

Buried in bullshit

Tom Farsides and Paul Sparks smell trouble

The amount of energy needed to refute bullshit is an order of magnitude bigger than to produce it. (Alberto Brandolini)

According to Frankfurt (2005), 'liars' and 'bullshitters' both falsely represent themselves as prioritising truth. They differ because liars actively try to hide the truth whilst bullshitters care less about the truth than they do about other things that are potentially in conflict with it.

Let's use the term 'scholars' for people who sincerely prioritise truth.

Note that this cast list is compiled by intentions and endeavours, not by outcomes. All three characters may communicate truth or falsehood irrespective of whether they do so unintentionally, incidentally or purposefully. Note also that there may not be strong relationships between character and competence. People can fall anywhere between ineptitude and finesse at lying, bullshitting and scholarship.

There is a worrying amount of outright fraud in psychology, even if it may be no more common than in other disciplines. Consider the roll call of those who have in

recent years had high-status peer-reviewed papers retracted because of confirmed or suspected fraud: Marc Hauser, Jens Förster, Dirk Smeesters, Karen Ruggiero, Lawrence Sanna, Michael LaCour and, a long way in front with 58 retractions, Diederik Stapel. It seems reasonable to expect that there will be further revelations and retractions.

That's a depressing list, but out-and-out lies in psychology may be the least of our worries. Could most of what we hold to be true in psychology be wrong

(Ioannidis, 2005)? We now turn to several pieces of evidence to demonstrate compellingly that contemporary psychology is liberally sprayed with bullshit (along with some suggestions of a clean-up).

Lies, damned lies and statistics

Almost all published studies report statistically significant effects even though very many of them have sample sizes that are too small to reliably detect the effects they report (Bakker et al., 2012; Cohen, 1962). Similarly, multi-study papers often report literally unfeasible frequencies of statistically significant effects (Schimmack, 2012).

In addition, many of the analyses and procedures psychologists use do not justify the conclusions drawn from them. A striking and common example is failing to correct for multiple tests. If there is a fixed chance of obtaining a statistically significant result (e.g. $p \leq .05$) when there is no genuine phenomenon, the chance of obtaining misleading statistical

significance increases with the number of tests performed. Psychologists routinely fail to correct for multiple comparisons (see Cramer et al., 2014). Apparent results, such as associations between astrological star signs and particular medical conditions, often disappear once appropriate corrections are made (Austin et al., 2006).

Dodgy dealings

So-called 'p hacking' also remains rife in psychology. Researchers make numerous decisions about methods and analysis, each of which may affect the statistical significance of the results they find (e.g., concerning sample size, sample composition, studies included or omitted from programmes of



references

Austin, P.C., Mamdani, M.M., Juurlink, D.N. & Hux, J.E. (2006). Testing multiple statistical hypotheses resulted in spurious associations. *Journal of Clinical Epidemiology*, 59(9), 964–969.

Bakker, M., van Dijk, A. & Wicherts, J.M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7(6), 543–554.

Belluz, J. (2015). How researchers dupe the public with a sneaky practice called 'outcome switching'. Vox. <http://tinyurl.com/zpphys>

Bevan, W. (1991). Contemporary psychology: A tour inside the onion. *American Psychologist*, 46(5), 475.

Billig, M. (2013). *Learn to write badly*. Cambridge: Cambridge University Press.

Blattman, C. (2015). Why I worry experimental social science is headed in the wrong direction. tinyurl.com/zf8384

Bloom, P. (2015). Scientific faith is different from religious faith. *The Atlantic*. tinyurl.com/j3qsmay

Carter, G. (2015). Goals of science vs goals of scientists. <http://tinyurl.com/gv9orld>

Cohen, J. (1962). The statistical power of abnormal-social psychological research. *Journal of Abnormal and Social Psychology*, 65(3), 145–153.

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science*, 1(3), 140216.

Chambers, C.D., Feredoes, E., Muthukumaraswamy, S.D. & Etchells, P. (2014). Instead of 'playing the game' it is time to change the rules. *AIMS Neuroscience*, 1(1), 4–17.

