeing have eroded traditional standards for making extraordinary claims.

Particle physics and astrophysics pioneered

the open release of data and publications more than two decades ago; other disciplines are following their lead. The scientific community must now address the habits that have crept in to ensure that enticing reports of false discoveries do not overwhelm more sober accounts of genuine scientific breakthroughs.

SHIFTING PRACTICES

Three changes in the ways that scientific studies are done and reported are fuelling this rash of false discoveries.

First, statistical standards have fallen. Extraordinary claims demand extraordinary proof. In particle physics, the usual threshold is '5 sigma': a signal 5 times stronger than the average noise level (sigma), which translates to a roughly 1-in-3.5-million probability ▶

► that the results were due to chance. But 5-sigma claims are becoming rare as scientists rush to assert priority with exciting but tentative results. The July 2012 official announcement of the discovery of the Higgs boson with the Large Hadron Collider at CERN, Europe's particle-physics lab near Geneva, Switzerland, was preceded by press releases of weak but suggestive indications even though there was no competition.

That scientists change the wording in their papers from 'discovery' to 'evidence' or 'indication' has little influence on how the results are used. Take the latest dark-matter discovery claim. On 8 March, astronomers posted a preprint of a paper in the arXiv repository, and their university issued a press release reporting what the authors called a "tantalizing" sign of γ -rays coming from a recently found dwarf-galaxy companion to the Milky Way that is allegedly rich in dark matter². The γ -ray signal, found in images from the Fermi γ -ray satellite's Large Area Telescope (LAT), seemed to be consistent with high-energy radiation produced when particles of dark matter annihilate. But the photon excess of only 3–4 times the noise level was inconclusive, as the authors acknowledged.

Another paper posted on arXiv the same day disfavoured the discovery. A more comprehensive re-analysis of the same data by the Fermi-LAT instrument team³ — using updated software 30–40% more sensitive — recorded no signal beyond noise. The authors of the first paper acknowledged that the software upgrade was imminent and would confirm or refute their claim, but did not wait for it.

Detecting a noise fluctuation is nothing new, but the possibility that the 'detection' might have been dark matter meant that it was widely reported in the media. Even balanced reporting raises the issue in the public's mind; the account in *The New York Times*⁴ mentioned the non-detection, but the hint of excitement drove the story.

Second, the greater use of public data sets increases the risk that some researchers will make spurious detections near the edge of an instrument's sensitivity. More brains may be picked to mine the data. But analysis is difficult without inside information from those who built and calibrated the instrument.

That was the case with the Fermi-LAT dark-matter detection. The released Fermi-LAT data — public since 2009 — are the product of complicated algorithms and calibrations that turn the electronic signals of detectors into quantities that any physicist can in principle analyse. The instrument builders, however, have the know-how to push the noise limits down.

The risk that someone will misuse data also grows when more people have access to them. Even the largest collaborations cannot police discoveries made by outsiders using

their data. Even if they internally re-run the analysis, the damage is done once an erroneous result has been made public.

Third, many more papers are now released on preprint servers such as arXiv (which had about 100,000 submissions in 2014), and press releases are sent out before peer review. Competition for positions, funding, career metrics such as the h -index and prizes drives

“Metrics need to be devised that distinguish citations of discredited claims.”

the rush to publish prematurely and publicize results. Incorrect papers posted on arXiv do more than add to the noise of irrelevant results. Funding decisions are skewed; theorists waste a lot of time trying to devise explanations; and the public is misled through news reports.

A striking example of a premature claim released online before peer review was the report last year of evidence for gravitational waves and cosmic inflation — the Universe's rapid expansion in the instant after the Big Bang — by the BICEP2 microwave telescope at the South Pole. The detection of a swirled polarization pattern, known as a B-mode, in the cosmic microwave background (radiation left over from the Big Bang) was not in doubt — it had a 7-sigma signal⁵. But its supposed cosmic origin turned out to be false. It was shown six months later — with data from the European Space Agency's Planck satellite — to be warm dust in the Milky Way⁶. Again, I believe that the authors of the original paper must have known of the impending Planck data but chose to blow their trumpet ahead of confirmation. Eventually, the BICEP2 and Planck collaborations worked together to arrive at a solid result⁷, an approach that should have been considered from the beginning.

QUALITY CONTROL

To avoid further weakening of scientific standards and reputations, researchers need to stick to scientific best practice.

A first step would be for physicists to make sure that they apply the 5-sigma rule (or an equivalent) for firm discoveries. Online posting should not be elided with publication. It is premature to announce an important finding to the public at the same time as it is announced to scientific peers. Critical examination by peers is necessary — not least to avoid personal biases.

As long as online posting is confused with the release of deeply scrutinized results, quality assurance of preprints posted online should become stricter. An 'endorsement system', whereby users must be endorsed by other users before posting a paper, has been developed by arXiv to ensure that non-scientific pieces are not hosted there. More is needed for extraordinary claims. Named

reviewers for major discoveries would reassure the readers and authors, as well as crediting the reviewers. Journals should discourage the referencing of arXiv papers.

Instrument builders and specialists who collected the original data should review major claims that are based on public data sets, either as referees, advisers or collaborators. Other teams with ancillary data that could refute or prove a claim should be involved in checking major results before release. This will require voluntary good conduct by competitors, which again could be encouraged by naming reviewers on breakthrough papers.

A system needs to be established to reward best practice. Collaborations should establish a way to ensure that a data team working with an individual scientist will not competitively sink the scientist's publication nor diminish their visibility. Internal review should precede announcements of major results at conferences. Policies should be devised for author lists to give proper credit.

Journals and arXiv should find a strategy for allocating credit to the lead scientists in such collaborations. The BICEP2 team, for example, did work with the Planck collaboration later; had they been able to mark their priority better they might have delayed a press conference.

Not surprisingly, the original BICEP2 paper has ten times more citations than the final word; many incorrect papers are more highly cited than counter cases. Academic metrics need to be devised that distinguish citations of discredited claims so that it is not more advantageous to state and retract a result than to make a solid discovery.

Physicists' associations (such as the American Physical Society or the International Union of Pure and Applied Physics) should lead a movement akin to the biology community's reproducibility initiatives. Scientists, publishers and representatives of funding agencies must convene to discuss improvements to norms such as peer review, metrics, use of databases, quality assurance and codes of conduct. ■

Jan Conrad is professor of astroparticle physics at the Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, Stockholm, Sweden.
e-mail: conrad@fysik.su.se

1. Ellis, G. & Silk, J. *Nature* **516**, 321–323 (2014).
2. Gerlinger-Sameth, A. et al. Preprint at <http://arxiv.org/abs/1503.02320> (2015).
3. Fermi LAT Collaboration & DES Collaboration. Preprint at <http://arxiv.org/abs/1503.02632> (2015).
4. Overbye, D. *The New York Times* 'Gamma Rays May Be Clue on Dark Matter' (11 March 2015).
5. Ade, P. A. R. et al. *Phys. Rev. Lett.* **112**, 241101 (2014).
6. Planck Collaboration. *Astron. Astrophys.* **576**, A107 (2015).
7. Ade, P. A. R. et al. *Phys. Rev. Lett.* **114**, 101301 (2015).