OPTICON MDCCCXXVI

Hardwicke, T E et al 2014 Only Human: Scientists, Systems, and Suspect Statistics. *Opticon1826,* (16): 25, pp. 1-12, DOI: http://dx.doi.org/10.5334/opt.ch

REVIEW

Only Human: Scientists, Systems, and Suspect Statistics

A review of: Improving Scientific Practice: Dealing With The Human Factors, University of Amsterdam, Amsterdam, September 11, 2014

Tom E. Hardwicke*, Leila Jameel*, Matthew Jones*, Eryk J. Walczak* and Lucia Magis-Weinberg*

It is becoming increasingly clear that science has sailed into troubled waters. Recent revelations about cases of serious research fraud and widespread 'questionable research practices' have initiated a period of critical self-reflection in the scientific community and there is growing concern that several common research practices fall far short of the principles of robust scientific inquiry. At a recent symposium, 'Improving Scientific Practice: Dealing with the Human Factors' held at The University of Amsterdam, the notion of the objective, infallible, and dispassionate scientist was firmly challenged. The symposium was guided by the acknowledgement that scientists are only human, and thus subject to the desires, needs, biases, and limitations inherent to the human condition. In this article, five post-graduate students from University College London describe the issues addressed at the symposium and evaluate proposed solutions to the scientific integrity crisis.

Introduction

The success of science is often attributed to its objectivity: surely science is an impartial, transparent, and dispassionate method for obtaining the truth? In fact, there is growing concern that several aspects of typical scientific practice conflict with these principles and that the integrity of the scientific enterprise has been deeply compromised. The diverse range of issues include cases of serious researcher fraud (e.g., Levelt-Noort-Drenth Committee, 2012; RIKEN Research Paper Investigative Committee, 2014), substantial publication bias towards positive findings (Rosenthal, 1979; Fanelli, 2012), a preponderance of statistically underpowered studies that produce inflated and/or unreliable effects (Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson, & Munafo, 2013), incomplete or erroneous reporting of study methodology (Carp, 2012; Vasilevsky, Brush, Paddock, Ponting, Tripathy et al., 2013), failure to comply with data access requests (Wicherts, Borsboom, Kats, & Molenaar, 2006), and the widespread prevalence of 'questionable research practices' that can insidiously generate false-positive findings

* Department of Experimental Psychology, University College London, UK
t.hardwicke.12@ucl.ac.uk, l.jameel@ucl.ac.uk, m.jones.12@ucl.ac.uk, eryk.walczak.11@ucl.ac.uk, lucia.magis-weinberg.12@ucl.ac.uk

Corresponding author: Tom E. Hardwicke

(Simmons, Nelson & Simonsohn, 2011; John, Loewenstein, & Prelec, 2012). In 2005, a devastating statistical proof was published, which claimed that 'most published research findings are false' (Ioannidis, 2005) and subsequent efforts to replicate existing findings have suggested that the suspected 'reproducibility crisis' is not just theoretically plausible; it is an empirical reality (Begley & Ellis, 2012; Prinz, Schlange, & Asadullah, 2011; Gilbert, 2014).

How is the scientific community to begin addressing these issues? For the organisers of a recent symposium, 'Improving Scientific Practice: Dealing with the Human Factors' hosted by The University of Amsterdam, the first step is to recognise that science is fundamentally a human endeavour, and thus subject to the limitations and biases that underlie human behaviour. Can we design a scientific ecosystem that acknowledges scientists are only human?

1. The damaged scientific ecosystem

The utopian idea of a 'pure' scientist is that of an individual motivated solely by the acquisition of knowledge. However, in reality scientists have human needs, desires, and motivations just like non-scientists (Mahoney, 1976). The scientific ecosystem which researchers inhabit is built and maintained by several organisations including universities, industry stakeholders, funding bodies, and publishers who also have interests that diverge from pure knowledge acquisition. Unfortunately, the present system does not adequately account for these human factors in science, and rewards individuals who are lucky or willing to 'play the game' (Bakker, van Dijk, & Wicherts, 2012). In this first section, we examine how the scientific ecosystem not only fails to guard against the inherent fallibilities of human behaviour, but actively perpetuates them.

1.1 Pressure to publish

Much of the scientific ecosystem revolves around a *de facto* principal commodity: the published research paper (Young, Ioannidis, & Al-Ubaydli, 2008). A metric called the *h*-index is sometimes used to evaluate scientists for hiring, promoting, and funding decisions (Hirsch, 2005). The *h*-index attempts to measure both productivity and research impact by combining number of publications and number of citations to these publications. However, citation rates are not necessarily indicative of quality or reliability: articles are also cited when they are critiqued, or when other researchers are unable to replicate the original finding. There was concern at the symposium that a single-minded drive for productivity is not conducive to the production of reliable research findings.