Cramer, A.O., van Ravenzwaaij, D., Matzke, D. et al. (2014). Hidden multiplicity in multiway ANOVA. arXiv

research, variables, potential outliers, statistical techniques). Simmons et al. (2011) vividly illustrate this by reporting a study that 'revealed the predicted effect [that] people were nearly a year-and-a-half younger after listening to *When I'm 64* than they were after listening to 'a control group tune that did not mention age' (p.1360).

For example, evidence is increasingly revealing that alarming numbers of psychologists are willing to admit having engaged in questionable research practices (Fiedler & Schwarz, 2015; John et al., 2012). Many published studies have selectively included or omitted evidence to support claims that authors must know are far from accurately representing the truth, the whole truth and nothing but the truth (Belluz, 2015; Franco et al., 2016; Neuroskeptic, tinyurl.com/j2patqu).

Unconvinced readers can discover for themselves how easy it is to 'Hack your way to scientific glory' by visiting an online tool (tinyurl.com/pjhh5m8) and selecting different sets of variables from a genuine database to find (or 'fail' to find) a significant relationship between the US economy and a particular party being in office.

Some are fighting back against these practices. Inzlicht (2015) blogged about a paper he acted as an editor for: 'emblematic of the old way of doing business, with 7 studies that were scrubbed clean to be near-perfect'. The revision disclosed an additional 11 existing studies, included more appropriate analyses and reported only two significant effects. 'I am a huge fan of this second paper,' Inzlicht wrote. 'I love all my children, but I would be lying if I said that this wasn't my favorite as editor. I love it because it is transparent; and because it is transparent, it allows for a robust science. This push for transparency, of revealing our warts, is exactly what the field needs.'

Why are more editors not following

Meet the authors

'A few years back, we became increasingly uncomfortable claiming expertise in our respective research areas. Increasing numbers of papers were being published, each with a growing number of studies and significant effects, and yet it was getting harder to identify precisely what was done and found in each. How could we be confident about which phenomena were (and were not) real if we couldn't keep up with or even comprehend much of the literature we were supposed to be expert in?

Although we occasionally stumbled across papers expressing dissatisfaction with this or that aspect of empirical practice (power, sample size, null hypothesis statistical testing, etc.), such matters seemed discussed only on the fringes of our discipline by methodologists and statisticians with interests other than understanding psychological processes, per se. Meanwhile, most people seemed to be getting on with business as usual. We did not realise how much our private grumblings were increasingly chiming with a growing zeitgeist.

And then one of us joined Twitter and it became immediately apparent that we were not the only ones struggling. Large parts of our discipline (among others) seem to be in a parlous state. Here, we summarise the problems, as well as various proposed solutions. We hope it will be useful to those who have still not quite grasped the severity of the situation we seem to be in.

Our most fervent hope, though, is that our colleagues can help us. Even if things improve in the future, we want to know what knowledge can we justifiably claim now, e.g. when teaching, making policy recommendations, or seeking grants? One prominent neuroscientist recently suggested that all findings in his field from before 2011 should be more or less dismissed. Should we do similar with swathes of psychological research? Can we continue to make claims based on existing findings from "the science of psychology"? Or will we be rightly called out for bullshit?'

Tom Farsides

is a Lecturer in Social Psychology at the University of Sussex
T.L.Farsides@sussex.ac.uk



Paul Sparks

is a Senior Lecturer in Social Psychology and Health at the University of Sussex
P.Sparks@sussex.ac.uk

Inzlicht's lead? Many researchers and reviewers simply do not have the methodological or statistical expertise necessary to effectively engage in science the way it is currently practised in mainstream psychology (Colquhoun, 2014; Lindsay, 2015). Scientists and reviewers also increasingly admit that they simply cannot keep up with the sheer volume and complexity of things in which they are allegedly supposed to have

expertise (Siebert et al., 2015).

This has long been a problem: Peters and Ceci (1982) changed author names and affiliations and resubmitted 12 manuscripts to 12 high-quality psychology journals which had published the original manuscripts 18 to 32 months previously. The deceit was spotted in three cases. Eight of the remaining nine were rejected, in many cases because of what were identified as 'serious

preprint, arXiv:1412.3416.

