

Kaleidoscope

Derek K. Tracy, Dan W. Joyce,
Sukhwinder S. Shergill

School bullying is a problem that has had an impact on most of us. We know it can have long-term harms, we know it needs to be curtailed, but how? Cyber-bullying is a novel issue on the rise, and the topic is particularly pertinent in the UK where rates of bullying are greater than in other European countries. Different interventions have been trialled, from whole-school policy, through conflict resolution training, to teaching of wider social skills. Writing in the *Lancet*, Bonell *et al* report on a cluster randomised trial across 40 English schools, commencing in year 7, using the ‘Learning Together’ intervention that had aspects of each of these.¹ It involved a facilitated school action group, staff coaching in ‘restorative practice’ and embedding emotional skills in the curriculum. The restorative practice aspect specifically facilitates victims communicating harms endured, and bullies acknowledging and amending their behaviour; it is increasingly used more globally to reduce antisocial behaviour, but had not been subjected to rigorous evaluation in schools.

The 20 schools in the active arm showed a significant reduction in bullying at the 3-year end-point, although the effect size was small and there was not a reduction in overall student reports of aggression. There were also improvements in mental health, well-being and quality of life – particularly for boys – as well as reductions in smoking, alcohol and drug use. The cost was £58 per pupil, a relatively insignificant sum when considering the adverse psychological outcomes of this behaviour. This study is the first whole-school randomised controlled trial on the topic; the known long-term adverse results of bullying mandate follow-on work.

To a trauma treatment – post-traumatic stress disorder. Guidelines tend to favour exposure-based therapies over medication, although there are few head-to-head trials, and even less data on their combination. Rauch *et al* rectify this, with a three-arm study comparing sertraline, prolonged exposure therapy and their combination.² Noteworthy, the design meant that all getting therapy (which cannot be masked to participants) also received a pill (which can be masked): either sertraline (the combination group) or placebo. Interestingly, over the 24-week treatment, there were no differences between the two active conditions, and perhaps even more surprising, no gain from their combination. The results go against meta-analyses that typically show trauma-focused therapies to be superior, and hence current guideline recommendations. The authors address this point, noting how such pooled data normally reflect differences in effect sizes rather than direct comparison as undertaken here, and thus – they argue – those differences may be more reflective of differences in study design than effectiveness.

It is not clear why treatment combination – which typically offers added gains in depression and anxiety disorders – did not translate into greater improvements here. Of note, there was no control arm, and the sample was a veteran cohort with combat-related post-traumatic stress disorder. Astonishingly, this is the first direct comparison of two of the most commonly used interventions in this cohort.

Like gambling eels, phenotypes in psychiatry are slippery, probabilistic things. In recent years, we have heard about precision medicine, where features of an individual’s phenotype enable us to target treatment. Being optimists, recognising the burden of

mental ill health, and frustrated by the often-contradictory results from randomised control trial evidence in clinical practice, psychiatry really wants personalised medicine to work. But is our enthusiasm leading to (or derived from) poor statistical methods? To quote Senn ‘Statistics: a subject which most statisticians find difficult but in which nearly all physicians are experts’.³ In his recent piece in *Nature* he cautions us on how our mistakes might lead to overoptimism for precision medicine.

Senn articulates a core error, that on the surface, seems entirely plausible: if a subset of patients who received an active treatment improve, then this subset must be special (i.e. ‘responders’) compared with those who did not improve (‘non-responders’). The reason we make this error is because we do not seek sources of variation that might otherwise explain this. He singles out psychiatry, citing our fondness of defining response as ‘percentage change’ from baseline, using an overall scale, and failing to acknowledge that this might represent natural variation rather than a property of the ‘responding patient’. He takes aim at dichotomisation (‘dichotomania’) of response, where efficacy is artificially forced into a Bernoulli outcome of ‘remission’ versus ‘non-remission’ – often, he claims, we then need even more data than had we simply used the native continuous measure of the individual’s condition.

