

The Problem of Fraud in Parapsychology

“Oh, what a tangled web we weave, when first we practise to deceive.”

Walter Scott,
Marmion (1808)

In this article I would like to take how parapsychology is portrayed in mainstream academic textbooks as characterised by experimenter fraud, with the implication that many positive outcomes can be accounted for in terms of malpractice. For example, Gross (2010, p. 85) quotes Colman (1987) as describing the history of parapsychology as “disfigured by numerous cases of fraud involving some of the most ‘highly respected scientists’”. But such damning appraisals are not restricted to dyed-in-the-wool sceptics. Douglas Stokes (2015) has claimed that the body of evidence from parapsychological research



| by CHRIS A. ROE
University of Northampton

“conform[s] to the pattern that would be expected if a small minority of psi researchers has engaged in fraud” (p. 42), and for James Kennedy (2014) the situation is even worse: “Experimenter misconduct has occurred many times in parapsychology and is a constant threat. It detracts from the scientific acceptance of the field and hinders progress by divert-

ing resources to invalid hypotheses” (p. 9). In this article I’d like to take a closer look at how commonplace fraud has been in parapsychology and how this compares with other scientific disciplines so as to make a judgement about whether such claims have any substance.

Instances of Fraud in Parapsychology

To substantiate his assertion, Kennedy refers ominously but obtusely to 17 cases of fraud in parapsychology of which 15 are derived from Rhine’s (1974a) paper, “Security Versus Deception in Parapsychology.” In that article Rhine does indeed refer to “a dozen cases to illustrate fairly typically the problem of experimenter unreliability prevalent in the 1940s and 1950s” (i.e., in the wake of the popularisation of Rhine’s monograph, *Extrasens-*

I would like to take a closer look at how commonplace fraud has been in parapsychology and how this compares with other scientific disciplines so as to make a judgement about whether such claims have any substance.

sory Perception, which described research methods that had been developed at Duke University so as to allow replication attempts without the need for specialist equipment or extensive training in research methods). It seems as if the monograph's aim was successful, since these cases sound like reports of experiments received by the *Journal of Parapsychology* from people who were not academics ("seven did not have the doctorate") and not part of the parapsychological community ("several were persons of evident ability but were located (some of them abroad) where research in parapsychology was extremely hard to manage but not nearly so hard to fake"). There is no indication that any of these persons continued to conduct experiments nor that any of their work was published. It seems odd, then, for Kennedy to portray them as if they were typical or representative of the parapsychological community at large. Rhine does refer to four additional cases involving people who were "all better qualified for

psi research than ... the 'dozen'. They all knew the rules and standards that had been developed through the years" (p. 105), and these are of more concern. Unfortunately, all cases are described in general terms, with no information included that might allow those suspected of "experimenter unreliability" to be identified. Their non-adherence to generally accepted security standards is described, though it is not altogether clear that they constitute misconduct. For instance, the first example involves a comparison of performance by participants at psi and non-psi tasks in which two experimenters were responsible for different tests. Rather than ensure that the tests were scored independently while masked to the outcomes from other tests, the experimenters actually exchanged information when participants were scoring particularly well. This could certainly result in an expectancy bias when scoring performance that could inflate any correlation in scores between the tasks, but pales in comparison with the types of fraud discussed later in this article. Examples 2 and 3 present a stronger circumstantial case for experimenter fraud, in which effects only occurred when one experimenter had unsupervised access to raw data records. Importantly, colleagues raised suspicions about the work and none of it had been published when the researchers

withdrew. The final example Rhine includes is a study that adopts the Screened Touch Matching method described by Pratt and Woodruff on the grounds that it had been criticised by Mark Hansel as allowing for fraud by exploiting inadequate matching, and so is included because "trickery was a conceivable possibility" rather than because someone had been caught *in flagrante delicto*. Taken together, these 16 cases (accepting that Kennedy includes only the first 15) do not make a compelling case for Kennedy's assertion that experimenter misconduct is commonplace in parapsychology. He also raises concerns about other researchers in Rhine's laboratory, but these amount to little more than pernicious insinuation of the tittle-tattle variety, and I will not spend time on them here.