Part of the problem is that the emphasis on productivity adversely interacts with the personal career goals of individual scientists. For example, short-term contracts are common in academia and it has been suggested that, 'the research system may be exploiting the work of millions of young scientists for a number of years without being able to offer continuous, long-term stable investigative careers to the majority of them' (Ioannidis, Boyack, & Klavans, 2014). Consequently, there is a climate of fierce competition for increasingly limited funding (Anderson, Ronning, Vries, & Martinson, 2007b). This 'publish or perish' culture places inappropriate demands on a research process that should ideally be impartial and puts the integrity of the scientific enterprise in significant jeopardy (Fanelli, 2010).

1.2 Great expectations

Why has productivity become such an important factor in the scientific ecosystem? At the symposium, John Ioannidis, Professor of Medicine at Stanford University, argued that science's great success stories have led to unrealistic expectations. There is a pressure for research papers to be coherent, flawless narratives, but this only masks the scientific process in a veneer of perfection; the findings of most scientific studies are in reality highly nuanced (Giner-Sorolla, 2012). The problem is compounded by an increasing trend towards publication of short empirical reports. This mode of publication might facilitate the rapid dissemination of new findings, but it could also incur the cost of inflated false-positive rates, reduced integration of new studies with the existing literature, and promote an unhealthy focus on seeking eye-catching 'newsworthy' effects over rigorous theory-driven experimentation (Ledgerwood & Sherman, 2012). The constant call for new discoveries, life changing innovations, and a publishing system that strongly favours positive results, puts unreasonable pressures on scientists that may encourage or coerce them to engage in behaviours that benefit their careers, but are inconsistent with good scientific practice (Fanelli, 2010).

1.3 Widespread questionable research practices

On a spectrum of scientific behaviours ranging from intentional misconduct (e.g., falsification, fabrication, and plagiarism) to flawless research conduct, the remaining 'grey area' is populated by a variety of questionable research practices (QRPs). QRPs describe a range of activities that intentionally or unintentionally distort data in favour of a researcher's own hypotheses (John et al., 2012; Simmons et al., 2011). These include 'cherry picking': omitting outcomes, variables, or conditions that do not support the author's own beliefs (Chan, Hrobjartsson, Haahr, Gotzsche, Altman, 2004); 'HARKing': hypothesising after the results are known to give the more compelling impression that findings were predicted *a priori* (Kerr, 1998); and 'p-hacking': prematurely examining data and exploiting techniques that may artificially increase the likelihood of meeting the standard statistical significance criterion (typically α = .05), for example, making stop/continue data collection decisions (Armitage, McPherson, & Rowe, 1969), or engaging in post-hoc exclusion of outlier values (Bakker & Wicherts, 2014).

It seems clear that the damaged scientific ecosystem is partly to blame for the widespread engagement in QRPs (Fanelli, 2010; Bakker et al., 2012). However, they could also be an inevitable consequence of biases inherent to human cognition and thus difficult to overcome. For example, confirmation bias describes an effect whereby an individual preferentially seeks, interprets, and remembers information in a way that is consistent with their pre-existing beliefs (Nickerson, 1998). Although confirmation bias has only been sparsely investigated in scientific situations, some evidence indicates that scientists tend to discount evidence that might disconfirm their theoretical preferences (Brewer & Chin, 1994).

It is of great concern that QRPs are not just confined to a small subsection of the scientific community, but rather widespread and considered by many to be 'defensible' (John et al., 2012; Martinson, Anderson, de Vries, 2005). When used in a single study, these QRPs increase the likelihood of making a false-positive finding (Simmons et al., 2011). When employed on a large scale, such practices could have a devastating impact on the validity of the entire field of scientific inquiry (Ioannidis, 2005).

1.4 Unwillingness to share data

Even when it is accepted that false-positive findings are an inevitable by-product of the research process, much faith is placed in the notion of science as a 'self-correcting' enterprise (Merton, 1942). The idea is that spurious findings will eventually be exposed and purged whilst accurate findings will prevail. In order to facilitate self-correction, it is essential that scientists are open about their work so that it can be checked and repeated by their peers. Transparency is often considered to be a fundamental tenet of scientific investigation and many scientists subscribe to the norm of *communality*, which entails 'common ownership of scientific results and methods and the consequent imperative to share both freely' (Anderson, Ronning, De Vries & Martinson 2010; Merton, 1942). Unfortunately, at the symposium, Dr Jelte Wicherts (Tilburg University) depicted an aspect of the scientific ecosystem that contrasts vividly with this norm.