Further, with only a pre- and post-treatment measure, if someone has a bad day pre-treatment (higher symptom burden), but a better day post-treatment, then we cannot know that it is the intervention driving the change. Worse still, this natural variation could ‘switch’ them over (or under) the ‘responder’ line in a dichotomised outcome. Perhaps the most telling example is our failure to not ‘think counterfactually’: in a trial, we want to establish what happens when a patient is treated, versus the counterfactual, if they were not treated. Senn argues we often incorrectly use the baseline (pre-treatment) as our default counterfactual evidence because in the untreated group, we would expect no difference from baseline. However, regression to the mean and the ‘bad day’ argument could still yield differences, and then a change from the baseline measurement cannot be robust evidence of differences in response attributable to being treated (or not).

There is a way forward: always looking for the real sources of variation (rather than assuming heterogeneity in response) and to measure the patient’s response to the same intervention or control a number of times – for example repeated AB designs in *N-of-1* trials – and only then go looking for a group that appears to consistently respond when others do not.

‘There was a trend to significance with a *P*-value of 0.06’. Urgh. Usually, this occurs when someone wants you to believe that a hypothesis test should have supported rejecting the null (for example, if only we had more data). Goodman gives the example of a hurricane:⁴ if your weather forecast reported ‘Tallahassee will be hit, $P = 0.03$ ’, you would find that tough to interpret. What you want is a statement about the certainty of the claim or effect given the information available, for example, from previous studies, the quality of the data and the reported model. He suggests that we should report a confidence index that formally quantifies our judgement about a claim and its plausibility but notes that the familiar confidence interval is not the same; you can have two similar confidence intervals reported in two separate studies, but one might be trustworthy, the other less so. The confidence index adds to this interpretation and he suggests methods like sensitivity analyses (or Bayesian methods) in which we aim to show how robust the result is to variations and uncertainty in the data and the model. Then, we report our hypothesis, results and claims accompanied by a numerical summary of how we tested the model to establish confidence.

Maybe it is not size that counts, it is what you do with it that matters. Nevertheless, without wishing to get too Freudian, an enormous neocortex is our species' defining characteristic. From an evolutionary perspective this growth comes at a high metabolic cost – a quarter of our caloric intake. There are cogent arguments of facilitatory greater access to energy rich foods (meat) and complex social supports in our ancient past (*cf.* the next piece), but there is always a Darwinian pressure on maximising resources. Sneve *et al* discuss the beneficiary cognitive gains the 'massive and disproportionate expansion' of parts of the neocortex conferred upon us – supporting selective brain growth – through cross comparison of histological and neuroimaging data from several primate species, including *sapiens*.⁵ High-expanding cortical regions – notably the lateral temporal, parietal and prefrontal cortices – are characterised by significant internetwork connectivity and flexible recruitment across a range of different cognitive tasks. The authors propose that it is the sheer flexibility of these expanding cortical 'hotspots' to connect with and integrate multiple different brain regions depending on cognitive and emotional needs – supramodal cognition – that sustained and supported their evolutionary selection over the ages. From phylogeny to ontogeny, the emergence of this ability is also seen across the lifespan of individual humans, starting in childhood, levelling by early adulthood. Perhaps the opening aphorism could be modified to 'it is not *just* size that counts'...

'Man is the only animal that blushes. Or needs to' taught Mark Twain, and some emotions seem particularly human. Whither shame? Through our evolutionary past, hominins have had a very high dependency on others in the social group, far more so than most other species; mutual aid being a critical part of survival in an ancient world of disease, predation, food scarcity and high earlier-life mortality. Shame might be a mechanism to prevent disrupting these supports through avoiding socially damaging behaviour, but the cognitive processes behind it are complex: an individual needs to be able to predict what others will think of bad behaviour, and weigh that against any gains. To remove, or not to remove, the piece of cake in the office fridge that clearly has a 'property of Shirl' post-it on it – that is the question.

Earlier work had typically been on Western(ised) cohorts, and there were counterarguments about shame being largely a cultural construct (not everyone has post-its, appreciates Shirley's cake and so forth – you get the point). Sznycer *et al* tested this by carrying out an experiment on almost 900 participants across 15 small-scale and varied communities from around the world.⁶ Multiple scenarios were tested to elicit rated reactions about a hypothetical other, or about something they might have done, for example 'He/You steal(s) from members of his community'. The findings were absolutely consistent across languages, cultures and different modes of subsistence: the more a group devalued an act, the greater the intensity of the shame felt. Culture is not the driver, genes are.

