New scepticism about science

CARRIE FIGDOR ON A PRESSING DEBATE IN MEDIA ETHICS

"Journalism's first obligation is to the truth." In *The Elements of Journalism*, a handbook of standard journalistic norms and practices, Bill Kovach and Tom Rosenstiel privilege this principle as the goal of objective, "just-the-facts" journalism. Newspeople have developed, and continually refine, professional practices for producing news reports that can reach this standard, as well as professional norms that guide the proper conduct of these practices. For example, getting at the truth requires verification – in epistemological terms, justification – in order to sift the credible from the unlikely. News stories produced by following this and other norm-guided practices are deemed fit to print; they contain the most credible information available and are presented as such.

Reporters are well aware of human epistemic fallibility, but those who have articulated the norms of objective news often point to science as proof that their quest for truth is not quixotic. Of course, the practices of science differ from those of journalism; the latter is largely a matter of transmitting knowledge rather than producing it, and norms for reliable testimony are not the same as those for reliable fact-finding. But the shared goal of objectivity prescribes for both a norm of independent confirmation: if some fact is true, then distinct people ought to be able to verify it independently. In science, this norm is applied at the group level to such verification-motivated practices as replication: the credibility of a published research finding that cannot be independently found by others following the same procedures is undermined. In journalism, it is embodied at the individual level in the verification-motivated practice of independent sourcing within a story: the credibility of a single source that cannot be corroborated by independent sources is undermined.

The problem is not fraud

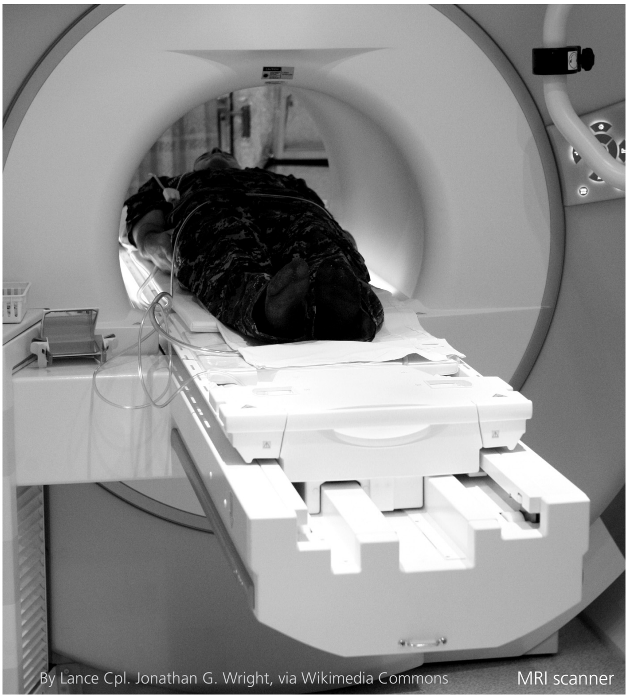
It is therefore of fundamental concern to science reporters, and by extension the public, when scientists themselves say that across many scientific fields there are systematic violations of the norms guiding scientific practices. The problem is not fraud, such as when scientists break rules by faking raw data. Instead, scientists are following the rules for conducting individual studies, but are doing so in ways that collectively violate group-level norms guiding their practices. The result is a collective epistemic disaster, in which most published findings in the affected fields are very likely to be false. They are not preliminary in the sense of "credible but in need of independent confirmation"; they are not credible. Moreover, the affected fields include biomedical research, psychology and cognitive neuroscience – fields whose findings are the bread-and-butter of science news.

This new scepticism stems from various known biases affecting standard practices for generating, submitting and publishing findings. Foremost is the fact that which findings get published is itself a biased sample. Academic science journal editors (as well as researchers) prefer new findings rather than replications, surprising findings, and positive findings. A positive finding is an outcome that is so statistically unlikely to have occurred by chance that one can confidently reject the "null" hypothesis that it did occur by chance. When the probability that an outcome is more or less likely to get published is not independent of factors irrelevant to the nature of the outcome, then the publication record is very likely to be biased. Given the existing preferences, it will contain more false positives than it should. Almost every published finding in the field loses credibility as a result – even if we know that it is likely that the publication record in that field contains many false positives, we do not know which ones they are.

Editorial preferences have distorting or exacerbating effects on the behaviour of scientists. If you want to publish, it is rational to generate and submit as many of the sorts of outcomes that are more likely to be published as you can. As a result, the findings generated and submitted for publication are also more likely to be a biased sample. To borrow from the parable of the blind men and the elephant, it is as if the blind men are given incentives to poke in the area of the trunk: even if each poke is done by the rules, their collective claims about the shape of the elephant are not credible.