Data sharing is an important aspect of self-correcting science because it allows scientists to verify original analyses, conduct novel analyses, or carry out meta-analyses that can establish the reliability and magnitude of reported effects (Sieber, 1991; Tenopir, Allard, Douglass, Aydinoglu, Wu, Read, Manoff, & Frame, 2011). Wicherts described a 2006 paper in which attempts were made to access the data of 141 articles published in prominent psychology journals (Wicherts et al., 2006). Despite guidelines from the American Psychological Association (APA, 2001: 396) that compelled them to do so, 73% of authors did not share their data (for a similar finding in the biological sciences see Vines, Albert, Andrew, Débarre, Bock et al., 2014). Another concerning finding emerged when a subset of these papers was examined in greater detail: unwillingness to share data was associated with a higher prevalence of statistical reporting errors, particularly when those errors favoured an interpretation of the study's findings as statistically significant (Wicherts, Bakker, & Molenaar, 2011).

More generally, Wicherts and colleagues have found that statistical reporting errors are commonplace in the psychological literature (Bakker & Wicherts, 2011). Based on a reanalysis of 281 articles, the researchers estimated that around 18% of statistical results in the psychological literature are incorrectly reported. Similar to the findings outlined above, the majority of these errors support an interpretation of the study's results as statistically significant even though they were not. This is troubling as it suggests that researcher errors do not simply add noise to the research process, they introduce a systematic bias towards positive findings (Sarewitz, 2012). A field riddled with suspect statistics, QRPs, and a concomitant unwillingness to share data, is in danger of perpetuating falsehoods rather than establishing truths (Ioannidis, 2012).

1.5 Bad apples and a rotten barrel

At the symposium, Melissa Anderson, Professor of Higher Education at the University of Minnesota, argued that historically research governance has largely relied upon the self-regulation of scientists. This tendency has been motivated by faith in the scientific process to recruit individuals who are fit for the job and to weed out any 'funny business'. Underpinning this is a set of assumptions about the integrity and infallibility of scientists. Firstly, there is an implicit supposition that scientists are 'good people', motivated largely by the pursuit of knowledge. Scientists are also considered to be highly trained professionals who have undergone rigorous examinations and interviews. It is often assumed that rare cases of misconduct will be addressed by science's various mechanisms of self-correction: procedures such as peer-review, ethics committees, and study replication are all expected to filter 'bad science' from the system (Anderson et al., 2010; Merton, 1942).

Scientists are also subject to various legal and ethical protocols intended to promote research integrity. However, a recent examination of these protocols, presented during a symposium poster session, suggested that in Europe there is a complex system of overlapping regulatory bodies providing guidelines that vary considerably between countries and institutions (also see Godecharle, Nemery, & Dierickx, 2013). For example, there was considerable heterogeneity in the definition of 'misconduct' and the proposed mechanisms for dealing with it. It is hard to see how regulations characterised by such disunity and incoherence can provide effective oversight of integrity in the day-to-day workings of science.

It is also noteworthy that regulatory regimes are largely focused on dealing with researchers who engage in intentional misconduct. Anderson outlined how regulation is geared towards protecting scientific integrity from these 'bad apples'. However, she also highlighted the critical difference between 'misconduct' and 'misbehaviour'. According to US federal law, research misconduct is defined as fabrication, falsification, or plagiarism (Office of Science and Technology Policy, 2000). It must represent a 'significant' departure from 'proper practice' and be 'intentional'. Research misbehaviour on the other hand, comprises more ambiguous activities, such as the QRPs highlighted earlier in this article (see Section 1.3). Throughout the symposium there was a general consensus that the scientific establishment should not only be concerned with the 'bad apples' that propagate full-blown research misconduct, but apply greater focus to the 'rotten barrel' that leads scientists to (perhaps unwittingly) engage in research misbehaviour.

2. Rehabilitating the scientific ecosystem

Whilst the symposium began by outlining threats to scientific integrity and possible causes, a variety of solutions were also proposed, some with fairly broad aims and others targeting specific issues. Many of the speakers stated that no single solution would provide a panacea, and suggested that multiple initiatives would be required. Several speakers and delegates proposed that funding should be invested in an empirical examination of research practices and potential solutions in order to ensure their effectiveness. Perhaps it is time for scientists to turn their microscopes upon themselves and examine how their own behaviour, intentional or otherwise, distorts the scientific process? Other attendees of the symposium were keen to seize upon the current momentum for change and begin repairing the scientific ecosystem as soon as possible. In practice, many of the solutions outlined below are already being implemented, but in an incremental and voluntary fashion. In this section, we evaluate the solutions proposed at the symposium and examine the idea that, ultimately, rehabilitation of the scientific ecosystem will require considerable cultural change.