Of more concern are two particular cases to which Kennedy (2014) and Stokes (2015) refer that are generally accepted as involving calculated and systematic fraud by the experimenter. The first of these involves Walter J. Levy Jr., a medical school graduate who had joined J. B. Rhine's Institute for Parapsychology and had been so prolific and highly regarded that he had been appointed Director of the Institute; he was widely expected to succeed Rhine on the latter's imminent retirement (Stokes, 2015). Levy was particularly interested in animal psi and had designed

ingenious experiments that tested psi abilities in gerbils, rats, and chick embryos such that successful psi performance would meet the animals' basic needs (avoidance of pain, increase in experiences of pleasure, maintenance of optimal body temperature). One of the great strengths of Levy's experiments was their automaticity — once set up, the apparatus could run independently, monitoring the behaviour of random event generators (REGs) that the animals needed to influence in order to produce the desired outcome, and creating a physical recording of the outcomes for analysis. Three of Levy's research colleagues (Kennedy, Jim Davis, and Jerry Levin) became suspicious, then, of the amount of time Levy seemed to spend in the vicinity of the equipment while the experiments were in progress. They secretly wired up the computer so that it would produce a second record of REG output, and to their consternation discovered that the REG output was perfectly random while Levy's official record showed a deviation in the predicted direction (i.e., the rats were getting more stimulation of their neurological pleasure centres than would be expected by chance). Rhine was presented with the evidence and confronted Levy, who admitted having falsified the confirmatory study but insisted that his original research was sound and the

Levy was particularly interested in animal psi and had designed ingenious experiments that tested psi abilities in gerbils, rats, and chick embryos such that successful psi performance would meet the animals' basic needs [...]

data falsification had only begun when the genuine effect could not be repeated (Kennedy, 1975). This line of research had not yet been published (because of the need, in Rhine's view, for replication in order to confirm findings as evidential). Levy defended his other, published, work, pointing out that it had been independently replicated both at the Institute for Parapsychology and elsewhere. Nevertheless, in writing about the affair in the next issue of the *Journal of Parapsychology*, Rhine (1974b) sagely advised "although his single known violation involves only one of his many experimental lines, it unavoidably casts a reasonable doubt on all of his work individually and jointly conducted during the five years he has been in parapsychology" (p. 220). Rhine took comfort from the principle that, in the long run, independent replication would differentiate between sound and unsound findings. He immediately wrote to all those he believed were plan-

ning to refer to Levy's work in designing and writing up their own studies to inform them that the data were suspect, ensuring that this instance of fraud was dealt with swiftly and publicly. Fifteen months later Rhine (1975) gave an update which identified other suspicious practices by Levy and led him to conclude rather reluctantly that, despite double masked and multi-experimenter designs in some cases, no study completely eliminated the possibility of dishonesty, and so there was no option but to write it all off.

The second generally accepted case involves Samuel G. Soal, a mathematics lecturer at Queen Mary College (part of the University of London), who had been one of the principal exponents of forced choice ESP testing in the UK, but the consistent failure of those tests had led him to become among Rhine's severest critics (Beloff, 1993). That is, until (on the advice of Whately Carington) he reanalysed his data to look for displacement effects, instances in which the participant's call corresponded not with the target symbol but with the preceding or subsequent symbol. Two participants who showed evidence of these effects, Basil Shackleton and Gloria Stewart, were invited back for further tests, which resulted in a steady stream of significant results. Soal was a mathematician by profession but achieved his D.Sc. for psychical research. Al-

though his reputation remained intact during his lifetime, suspicions had been raised but these were quickly put down by threats of legal action (Beloff, 1993, p 147). In a detailed account, Markwick and West (unpub.) have described how one of the accusers was Mrs. Gretl Albert, one of the successful agents for Basil Shackleton. She confided to her friend Mollie Goldney (Soal's co-experimenter) that after the test she had seen him altering figures on a score sheet (the records were in the form of the numbers 1 to 5 rather than as symbols). Goldney asked to see the score sheets, but could find no signs of alteration. The nature of the alterations was not obvious from simple scrutiny with a magnifying glass; it required statistical analysis. Scott and Haskell (1974) speculated that if Soal was converting target 1s into 4s or 5s so that they matched participant calls, then there should be an excess of hits on trials where the target was apparently a 4 and 5 and a deficit of target 1s where the guess was 4 or 5. Both effects were confirmed so dramatically as to suggest that most target 1s that fell opposite a guess 4 or 5 had been altered to produce false hits. However, there was no overall deficit of target 1s or excess of target 4s and 5s, which Barrington (personal communication, cited in Markwick & West, unpub.) interpreted as suggesting that the target sheets had been prepared with