Submissions can be biased by withholding negative findings. Failures to replicate a published finding may be quietly slipped into a file drawer (the "file drawer effect"). Negative findings can also include outcomes that do not support a hypothesis ("selective outcome reporting") and outcomes of methodological and statistical analysis pipelines that do not pass the threshold for statistical significance ("selective analysis reporting" or "p-value fishing"). In the latter case, since raw data can be statistically manipulated and analysed in numerous legitimate ways, positive outcomes are bound to be discovered if one searches enough.

Joshua Carp, in his 2012 study "On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of fMRI Experiments", modelled 34,560 legitimate methodological and analytical routes from the same set of raw data to reportable findings and found that locations and levels of statistically significant neural activation differed depending on the route chosen. Such analytic flexibility is more prevalent in fields where tools for generating and analysing raw data are being developed, but it also occurs between the experimental design and outcome reporting of gold-standard randomised controlled trials.



By Lance Cpl. Jonathan G. Wright, via Wikimedia Commons MRI scanner

Collective bias arises if routes that yield positive findings are more likely to be reported. However, published papers typically do not mention alternative pipelines that also may have been tried.

The generation of results can also be biased in various ways. Replications are avoided, so there's no properly working system for flushing out false positives. Moreover, in some fields, many labs are doing similar studies, but not replications, in

There are violations of norms guiding scientific practices

an effort to be the first to get an exciting positive result. This collective practice is guaranteed to produce more false positives, and publication bias makes it likely that more of them will make it into the publication record. Even when replications are done, there is what Michael Smithson (a psychologist at the Australian National University and science blogger) calls "positive-finding replication bias": a first-of-its-kind study reporting a significant result is more likely to be replicated than one that does not, even though both should be at equal rates.

This kind of scepticism about science should be distinguished from the pessimistic meta-induction, a staple topic in the philosophy of science. The pessimistic meta-induction holds that we have no good reason to think that our current scientific theories are true because in the past even our best theories have turned out to be false. The meta-induction tells us to doubt the truth of (for example) our current theories of planetary motion, not to doubt that planets exist (which might be filed under "scepticism about the external world"). The current problem largely affects sciences where the targets of theory are patterns, not objects, and where most research efforts are aimed at trying to find patterns, not theorising about patterns that are already agreed to exist. Relatively speaking, the blind men have it easy trying to discover the elephant's shape, since at least they agree it exists.

Scepticism stems from various known biases

For example, observing George smoking is not observing George's smoking habit, which is a pattern or regularity of behaviour. Nor is observing that George has a smoking habit and that he has lung cancer observing a pattern in the relationship between smoking habits and lung cancer, since George may just be unlucky. Nor is merely observing many people over time sufficient to yield justified claims about the existence of a pattern between smoking habits and lung cancer diagnoses. If we are justified in believing there exists a pattern in this relationship – about which we can go on to theorise as to its causes – it is because in a sufficiently large random sample of the population more smoking habits are correlated with more cases of lung cancer than should occur just by (bad) luck. If our sample is not random – if, say, every observed smoker comes from a cancer-ridden family – we're no better off than the blind men arguing about whether the elephant exists.

John Ioannidis, whose 2005 article "Why Most Published Research Findings are False" forcefully raised scientific awareness of the epistemic consequences of group-level bias, calls a field in which there are no genuine correlations to be found a "null field" – a field where observed correlations are actually random and no pattern exists. Astronomy is not a null field, even if the meta-induction tells us that our theory of the planets is false. Astrology is: its targets are patterns – correlations of birthdays, character traits, and planetary motions – and there are no such patterns to be wrong about. Which of the affected fields are also null fields is an open, and understandably threatening, question.

The nature of the target phenomena can also make it tricky to interpret what we should be sceptical about. In what is sometimes called the decline effect, initially strong and well-publicised results are followed by attempted replications that show a much weaker effect or none at all. In his 2010 article "The Truth Wears Off", Jonah Lehrer discussed this effect using studies of second-generation antipsychotic drugs: early exploratory studies found them to have substantial success in reducing psychotic symptoms, but later studies – especially large randomised clinical trials – showed them to be increasingly ineffective. The decline effect is not necessarily a problem with the scientific method, but with the fact that many (especially small) studies in which the target is a pattern are treated as significant in isolation. Since patterns cannot be discovered in one small go, it's not that the truth *wears off* but that it *takes time*.

Scientists are taking measures to bring their practices back into conformity with the norms guiding them. One such measure is to have scientists pre-register the research methods and analytical pipeline they will be using – randomised controlled trials have followed such reporting guidelines for a decade. New journals dedicated to publishing negative findings and replications have appeared and are being promoted. The most publicised recent effort is the Reproducibility Project in psychology, which aims to generate replications of a sample of studies published in three major psychology journals in 2008 to determine how much publication bias may exist in psychology.