The authors note that the response to another finding out reputation-damaging information about us is universal and ancient, with a stereotyped non-verbal show of subordination that displays it is acceptable to support us less, and subsequent compensatory over-cooperative behaviour to the wider group when they learn of what we have done. Twain was correct – it is uniquely part of being human and a deep part of our evolutionary biological heritage.

Finally, to lighten our load after discussing shame: people probably like you more than you think. You know that feeling of

anxiety anticipating meeting someone new, especially if you want things to go well? How, after you chat to them, you come away irritated by your clumsy conversation – the awkward gaps, the sense of talking too much, the cringe-worthy anecdote you regretted telling? We do. We think you do too: if you cannot remember, perhaps the phrases 'first date' or 'the interview panel will of course be in touch' might ring a few bells. Boothby *et al* propose this is a universal illusion, which they label the 'liking gap'.⁷ They observed new acquaintances among first-year college students, members of the public, and a laboratory team, and the findings were consistent: people systematically underestimate how much others like them and how much their company was actually enjoyed. It persisted across conversations of varying length, and could last months. We were delighted by the phrase in the paper that 'conversations are conspiracies of politeness in which people do not reveal their true feelings'. The authors have a positive message: it is not you, it is us – all of us – and people like you more than you think.

Dunne and colleagues ask how to manage this, specifically 'choking under pressure' as they tested participants during a task where they could win money for completing a difficult motor task.⁸ Choking tends to be proportional to the potential win, and their technique was straight forward reappraisal: instead of thinking 'do well and I win money', they changed thinking to consider that they lose money by doing badly. It is almost too simple to credit, but the results support the technique, and it was associated with significantly reduced choking in performance. Neuroimaging data showed that the switch also led to differential brain activity – it is getting the brain to work differently – and, fascinatingly, differences in sympathetic arousal. This last factor may explain the literal description of 'choking' as a real sensation from altered autonomic nervous system functioning in the face of pressure.

So, moving forward, you need to reframe pending stresses to think 'this person I am talking to will not get the life-changing opportunity to date/hire Amazing Me if I do not show them my stuff!' However, the jury is out as to whether this technique can be successfully applied to the English football team's penalty takers.

References

- 1 Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, et al. Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial. *Lancet* 2018; **392**: 2452–64.
- 2 Rauch SAM, Myra Kim H, Powell C, Tuerk PW, Simon NM, Acierno R, et al. Efficacy of prolonged exposure therapy, sertraline hydrochloride, and their combination among combat veterans with posttraumatic stress disorder. a randomized clinical trial. *JAMA Psychiatry* 5 Dec 2018 (doi:10.1001/jamapsychiatry.2018.3412).
- 3 Senn SS. Statistical pitfalls of personalized medicine. *Nature* 2018; **563**: 619–21.
- 4 Goodman SN. How sure are you of your result? Put a number on it. *Nature* 2018; **564**: 7.
- 5 Sneve MH, Grydeland H, Rosa MGP, Paus T, Chaplin T, Walhovd K, et al. High-expanding regions in primate cortical brain evolution support supramodal cognitive flexibility. *Cerebral Cortex* 24 Oct 2018 (doi: 10.1093/cercor/bhy288)
- 6 Sznycer D, Xygalatas D, Agey E, Alami S, An XF, Ananyeva KI, et al. Cross-cultural invariances in the architecture of shame. *PNAS* 2018; **115**: 9702–7.
- 7 Boothby EJ, Cooney G, Sandstrom GM, Clark MS. The liking gap in conversations: do people like us more than we think? *Psychol Sci* 2018; **29**: 1742–56.
- 8 Dunne S, Chib VS, Berleant J, O'Doherty JP. Reappraisal of incentives ameliorates choking under pressure and is correlated with changes in the neural representations of incentives. *Soc Cogn Affect Neurosci* 27 Nov 2018 (https://doi.org/10.1093/scan/nsy108).