2.1 Changing incentives

Section 1.1 commented on how the scientific ecosystem's incentive structure is grossly misaligned with the principles of good science. At the symposium Professor Ioannidis proposed an ambitious scheme for appraising and rewarding research: a new metric that captures productivity, quality, reproducibility, shareability, and translatability (PQRST; Ioannidis & Khoury, 2014). The idea is to diversify the types of scientific activity that are rewarded in order to prevent productivity becoming scientists' principal goal. Practically speaking, it should be reasonably straightforward to estimate productivity using existing measures (for example, the proportion of registered clinical trials on ClinicalTrials.gov published two years after study completion), but the remaining parameters would require adding new features to scientific databases. For example, to calculate a 'shareability' index databases would need to monitor whether authors have uploaded their data to a public repository. Given the conflicting interests that influence the scientific ecosystem, it seems that reaching agreement on which quality standards are appropriate to use will be a more considerable barrier to change. Ioannidis hopes that realigning incentive structures with principles of good science will reduce the prevalence of scientific misbehaviours like QRPs and unwillingness to share data.

2.2 Scientific integrity training

Changing incentive structures may help to address intentional engagement in QRPs; however, it is also plausible that many QRPs are employed unwittingly simply because researchers are not fully aware of the extent to which these practices are problematic. The issue of integrity training was raised repeatedly at the symposium, but interestingly these proposals were largely directed at educating junior scientists. The poster on European research misconduct regulations, presented by Godecharle and colleagues, also reflected this: only Irish guidelines mentioned providing training to senior scientists (see Godecharle et al., 2013).

Some research suggests that the effectiveness of formal ethical training might be limited in comparison to the influence of lab culture or mentoring (e.g., Anderson, Horn, Risbey, Ronning, DeVries, & Martinson, 2007a). At the symposium, Anderson proposed that principal investigators should improve awareness of research integrity amongst junior researchers through lab-based discussions, and should seek to engage students by employing relevant real-life examples. For instance, mentoring sessions could utilise role-play in which researchers confront ambiguous research scenarios they might actually find themselves in. This would constitute a shift away from simply briefing students on the regulations and protocols that they are expected to follow as researchers. Instead it would concentrate on highlighting the difficulties of conducting research and show them how to solve problems in a realistic environment. However, we note that focusing training efforts solely on junior scientists may not be sufficient to address the present threats to scientific integrity whilst engagement in QRPs is widespread amongst senior scientists (John et al., 2012).

2.3 Preregistration of study protocols

Even if changing incentives and introducing training schemes are effective in improving scientific integrity, they may not be sufficient to eliminate the influence of QRPs that could arise as a consequence of biases inherent in human cognition (see Section 1.3). A potential solution to this problem, *preregistration*, was introduced to the symposium by Eric-Jan Wagenmakers, Professor of Cognitive Science at the University of Amsterdam. The central premise of preregistration is that researchers specify a methodology, sample size, and data analysis plan prior to conducting a study. This preregistration document can be uploaded to a public repository, such as The Open Science Framework (OSF) and referred to in any subsequent paper that reports the study. A stronger version of preregistration involves the submission of the preregistration document to a journal where, assuming the study is of satisfactory methodological quality, it will be accepted on the basis of the preregistration alone, and the journal would be committed to publishing the study *regardless of the results*.

The anticipated benefits of preregistration are two-fold. Primarily, it would prevent researchers from engaging in many QRPs because they are held to account by their own preregistration document. For example, it would be impossible for a researcher to engage in 'cherry picking', inappropriate post-hoc outlier exclusion, data 'peeking', or HARKing (see Section 1.3; John et al., 2012), when the relevant parameters have been specified prior to data collection. Furthermore, journal-based preregistration would help to address publication bias by ensuring that publication is dependent primarily upon methodological quality rather than the nature of the results (Chambers, 2013). This would help to reduce the 'file-drawer' problem (Rosenthal, 1979) whereby findings that do not achieve statistical significance are considerably less likely to be published—a state of affairs that drives the current publication bias towards positive findings (Fanelli, 2012) and undermines the validity of the academic literature (Ioannidis, 2005).