But correction takes time – years, no doubt. In the short run, science reporters face a nasty dilemma. To a journalist following the norms of objective reporting, an official who is known to lie a lot may remain a useful source of tips, but what that person says would not be reported without getting independent confirmation. Even if such a story was turned in, an editor would sit on it until such confirmation was provided. Systematic violations of this norm raise the likelihood of publishing stories that are false. Currently, science journalists are in systematic violation of this norm by reporting most science findings as true or most credible or even preliminary, when they are not fit to print at all. The dilemma arises regarding what they should do to once again be faithful to the first principle of objective journalism.

The first horn of the dilemma is to modify the practices to hew to the norms. Science journalists can simply stop reporting most results in the affected fields until the scientific practices producing these results are norm-conforming. It's not just that *for journalists* the news is what is verifiably likely to be true, not what is verifiably likely to be false, but that this is what the public expects the news to be as well. As usually interpreted, the journalistic norms provide clear guidance: don't touch most scientific results – wait for publication bias to be eliminated. Of course, this is how stories get killed, not published.

Science reporters can still write up large studies and meta-analyses, which analyse the findings of many (typically small) studies. But by selectively avoiding reporting small studies, studies that show small effects, and studies that test the authors' own hypotheses, they would eviscerate non-academic science publications. Moreover, they would likely antagonise the scientific and academic worlds, since careers, professional and financial rewards and university prestige are frequently enhanced by publicity about scientific research. Jobs and relationships with sources would be severely threatened.

Failures may be quietly slipped into a file drawer

The other horn is to modify the norms to hew to the practices. They can continue to report most findings while explicitly emphasising that they are very likely false. This means most science reports would embody a kind of Moorean paradox. In this paradox, one asserts a proposition as true and then asserts that one does not believe it. The usual form of the paradox is "P, but I don't believe that P". In this case, the form of many science news stories would be: "P, but P is very likely to be false". Typically, reports of this form are limited to predictions of the end of the world and those in satirical publications, such as *The Onion*. They are fit to print because the public can be relied on to not take them seriously. Seizing this horn means accepting "P, but P is very likely to be false" as the new normal for science news as well as satire.

It also raises the significant problem that science journalists may exacerbate public scepticism about science at a time when policymakers are already constrained in their ability to adopt policies justified by science. An obvious example is global warming, which is a topic of concern because of the policy implications of accepting not just that the Earth is warming but that human activity is the primary cause. Scientists generally don't do policy; they provide facts. But if many scientific facts (not the global warming facts, which are not undermined by collective bias) are clearly reported as very likely to be false, the public may be even less likely to accept what scientists say in general.

Moreover, the most prominent epistemic alternative to science is religious faith, a form of belief backed by divine authority. In *In Praise of Reason*, Michael Lynch argues that a dogmatic attitude – in which one refuses to provide reasons for belief beyond citing authority – is anti-democratic, because democracy relies on the treatment of citizens as people to whom reasons should be given. If Lynch is correct, by seizing this horn, science journalists risk not just turning off the public but undermining democracy. Since a free press is part of a democratic society, by adjusting their practices in this manner, science journalists may undermine the social conditions in which their jobs are possible.

Other norms guiding science reporting also need to be adjusted in the light of the new scepticism. For example, the practice of balance – providing multiple perspectives on a given issue – does nothing to correct for systematic scientific bias (or for the preliminary nature of many findings, for that matter). This norm is often poorly followed, but the idea behind it is valid enough: to signal caution by showing distinct plausible ways of synthesising the available facts. In science reporting, however, sometimes a scientist from another lab is quoted expressing caution about the study that is the focus of the article, and study authors themselves often caution that the results are "preliminary". But when most published results are likely to be false, quoting a "balancing" scientist or a caution about preliminary results is quite beside the point. True balance would involve seeking out negative findings, among other things, although this would not counteract bias arising from analyses that are not written up or replications that have not been done. And if scientists themselves don't have access to other people's file drawers, it is unreasonable to think journalists will.

However, the first step towards facing this dilemma is to raise awareness among science reporters of the implications for science reporting of the new scepticism. Ioannidis was profiled in *The New York Times* by David Dobbs in 2006, and the *Chronicle of Higher Education* published a story on the Reproducibility Project in April 2012. (Science bloggers have also weighed in.) However, *Science Communication* (as of June 2012) and *Columbia Journalism Review* (searched online, Aug. 6, 2012) had no discussion of Ioannidis' 2005 article or quite generally on how science-based scepticism about science might impact science reporting. *CJR* has had articles on efforts to be more balanced regarding how to cover science that sparks public debate, such as climate science. Such concerns about public scepticism of science are important, but they presuppose that the problem lies only with how to transmit the facts. The deeper problem is with the facts themselves.

Carrie Figdor is assistant professor of philosophy at the University of Iowa; her research interests include philosophy of mind, metaphysics, neuroethics and media ethics