Cumming, G. (2014). The new statistics. *Psychological Science*, 25(1), 7-29.

Dienes, Z. (in press). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*.

Earp, B.D. & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, 621.

Engel, C. (2015). Scientific disintegrity as a

public bad. *Perspectives on Psychological Science*, 10(3), 361-379.

Fiedler, K. & Schwarz, N. (2015).

Questionable research practices revisited. *Social Psychological and Personality Science*.

Franco, A., Malhotra, N. & Simonovits, G. (2016). Underreporting in psychology experiments. *Social Psychological and Personality Science*, 7(1), 8-12.

Frank, M. (2015). The ManyBabies Project.

tinyurl.com/jzj5gck

Frankfurt, H.G. (2005). On bullshit. tinyurl.com/6plgw9k

Funder, D.C., Levine, J.M., Mackie, D.M. et al. (2014). Improving the dependability of research in personality and social psychology. *Personality and Social Psychology Review*, 18(1), 3-12.

Garfield, E. (1986). Refereeing and peer review, Part 2. *Current Contents*, 32, 3-12.

Giner-Sorolla, R. (2012). Science or art? *Perspectives on Psychological Science*, 7(6), 562-571.

Ioannidis, J.P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124.

Inzlicht, M. (2015). A tale of two papers. tinyurl.com/hcym3ew

Jarrett, C. (2008). Foundations of sand? *The Psychologist*, 21, 756-759.

John, L.K., Loewenstein, G. & Prelec, D.

opinion

methodological flaws'. But as journals proliferate and incentives to publish increase, academic bloggers like Kevin Mitchell have noted that it becomes even more likely that quantity overwhelms quality (tinyurl.com/zg8sg3k).

Replication and revisionism

Few successful attempts have been made to rigorously replicate findings in psychology. Recent attempts to do so have suggested that even studies almost identical to original ones rarely produce reassuring confirmation of their reported results (e.g. Open Science Collaboration: see www.https://osf.io/vmrgu).

The task of replication is made tougher because researchers control what information reviewers get exposed to, and journal editors then shape what information readers have access to. If readers want further information, they usually have to request it from the researchers and they, their institution or the publishing journal may place limits on what is shared. One consequence of this is that other researchers are considerably hampered in their ability to attempt replication or extension of the original findings. James Coyne blogged last year (<http://tinyurl.com/hjohyp6>) about unsuccessful freedom of information requests to prompt the release of data to allow independent re-analysis of a study that was published in an outlet that explicitly promises such a possibility.

On the positive side, classic findings and interpretations of them that have until now been more or less accepted as 'common knowledge' in psychology are increasingly being challenged and revised (Jarrett, 2008). Yet established and often cherished beliefs are difficult to change. Even when incorrect claims are exposed in ways that should be fatal, they continue to have an influence on subsequent scholarship (Lewandowsky et al., 2012; Tatsioni et al., 2007). Trust in others' testimony is essential in science: this leads researchers and communicators



Panning for gold? Most prestigious journals also have a strong preference for novel and dramatic findings over the replications and incremental discoveries that are typical in an established science.

to report as truths phenomena and theories that they would almost certainly not believe if they critiqued them more thoroughly.

The system is screwed

Traditionally, researchers are much less likely to submit manuscripts reporting experiments that did not find an effect, and journals are far less likely to accept them if they do (Cohen, 1962; Peplow, 2014). Most prestigious journals also have a strong preference for novel and dramatic findings over the replications and incremental discoveries that are typical in an established science. If researchers want to be published in high-ranking peer-reviewed journals, therefore, they are highly incentivised to present highly selective and therefore misleading accounts of their research (Giner-Sorolla, 2012).

The current mechanisms of science production, then, place individual researchers in a social dilemma (Carter, 2015). Whatever others do and whatever the collective consequences, it is in the individual researcher's best economic interest to downgrade the importance of truth in order to maximise publications,

grants, promotion, media exposure, indicators of impact, and all the other glittering prizes valued in contemporary scientific and academic communities (Engel, 2015). This is especially the case when organisations and processes that might otherwise ameliorate such pressures instead exacerbate them because they too allow concerns for truth to be downgraded or swamped by other ambitions (e.g. journal sales, student recruitment, political influence, etc.) (Garfield, 1986).

