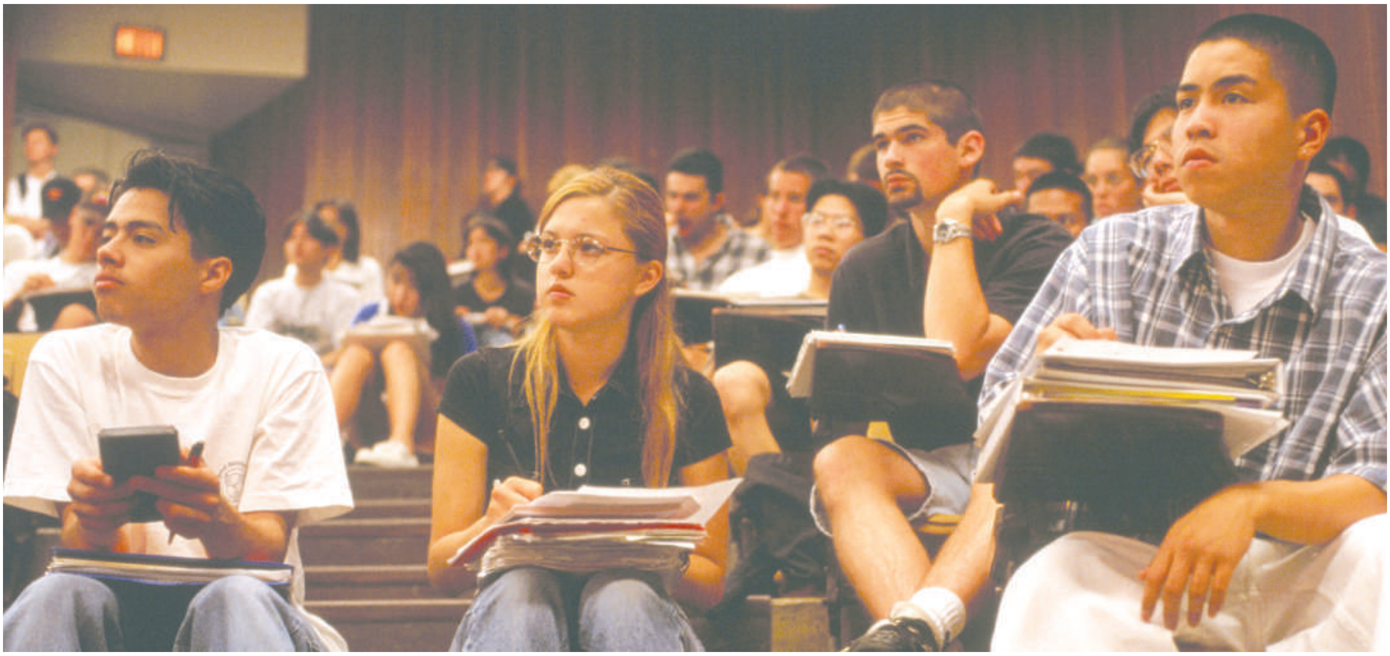
nature

# NEWS & VIEWS

D. BUTOW/CORBIS SABA



INTELLIGENCE

# A gender bender

Steve Blinkhorn

**The conclusion of a number-crunching exercise on various data sets is that male university students have significantly higher IQs than their female counterparts. But the methodology used is deeply flawed.**

"Finding bad reasons for what we believe on instinct" is how the philosopher F. H. Bradley famously defined metaphysics[1]. Meta-analysis — the deployment of statistical methods on a range of other people's research results to draw stronger conclusions than any individual study can bear — is in jeopardy of coming into that kind of disrepute when it is treated in the manner of a paper just published in the *British Journal of Psychology*[2].

The paper is by Paul Irwing and Richard Lynn, and was widely trailed in the British press at the end of August. It claims to show an average IQ advantage of 4.6 points in favour of males among university students — about one-third of the standard deviation of IQ in the population as a whole, which by convention is set at 15. If this were true, it would really matter, in a way that resonates with the recent controversy at Harvard University concerning the representation of women in science[3]. It would also overturn a consensus of more than 50 years' standing, that the only sex difference in IQ is a possible slightly greater variance among males.