manipulation in mind by beginning with too many 1s and too few 4s or 5s. Scott and Haskell acknowledged that their findings could not explain the above chance scoring in most of Soal's sessions with Shackleton, but felt it was more likely that other falsification methods had been used than that results were a mixture of the genuine and the fraudulent. Evidence of those other methods came from Markwick (1978), who found that Soal's pre-prepared lists of random numbers (supposedly drawn from published sources) contained repetitions of sequences of up to 25 digits at a time, sometimes in reverse order. These repeated sequences contained an occasional extra digit, and 75% of these gave hits, which could suggest that placeholder digits (e.g., 1) had been entered with the intention of adjusting them later to match the call. Despite the general suspicions surrounding Soal's reported levels of success (according to Stokes, 2015, J. B. Rhine had long suspected that his research was fraudulent), other aspects of his behaviour are puzzling. Markwick and West (unpub., p. 176) conclude, "our revelation of his further deceptions may boost some sceptics' assumptions that all seemingly convincing claims for the paranormal must be fraudulent. That is not our view. Soal was exceptional in his secretiveness, his resistance to outside interference and his unwillingness

to have his subjects tested by other experimenters."

Cases of Fraud in Other Disciplines

So how does this portrait of fraud in parapsychology compare with other sciences? Rhine himself felt that parapsychology lagged behind: "most other branches of science have already matured to the point where the problem of experimenter trickery causes no great concern. This is partly because deliberate fraud would be too quickly spotted and exposed" (p. 112). This confidence in the checks and balances of mainstream scientific practice has been quite typical. Braud and Wade (1982, p. 11) reproduce testimony by Philip Handler, President of the National Academy of Sciences, before the US House Committee on Science and technology who described the problem of scientific fraud as "grossly exaggerated" by the press, and even when it does transpire "occurs within a system that operates in an effective democratic and self correcting mode". But that system has proven to be imperfect and the collection of cases of scientific misconduct has grown steadily, so much so that the National Science Foundation differentiates between three types: fabrication, falsification, and plagiarism (cf. Gross, 2016, p. 694). Plagiarism is the appropriation of another person's

ideas, processes, results, or words without giving appropriate credit. Falsification is manipulating research materials, equipment, or processes, or changing data or results. Fabrication is making up data or results. To illustrate these three types and to evidence their pervasive occurrence I shall briefly refer to major figures in the history of science who are now generally accepted as having engaged in fraud. This listing is far from exhaustive (for further examples, see Braud & Wade, 1982; Grant, 2007; Judson, 2004; Kohn, 1986).

Perhaps the earliest established case of plagiarism is that of Claudius Ptolemy, whose *Almagest* presents astronomical observations that provide the basis for a mathematical model to describe the movements of celestial bodies around the Earth, and was hugely influential until superseded by the heliocentric model proposed by Copernicus and others. Ptolemy claimed to have made these observations himself in Alexandria, Egypt. However, later back-calculations from the planets' current positions suggested that many of these observations were very poor even by the standard of the day and accorded much more closely with what Ptolemy's predecessor Hipparchus could have observed from Rhodes some 278 years earlier (Newton, 1977). Ptolemy's observations include curious omissions —of the 1,025 stars he documented, none are

from the five degrees band visible from Alexandria but invisible from Rhodes. As Grant (2007, p. 20) summarises, “rather than go out and make observations, it seems Ptolemy spent his time in the Library at Alexandria cribbing many of Hipparchus's results and claiming them as his own.”

Isaac Newton may be the most eminent person to be accused of falsification. He was an irascible personality who regularly had spats with contemporaries. His greatest adversary was probably Leibniz, whose natural philosophy was at odds with Newton's theory of gravitation and laws of motion as outlined in his *Philosophiæ Naturalis Principia Mathematica*. Leibniz's influence in continental Europe meant that Newton's theory met with a lukewarm response there. In subsequent editions, Newton's case was strengthened by the adjustment of data, including his calculations of the velocity of sound, the precession of the equinoxes and the measurement of tides, so that they agreed precisely with his theory (Braud & Wade, 1982). Newton's reported measurements are given to six significant digits, a level of precision that is almost impossible even today (Westfall, 1973).

The Augustinian Friar, Gregor Mendel, is credited with making observations of the inheritance of characteristics across generations of pea plants in proportions that suggested a kind of discrete

transmission, so laying the foundation for a science of genetics. However, the proportions that Mendel reported fit so exactly with theoretical expectation that they drew the suspicion of R. A. Fisher, the eminent statistician who was responsible *inter alia* for the analysis of variance (ANOVA) test. Fisher closely examined Mendel's methods and data and found that the data were too good to be true, rather kindly concluding that Mendel's assistants may have adjusted figures in line with expectation. Others have suggested subconscious errors or selection (Judson, 2004, pp. 52-58), or less kindly suggested fudging (Anonymous, 1972).