Several members of the symposium audience pointed out potential problems with preregistration. For example, it was suggested that there would be nothing stopping a scientist from engaging in QRPs and then 'preregistering' a study that they had in fact already completed. This is true, countered Wagenmakers, but in a preregistration scheme, such practices would clearly be fraud, and thus only likely to be committed by a small minority. A more practical criticism was that preregistration could increase workload because it involves two stages of peer review: prior to data collection to evaluate methodology and after data collection to evaluate adherence to the preregistration plan. However, Chambers et al. (2014) argue that journal-based preregistration could in fact save time. In the current publication system it is common for a manuscript to be submitted and reviewed at multiple journals, often being rejected several times based on either methodological problems or because the results are not deemed 'interesting'. However, in a preregistration scheme, studies are primarily judged on their methodological quality, which is established prior to the study being run. Thus, a more thorough reviewer-author interaction at the pre-data collection stage will ultimately reduce the likelihood that the research has to undergo several rounds of submission and review at multiple journals. A system that scrutinises research protocols and methods prior to commencing data collection could also be helpful for authors. Under the current system, irreparable methodological issues may only come to light when authors have already invested time and money in running the study. Whereas, in the new system authors would receive feedback about their proposals before commencing the study, allowing for improvements to be made. Overall then, the time-cost for authors, reviewers and editors could be negligible or even an improvement compared to the present system.

Other delegates objected on the grounds that preregistration may shackle science by outlawing creative post-hoc explorations of data or restricting observational research (see also Scott, 2013). But Wagenmakers argued that preregistered studies could still include post-hoc exploratory analyses that the authors and reviewers believe to be appropriate. By using preregistration, a clear distinction would be made between confirmatory analyses specified in the preregistration, and exploratory analyses inspired by the data (see Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Readers could then treat author claims with the appropriate degree of skepticism depending on the status of their analysis. Furthermore, fraudulent preregistration could backfire, as editors are likely to require revisions to the proposed protocol (Chambers, Feredoes, Muthukumaraswamy, & Etchells, 2014). Thus, even relatively minor changes to the experimental procedure would be impossible if the study had already been completed.

2.4 Transparency through data sharing

Preregistration may address integrity issues prior to and during data collection, but the studies described earlier by Jelte Wicherts and colleagues suggest a widespread unwillingness to share data with fellow scientists *after* findings have been published (Wicherts et al., 2006). Wicherts believes that his work describes a culture of secrecy in which misconduct can flourish, and he has built a strong case for obligatory data sharing in the scientific community (Wicherts, 2011).

However, there are practical and ethical issues to overcome. Martin Bobrow (2013) for example, agrees that there is, 'an ethical imperative…to maximize the value of research data' but also acknowledges the need to be cautious, as there is a risk of individuals being identified in sensitive datasets. Bobrow suggests that as more research is shared it is important to assess how these data are being used, to examine the risks, and to devise appropriate governance that balances privacy with public benefit.

In the neuroimaging community, concerns have been raised about the various technical issues involved in sharing large and complex brain imaging data (Nature Neuroscience Editorial, 2000). A more general issue that has arisen from this debate is that many researchers fear being 'scooped' if discoveries are made using their dataset before they have been able to finish analysing the data themselves. These concerns appear to be an unfortunate consequence of a scientific ecosystem that incentivises productivity in terms of publications and fails to account for other activities that contribute to credible scientific inquiry (see Section 1). The PQRST metric proposed by Ioannidis and Khoury (2014; see Section 2.1) explicitly incorporates 'shareability' as an index of scientific quality.

Generally speaking, professional guidelines, for example those provided by the American Psychological Association, do not appear to offer sufficient compulsion for authors to share their data. At present, data sharing policies vary substantially across journals (Alsheikh-Ali, Qureshi, Al-Mallah, & Ioannidis, 2011) and Wicherts recommends that journals require from authors to upload their data to a public repository (e.g., The OSF) along with a relevant codebook so that other researchers can navigate the dataset. Although this may generate additional work for the original author in terms of preparing the dataset for other users, Wicherts argues that data sharing in this manner is an essential component of transparent scientific practice.

2.5 Cultural change

Some of the solutions outlined above have either been met with resistance, or at least not fully embraced by the scientific community (e.g., Scott, 2013). There is some evidence suggesting that scientists are generally open to change, but wary of new schemes and regulations that might impose rigidity on their practice (Fuchs, Jenny & Fiedler, 2012). A more comprehensive solution to the current problems faced by science would comprise a wholesale rehabilitation of scientific culture, in tandem with some of the more practical initiatives proposed above.

Individual scientists rarely work in isolation, typically operating in teams, situated within departments and institutions, and interacting with colleagues in their disciplinary field through publications, attendance at conferences, and informal communications both public (e.g., social media) and private (e.g., e-mail). These different communities each have a cultural identity and establish proximal norms that influence the behaviour of community members. In order for any of the previously proposed procedures or regulations to be effective, the culture of science may need to shift so that individuals are supported by their colleagues to make the right decisions.