The prospect of further controversy focused on gender politics and the shortfalls of IQ tests does not seem remote. But the real discussion should concern whether this paper stands up to scrutiny. It does not.

The authors chose to investigate sex differences in scores on two versions of Raven's matrices, the standard and the advanced. These are great survivors among intelligence tests, being still widely used in the same form around 60 years after they were developed. Their particular claims to fame arise from the fact that the content is entirely diagrammatic (and so, according to some, less culturally biased), and from the results of factor analyses that suggest the ability they tap is close to pure *g*, variously called general intelligence, fluid intelligence or the ability to learn.

Irwing and Lynn searched the literature for studies on university students that reported average scores by gender. They found 22, with sample sizes ranging from 30 to more than 9,000, publication dates from 1964 to 2004, and locations as various as Egypt, India, Mexico, Belgium, Australia and the United States — but curiously nothing from Britain where the tests were developed. For the most part, the purposes of the studies identified were not to investigate sex differences.

The number-crunching phase of the meta-analysis — weighting differences by sample size — resulted in an estimated difference of 0.15 standard deviations in favour of males. There are technical reasons for supposing that even this is an overestimate, but the authors go on to find reasons for not accepting this result. First, they exclude the largest study of all, an explicit 'norming' exercise in Mexico that accounts for almost 45% of the data, on the grounds that it is an 'outlier', with a difference of only 0.07 of a standard deviation — males answering on average only 0.4 more items correctly out of 60.

Abandoning the principle of weighting results by sample size, Irwing and Lynn then take the median of the estimated differences (0.31), and multiply it by the general-population standard deviation of IQ (15.0) to get the estimate of a 4.6-point IQ advantage for males. Before going on to speculate on the neurology underlying the female inferiority that they claim to have demonstrated, they propose that this result is "strong corroboration" of other meta-analytical results they have obtained for the general population.

It is nothing of the sort. The ten studies with estimated differences above the median cover a total of only 2,591 participants, whereas the ten studies with differences below the median account for 15,735 participants — the four largest differences come from samples of 111, 173, 124 and 300, the four smallest from samples of 844, 172, 9,048 and 1,316. Choosing to use the median is a flawed and suspect tactic.

But even going along with Irwing and Lynn's approach produces a different outcome if more plausible parameters are chosen. Whatever the standard deviation of IQ among university students, it is highly unlikely to be as large as 15, the conventional figure for the population as a whole. Raven's matrices, by design, tend not to produce normal (gaussian) distributions. An estimated standard deviation of 10 IQ points among students would ensure that there are not too many genuinely mentally defective undergraduates, even if some of them do have learning difficulties. This suggests a mean difference of up to around 1.5 IQ points — $10 \times 0.15$ (the figure weighted by sample size). Were a perfect study possible, with proper attention to sampling and motivational issues, I would expect even that to turn out to be an overestimate.

Raven's matrices have been widely used for decades, and it is likely that the 'file-drawer' effect is at work here — studies with no significant or substantial results never find their way into a journal, with no reporting of separate statistics for each gender when there is no difference. My own file drawer turned out to contain an analysis of data from 752 applicants for places on one degree course during the 1970s, tested on the advanced matrices, which were designed for the top 20% of the population. This yielded an advantage of 0.07 standard deviations for females. The sample is larger than all but five of those found by Irwing and Lynn. Should it be included in a meta-analysis? Of course not: the data were collected for another purpose and, without attention to purpose and context, meta-analysis is just numerical manipulation for its own sake.

Where there are sex differences to be found, detailed study of the internal workings of the test tends to show why. That's not based on instinct, but on my professional experience in designing gender-fair tests. Meta-analysis is not a substitute for properly designed research, and sex differences in average IQ, if they exist, are too small to be interesting. ■

Steve Blinkhorn is at Psychometric Research and Development Limited, Brewmaster House, The Maltings, St Albans, Hertfordshire AL1 3HT, UK.
e-mail: steve@prd.co.uk

1. Bradley, F. H. *Appearance and Reality: A Metaphysical Essay* (Swan Sonnenschein, London, 1893).
2. Irwing, P. & Lynn, R. *Br. J. Psychol.* **96,** 505–525 (2005).
3. Singer, E. *Nature* **434,** 424 (2005).