Future perfect, bullshit present?

There are a lot of current initiatives that advocates claim will make psychology much more reliable and valid in the future. These include measures to improve researchers' methodological and statistical competence (Funder et al., 2014); change the sorts of statistical analyses they use (Cumming, 2014; Dienes, in press); provide pre-registration possibilities (Chambers et al., 2014); promote high-quality replications (Frank, 2015; Open Science Collaboration); facilitate open-access data and materials (Morey et al., 2015); encourage post-publication review (Nosek & Bar-Anan,

[2012]. Measuring the prevalence of questionable research practices. *Psychological Science*, 23(5), 524–532.
Koch, S. (1981). The nature and limits of psychological knowledge. *American Psychologist*, 36(3), 257.
Lakens, D. (2014). Grounding social embodiment. *Social Cognition*, 32, 168–183.
LeBel, E.P., Loving, T.J., Campbell, L. et al. (2015). Scrutinizing the costs

versus benefits of Open Science practices. tinyurl.com/hkfu67e
Lewandowsky, S., Ecker, U.K., Seifert, C.M. et al. (2012). Misinformation and its correction. *Psychological Science in the Public Interest*, 13(3), 106–131.
Lilienfeld, S.O. (2012). Public skepticism of psychology. *American Psychologist*, 67(2), 111–129.
Lindsay, D.S. (2015). Replication in psychological science. *Psychological*

Science.
Marcus, A. & Oransky, I. (2011). Science publishing. *Nature*, 480, 449–450.
Matthews, D. (2015). Secret dossier on research fraud suggests government concern over science. Times Higher Education. tinyurl.com/glrj3up
Meehl, P.E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66(1), 195–244.

Morey, R., Chambers, C.D., Etchells, P. et al. (2016). The peer reviewers' openness initiative. *Royal Society Open Science*. tinyurl.com/gsaqha2
Nosek, B.A. & Bar-Anan, Y. (2012). Scientific utopia. *Psychological Inquiry*, 23(3), 217–243.
Nosek, B.A., Spies, J.R. & Motyl, M. (2012). Scientific utopia II. *Perspectives on Psychological Science*, 7(6), 615–631.

2012); improve dissemination of information about corrections and retractions (Marcus & Oransky, 2011); and change incentive structures (Nosek et al., 2012).

Some are sceptical that all such initiatives will bring net gains (e.g. Blattman, 2015; Earp & Trafimow, 2015; LeBel et al., 2015; Sbarra, 2014). Although we have views on such things, our concern here is less with the future than with the present.

If a plethora of sweeping changes is required to achieve trustworthiness in psychological science in the future, what can psychologists claim on the basis of the research literature now? Are we lying or at best bullshitting when we tell students, grant-awarding bodies, policy makers, the public and each other about things that psychology has discovered (Lilienfeld, 2012; Matthews, 2015)? Are we disingenuous when we trumpet the epistemological superiority of so-called psychological science and its products (e.g. Bloom, 2015)? Given the multiple serious, widespread, and enduring problems we have, can we claim hand-on-heart to confidently know anything and, if so, how can we identify it among all the bullshit and lies?

As it happens, we do think that our discipline has a lot to offer. But we also think that norms of assessing and representing it need to change considerably if we are to minimise our at least complicit contribution to the collective production and concealment of yet more bullshit. Here are some provisional and tentative recommendations.

1. *Don't give up.* Meehl (1990) suggested that problems similar to those identified above make the psychological research literature 'well-nigh uninterpretable'. When convincing others of this, he reported that some gave up studying questions of importance and interest to

study things that were at least amenable to rigorous experimentation, while others used defence mechanisms so that they could carry on as normal and continue to reap rewards while avoiding a guilty conscience. Both strategies seem to us to be unattractive and unnecessary. We believe that psychology has the potential to make unique and important contributions to understanding important phenomena.