In a culture where scientists have to 'play the game' to survive (Bakker et al., 2012), it is hard for an individual scientist to prioritise the integrity of their research. Martinson et al. (2005) found a significant association between self-reported scientific misbehaviour and perceived inequities in the funding allocation process. These findings suggest that when people feel 'wronged' or are working in a climate they believe to be rife with competition and power games, they are more likely to prioritise the success of their own careers over behaviours that support credible scientific inquiry. Anderson also described a study (unpublished data) in which 7,000 mid-career and early-career researchers were asked whether they had ever engaged in either research misconduct or misbehaviour. A very modest number reported misconduct, but many reported misbehaviour. Researchers were also asked to report what they thought about other researchers' engagement in these practices. Interestingly, a positive correlation was identified between those who self-reported increased levels of research misconduct or misbehaviour, and the extent to which they perceived others were engaged in such practices. This depicts a scientific ecosystem in which individuals are more likely to engage in misconduct and misbehaviour if they think others around them are too.

At the symposium, Anderson proposed a number of initiatives that sought to challenge the current scientific culture. There is some evidence to suggest that signing an institutional research integrity oath or honour code, and receiving reminders of these agreements, could reduce research misbehaviour. In an experiment with students at MIT and Yale, Mazar, Amir and Ariely (2008) found that simply printing the statement 'I understand that this…falls under [MIT's/Yale's] honor system' on test papers significantly reduced cheating regardless of the incentive offered and despite no real honor code existing at these institutions. Whilst this is a promising finding, the idea remains to be investigated in research settings involving real research misconduct or misbehaviour, where the stakes are higher, and the factors influencing engagement in QRPs are diverse. Perhaps journal submission portals or PhD vivas could require researchers to sign a research integrity code when submitting a manuscript or thesis? It would also be important to consider how an honor code could be applied to complex 'grey area' behaviours, since the usual mechanisms are clearly insufficient for regulating research misbehaviour.

3. Conclusion

Whilst the issues faced by the scientific disciplines are alarming, it is exciting to be part of a community that is reflecting critically on an unsustainable status quo. Many of the current issues have been raised previously but change has not been forthcoming. The main differences this time are an increased awareness about these issues within the scientific community and widespread access to technological apparatus that can support inventive and accessible solutions.

However these are also unsettling times for young researchers finding their feet in a scientific system that appears to have drifted far from its principal goal of truth-seeking. In his book *Advice For A Young Investigator*, the neuroscientist Ramón y Cajal suggests that 'two emotions must be unusually strong in the great scientific scholar: a devotion to truth and a passion for reputation' (Ramón y Cajal,1897/1999: 40). Yet in a scientific ecosystem that rewards researchers for their productivity more than for their methodological rigor, a young investigator who is fully devoted to the truth cannot afford to be passionate about their reputation, and a young investigator passionate about their reputation cannot afford to be fully devoted to the truth. It is time to rehabilitate the scientific ecosystem, and the first step is to acknowledge that scientists are only human.

Acknowledgements

The authors would like to thank the Research Department of Experimental Psychology, University College London, for the provision of travel subsidies that enabled attendance at this event. We are grateful to David Shanks and Kari Vyas for their comments on an earlier draft of this article. We would also like to thank the organisers of the symposium addressed in this paper for arranging such an enjoyable and thought-provoking array of speakers on the topic 'Human Factors in Science'.

References

Alsheikh-Ali, A A, Qureshi, W, Al-Mallah, M H, and **Ioannidis, J P A** 2011 Public availability of published research data in high-impact journals. *PLoS ONE*, 6(9): e24357. DOI: http://dx.doi.org/10.1371/journal.pone.0024357

American Psychological Association 2001 Ethical standards for the reporting and publishing of scientific information. In *Publication Manual of the American Psychological Association* (5th ed., pp. 387–396). Washington, DC.

Anderson, M S, Horn, A S, Risbey, K R, Ronning, E A, De Vries, R, and **Martinson, B C** 2007a What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a national survey of NIH-funded scientists. *Academic Medicine: Journal of the Association of American Medical Colleges,* 82(9), 853–60. DOI: http://dx.doi.org/10.1097/ACM.0b013e31812f764c

Anderson, M S, Ronning, E A, De Vries, R, and **Martinson, B C** 2007b The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics*, 13: 437–461. DOI: http://dx.doi.org/10.1007/s11948-007-9042-5

Anderson, M S, Ronning, E A, De Vries, R, and **Martinson, B C** 2010 Extending the Mertonian norms: Scientists' subscription to norms of research. *The Journal of Higher Education,* 81(3): 366-393. DOI: http://dx.doi.org/10.1353/jhe.0.0095