ASTRONOMY

# Light on a dark place

Christopher Reynolds

**The sharpest images ever taken of matter around the probable black hole at the centre of our Galaxy bring us within grasp of a crucial test of general relativity — a picture of the black hole's 'point of no return'.**

Ever since its discovery in 1974, a strange source of radio waves in the constellation of Sagittarius has been suspected of flagging the presence of a massive black hole at the centre of our Galaxy[1]. On page 62 of this issue, Shen *et al.*[2] report the highest-resolution images yet of this source, Sgr A*. These observations provide strong evidence that Sgr A* is indeed a black hole, and afford a glimpse of the behaviour of the matter that is about to flow into it. They are also a further step towards attaining an image of the shadow around the edge of a black hole, a powerful and classic test of the general theory of relativity.

Black holes are perhaps the most exotic objects to impinge on the cosmic consciousness. They are formed when matter such as that from a dying massive star collapses in calamitously under its own gravity, forming a region of space in which the gravitational field is so strong that it swallows all matter and radiation that come near it. Delineating this region is the event horizon: the point of no return, beyond which no matter or light can ever escape.

Sgr A* is certainly an exotic object. Through observations at infrared frequencies of the bright stars speeding around it, astronomers have confirmed that it is four million times more massive than our Sun and confined to a region of space at the exact centre of the Galaxy[3,4] that is no bigger than the region enclosed by the orbit of Pluto. Such a high concentration of mass puts tight constraints on the possible nature of the object. A cluster of several million neutron stars, themselves collapsed dead stars, could be as heavy as that, but could only survive in such a compact form for 20,000 years or so — a blink of an eye in astronomical terms — before either collapsing further (to a black hole) or evaporating away into space. It is unlikely that we are observing the galactic centre just when such a bizarre neutron-star cluster happens to exist. There is only one other possibility, however, if standard physics — the standard model of particle physics, coupled with the general theory of relativity — is to hold. That is that Sgr A* harbours a supermassive black hole.

Shen *et al.*[1] use a technique known as Very Long Baseline Interferometry (VLBI), which correlates information from radio antennae at separate, remote locations, thereby increasing the spatial resolution for images of far-off objects. The authors used the Very Long Baseline Array, a system of ten radio telescopes scattered across the United States with a maximum separation of some 8,000 kilometres. They built up a picture of the radio emission at a wavelength of 3.5 centimetres from gas in a region just 8 light minutes across, centred on the putative black hole. Even with the most conservative assumptions, the authors find lower limits on the concentration of mass that are a factor of 100,000 higher than those derived from the motions of the stars surrounding it. This would reduce the lifetime of any neutron-star cluster there to a mere 100 years, a result that must dispel any lingering notions that the source at Sgr A* is a compact cluster of known objects.

But we should guard against complacency: nature might have some surprises in store. Could it be that standard physics is inadequate and that, other than a black hole, there are stable objects that have the compact, huge mass of Sgr A*? What is needed is a more discerning test than simply detecting something massive and compact: we need to find the event horizon, the defining property of a black hole.

As physical phenomena go, event horizons are tricky to observe. In fact, the existence of an event horizon is almost invariably argued for based on the absence of some observational signature[5,6]. High-resolution imaging, however, does provide a compelling way to search for an event horizon. If a black hole is surrounded by an almost spherical distribution of radiating matter, as in the case of Sgr A*, a sufficiently high-resolution image should reveal a shadow around it (Fig. 1). This dark circle is caused by radiation from sources behind the black hole that are being swallowed by the event horizon[7]. Surrounding this shadow would be a bright ring — the result of the strong deflection by the black hole's gravitational field of those light rays that do scrape past it.

The predicted diameter of the event horizon's shadow for Sgr A* is just 30 microarcseconds[7], or 120 millionth of a degree. This would be the apparent size of a tennis ball on the Moon when viewed from Earth, and is about a factor of four smaller than the scales probed by current VLBI experiments. Two main directions of research should eventually allow us to see an image of the event horizon — if indeed there is one.

First, the resolving power of VLBI can be improved by reducing the wavelength at which