2. *Prioritise scholarship.* Psychologists and their institutions should do everything within their power to champion truth and to confront all barriers to it. If we have to choose between maintaining our professional integrity and obtaining further personal or institutional benefits, may we have the will (and support) to pursue the former.

3. *Be honest.* Championing truth requires honesty about ignorance, inadequacies, and mistakes (Salmon, 2003). Denying flaws helps no one, especially if our denials are accompanied by poorly received assertions of invincibility and superiority. Acknowledgement of weakness is a strength. Expertise should be in service of scholarship, not prioritised above it. Expertise idolatry risks encouraging defensive bullshit from the anxious and generating blinkered, dogmatic bullshit from specialists (Frankfurt, 2005; Ottati et al., 2015).

4. *Use all available evidence as effectively as possible.* Important as they are, experiments are neither necessary nor sufficient for empiricism, scholarship or 'science' (see Robinson, 2000). To study important phenomena well, we need first to identify what they are and what central characteristics they have (Rozin, 2001). To study things thoroughly, we need to identify processes and outcomes other than those derived from our pet 'theories'. Evaluating the research literature may

well require skills different from those that have been dominant during much of its production (Koch, 1981). In particular, we have found particularly effective accurately describing others' procedures and outcomes in ordinary language and then examining how well these justify the usually jargonistic 'theoretical' claims supposedly supported by them (cf. Billig, 2013).

5. *Nurture nuance.* Experiments within psychology are usually (at best) little more than demonstrations that something can occur. This is usually in service of rejecting a null hypothesis but it is almost as often misreported as suggesting (or showing or, worst of all, 'proving') something much more substantial – that something *does* or *must* occur. Perhaps the single most important thing psychology can do to quickly and substantially improve itself is to be much more careful about specifying and determining the boundary conditions for whatever phenomena it claims to identify (Ferguson, this issue; Lakens, 2014; Schaller, 2015).

6. *Triage.* Given that at least some areas of psychology seem awash with bullshit, we would be wise to prioritise evaluating topics of centrality and importance rather than on the basis that some reported findings are, for example, recent or amenable to testing using online experiments (Bevan, 1991). 'Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise' (Tukey, 1962, pp.13–14).

The question we chose to head up this section is not rhetorical. We do not consider the recommendations we list as final or complete. Science is a social enterprise and we are interested to hear the views of others with perspectives different from ours. We are certain that something needs to be done, though. We're fed up with all the bullshit.

"can we claim hand-on-heart to confidently know anything?"

Ottati, V., Price, E.D., Wilson, C. & Sumaktoyo, N. (2015). When self-perceptions of expertise increase closed-minded cognition. *Journal of Exp. Social Psychology*, 61, 131–138.

Peplow, M. (2014). Social sciences suffer from severe publication bias. *Nature*. tinyurl.com/nw8bu8q

Peters, D.P. & Ceci, S.J. (1982). Peer-review practices of psychological journals. *Behavioral and Brain*

Sciences, 5(2), 187–195.

Robinson, D.N. (2000). Paradigms and 'the myth of framework'. *Theory & Psychology*, 10(1), 39–47.

Rozin, P. (2001). Social psychology and science. *Personality and Social Psychology Review*, 5, 2–14.

Salmon, P. (2003). How do we recognise good research? *The Psychologist*, 16, 24–27.

Sbarra, D.A. (2014). Forward thinking. *Perspectives on Psychological Science*,

9(4), 443–444.

Schaller, M. (2015). The empirical benefits of conceptual rigor. *Journal of Experimental Social Psychology*.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, 17(4), 551–566.

Siebert, S., Machesky, L.M. & Insall, R.H. (2015). Overflow in science and its implications for trust. *ELife*, 4, e10825.

Simmons, J.P., Nelson, L.D. & Simonsohn, U. (2011). False positive psychology. *Psychological Science*, 22, 1359–1366.

Tatsioni, A., Bonitsis, N.G. & Ioannidis, J.P. (2007). Persistence of contradicted claims in the literature. *JAMA*, 298(21), 2517–2526.

Tukey, J.W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 1–67.