Armitage, P, McPherson, C K, and **Rowe, B C** 1969 Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society Series A (General)*, 132(2): 235–244. DOI: http://dx.doi.org/10.2307/2343787

Bakker, M, van Dijk, A, and **Wicherts, J M** 2012 The rules of the game called psychological science. *Perspectives on Psycho-*

logical Science, 7: 543–554. DOI: http://dx.doi.org/10.1177/1745691612459060

Begley, C G, and **Ellis, L M** 2012 Drug development: Raise standards for preclinical cancer research. *Nature* 483(7391): 531–533. DOI: http://dx.doi.org/10.1038/483531a

Bobrow, M 2013 Balancing privacy with public benefit, *Nature,* 500 (7461): 123. DOI: http://dx.doi.org/10.1038/500123a

Brewer, W F, and **Chinn, C A** 1994 Scientists' responses to anomalous data: Evidence from psychology, history, and philosophy of science. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1: 304–313.

Button, K S, Ioannidis, J P A, Mokrysz, C, Nosek, B A, Flint, J, Robinson, E S J, and **Munafo, M R** 2013 Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14: 365–376. DOI: http://dx.doi.org/10.1038/nrn3475

Carp, J 2012 The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage*, 63: 289–300. DOI: http://dx.doi.org/10.1016/j.neuroimage.2012.07.004

Chambers, C D 2013 Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49(3), 609–610. DOI: http://dx.doi.org/10.1016/j.cortex.2012.12.016

Chambers, C D, Feredoes, E, Muthukumaraswamy, S D, and **Etchells, P J** 2014 Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, 1(1) 4–17. DOI: http://dx.doi.org/10.3934/Neuroscience2014.1.4

Chan, A W, Hrobjartsson, A, Haahr, M T, Gotzsche, P C, and **Altman, D G** 2004 Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA*, 291: 2457–2465. DOI: http://dx.doi.org/10.1001/jama.291.20.2457.

[Editorial] 2000 A debate over fMRI data sharing. *Nature Neuroscience*, 3(9): 845-846. DOI: http://dx.doi.org/10.1038/78728

Fanelli, D 2010 Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE*, 5: e10271. DOI: http://dx.doi.org/10.1371/journal.pone.0010271

Fanelli, D 2012 Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90, 891–904. DOI: http://dx.doi.org/10.1007/s11192-011-0494-7

Fuchs, H M, Jenny, M, and **Fiedler, S** 2012 Psychologists are open to change, yet wary of rules. *Perspectives on Psychological Science*, 7(6): 639–642. DOI: http://dx.doi.org/10.1177/1745691612459521

Gilbert, E A 2014 Reproducibility project: Psychology (preliminary) results. Retrieved from: https://osf.io/vtksf/

Giner-Sorolla, R 2012 Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, 7(6), 562–571. DOI: http://dx.doi.org/10.1177/1745691612457576

Godecharle, S, Nemery, B, and **Dierickx, K** 2013 Guidance on research integrity: no union in Europe. *The Lancet*, 381(9872): 1097–1098. DOI: http://dx.doi.org/10.1016/S0140-6736(13)60759-X

Hirsch, J E 2005 An index to quantify an individual's scientific research output. *PNAS*, 102: 16569–16572. DOI: http://dx.doi.org/10.1073/pnas.0507655102

Ioannidis, J P A 2005 Why most published research findings are false. *Perspectives in Psychological Science,* 7(6): 645–654. DOI: http://dx.doi.org/10.1371/journal.pmed.0020124

Ioannidis, J P A 2012 Why science is not necessarily self-correcting. *PLoS Medicine* 2(8): e124. DOI: http://dx.doi.org/10.1177/1745691612464056

Ioannidis, J P A, Boyack, K W, and **Klavans, R** 2014 Estimates of the continuously publishing core in the scientific workforce. *PloS*

One, 9(7), e101698. DOI: http://dx.doi.org/10.1371/journal.pone.0101698

Ioannidis, J P A, and **Khoury, M J** 2014 Assessing value in biomedical research: the PQRST of appraisal and reward. *JAMA*, 312(5), 483–4. DOI: http://dx.doi.org/10.1001/jama.2014.6932

John, L, Loewenstein, G, and **Prelec, D** 2012 Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science* 23(5), 524–32. DOI: http://dx.doi.org/10.1177/0956797611430953

Kerr, N L 1998 HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3): 196–217. DOI: http://dx.doi.org/10.1207/s15327957pspr0203_4

Ledgerwood, A, and **Sherman, J W** 2012 Short, sweet, and problematic? The rise of the short report in psychological science. *Perspectives on Psychological Science*, 7(1): 60–66. DOI: http://dx.doi.org/10.1177/1745691611427304

Levelt Committee, Noort Committee, and Drenth Committee 2012 Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. https://www.commissielevelt.nl/wp-content/uploads_per_blog/commissielevelt/2013/01/finalreportLevelt1.pdf

Mahoney, M J 1976 *Scientist as subject: The psychological imperative.* Cambridge, MA: Ballinger Publishing Company.

Martinson, B C, Anderson, M S, and **de Vries, R** 2005 Scientists behaving badly. *Nature,* 435(7034): 737–738. DOI: http://dx.doi.org/10.1038/435737a

Mazar, N, Amir, O, and **Ariely, D** 2008 The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research,* 45(6): 633–644. DOI: http://dx.doi.org/10.1509/jmkr.45.6.633

Merton, R K 1942 The Normative Structure of Science. In: Storer, N W 1973 *The Sociology of Science*. Chicago: University of Chicago Press. pp 267–278.

Nickerson, R S 1998 Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2: 175–220. DOI: http://dx.doi.org/10.1037/1089-2680.2.2.175

Office of Science and Technology Policy 2000 Federal Research Misconduct Policy. *Federal Register*, 65(235): 76260–76264.

Prinz, F, Schlange, T, and **Asadullah, K** 2011 Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery* 10: 712. DOI: http://dx.doi.org/10.1038/nrd3439-c1

Ramón y Cajal, S 1897/1999 Advice for a young investigator. Translated by Swanson N & Swanson L W. Cambridge, MA: MIT Press.

RIKEN Research Paper Investigative Committee 2014 Report on STAP Cell Research Paper Investigation. Retrieved from: http://www3.riken.jp/stap/e/f1document1.pdf

Rosenthal, R 1979 The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86(3): 638–641. DOI: http://dx.doi.org/10.1037/0033-2909.86.3.638

Sarewitz, D 2012 Beware the creeping cracks of bias. *Nature*, 485: 149. DOI: http://dx.doi.org/10.1038/485149a

Scott, S 2013 Pre-registration would put science in chains. *Times Higher Education.* Retrieved from: http://www.timeshighereducation.co.uk/comment/opinion/science-in-chains/2005954.article

Sieber, J E 1991 Openness in the social sciences: Sharing data. *Ethics & Behavior*, 1: 69–86, DOI: http://dx.doi.org/10.1207/s15327019eb0102_1

Simmons, J P, Nelson, L D, and **Simonsohn, U** 2011 False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11): 1359–1366. DOI: http://dx.doi.org/10.1177/0956797611417632

Tenopir, C, Allard, S, Douglass, K, Aydinoglu, A U, Wu, L, Read, E, Manoff, M, and **Frame, M** 2011 Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6: e21101. DOI: http://dx.doi.org/10.1371/journal.pone.0021101

Vasilevsky, N A, Brush, M H, Paddock, H, Ponting, L, Tripathy, S J, LaRocca, G M, and **Haendel, M A** 2013 On the reproducibility of science: unique identification of research resources in the biomedical literature. *PeerJ* 1: e148. DOI: http://dx.doi.org/10.7717/peerj.148

Vines, T H, Albert, A Y K, Andrew, R L, Debarre, F, Bock, D G, Franklin, M T, Gilbert, K J, Moore, J S, Renaut, S, and **Rennison, D J** 2014 The availability of research data declines rapidly with article age. *Current Biology*, 24: 94–97. DOI: http://dx.doi.org/10.1016/j.cub.2013.11.014

Wicherts, J M 2011 Psychology must learn a lesson from fraud case. *Nature*, 480: 7. DOI: http://dx.doi.org/10.1038/480007a

Wicherts, J M, Bakker, M, and **Molenaar, D** 2011 Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. *PLoS ONE*, 6(11). DOI: http://dx.doi.org/10.1371/journal.pone.0026828

Wicherts, J M, Borsboom, D, Kats, J, and **Molenaar, D** 2006 The poor availability of psychological research data for reanalysis. American Psychologist, 61(7): 726–728. DOI: http://dx.doi.org/10.1037/0003-066X.61.7.726

Young, N S, Ioannidis, J P A, and **Al-Ubaydli, O** 2008 Why current publication practices may distort science. *PloS Medicine*, 5: e201. DOI: http://dx.doi.org/10.1371/journal.pmed.0050201

How to cite this article: Hardwicke, T E, Jameel, L, Jones, M, Walczak, E J and Magis-Weinberg, L 2014 Only Human: Scientists, Systems, and Suspect Statistics. *Opticon1826,* (16): 25, pp. 1-12, DOI: http://dx.doi.org/10.5334/opt.ch

Published: 22 December 2014