Galileo Galilei is now thought to have fabricated the results of many of his experiments, which is somewhat ironic for the exemplary empiricist who privileged experimental observations over aesthetic or theoretical concerns. Tales of his testing the action of gravity by dropping objects from the Leaning Tower of Pisa are regarded as apocryphal but were invented later rather than by Galileo himself (Grant, 2007, p. 20). Other experiments that he claimed to have repeated “near a hundred times” with consistent results, could not have given that degree of homogeneity using the materials available at the time — as was found by contemporaries such as P re Mersenne who failed to replicate his findings (Braud & Wade, 1982, p. 26).

Fraud in Psychology

A high profile and extensive case of fraud from psychology involves Diederik Stapel. He was Professor of Social Psychology and Dean of the School of Social and Behavioural Sciences at Tilburg University. He had enjoyed a prolific career, receiving €2.2 million in grants from the Netherlands Organisation for Scientific Research (Enserink, 2012b), and producing 124 journal articles that had been cited a total of 1,756 times (Budd, 2013). One article, published in the flagship journal *Science*, reported an experiment conducted at the train station in Utrecht showing that a rubbish-filled environment tended to bring out racist tendencies in individuals (Stapel & Lindenberg, 2011). The cleaning staff had been on strike just before the summer vacation and this had provided an opportunity to compare responses from people who visited the station during the strike when the platform was unclean with those who were there once the cleaners had returned. He had speculated that expressed attitudes would be more stereotypical (“Brazilians are sexy, British people are polite, New Yorkers are pushy”) in the former, less orderly environment. Cleverly, he also designed a behavioural measure by inviting the predominantly white respondents to sit down on one of six seats laid out

in a row while they completed the measure, which happened to have a black person seated at one end, apparently already participating in the study. Stapel reported that participants did indeed express more stereotypical views in the untidy condition, and interestingly sat farther away from the already seated black person when their environment was messier, indirectly suggesting heightened racism. The problem for these elegant findings is that Stapel was so convinced that this phenomenon was a true property of the real world that he did not think it necessary to actually collect the data, which he instead invented at home so that they would give a cleaner result. He claimed that this was a consequence of his early experience with journal editors who found his real experimental data too complicated, with outcomes that were too messy, often asking him to leave out elements and make things simpler before they would publish — it was simpler to ensure the data were neat and consistent by making them up (Bhattacharjee, 2013).

Soon he was embroiled in an elaborate charade in which he would collaborate with research assistants or Ph. D. students on the development of research materials, such as questionnaires and bespoke equipment, but then (astonishingly) would insist on conducting the studies alone or taking the materials to contacts

of his in schools and colleges to administer. This allowed him to simulate the experiment at home to give a reasonable benchmark score and then create datasets around that figure that would give an unambiguous but believable confirmation of his hypothesis. He would bring the data sets or results of analyses back to colleagues and collaborate with them on the write-up. Stapel may have got away with his fabrications for so long because he ensured that his findings were in keeping with general expectations. “I didn’t do strange stuff, I never said let’s do an experiment to show that the earth is flat ... I always checked ... that the experiment was reasonable, that it followed from the research that had come before, that it was just this extra step that everybody was waiting for” (Bhattacharjee, 2013).

Ultimately his fraud was revealed when his collaborators asked about possible internal analyses (such as sex differences) that had not occurred to Stapel and so had not been concocted, or they asked for the raw data to conduct exploratory analyses but were told that they had been destroyed for lack of storage space. Suspicions were also raised by Stapel’s near faultless record of significance. One colleague was struck by how great the data looked, no matter what the experiment: “I don’t know that I ever saw that a study failed, which is

highly unusual ... Even the best people, in my experience, have studies that fail constantly. Usually, half don't work" (Bhattacharjee, 2013). An investigation by the universities that had employed him produced a final report that found Stapel had committed fraud in at least 55 of his papers, as well as in 10 Ph. D. dissertations written by his students (Levelt, 2012). By 2014, 58 of his published papers have been retracted.

Although Stapel may be labelled "perhaps the biggest con man in academic science" (Bhattacharjee, 2013), he is far from unique among psychologists: lack of space prevents me from discussing well documented cases involving Sir Cyril Burt, who invented data sets (and co-authors) to support the role of genetic inheritance on personality and intelligence (Tucker, 1997), Marc Hauser, who fabricated data and pressured his graduate students to reach his preferred conclusions (Wade, 2010), Dirk Smeesters, who had two high profile papers retracted when their data were found to be too good to be true (Enserink, 2012a), and Karen Ruggiero, who admitted to fabricating five experiments published in two articles and to doctoring research that appeared in a third (Price, 2010).

These detailed exposés can give the impression that they represent highly unusual aberrations perpetrated by individuals who can safely be regarded as

pathological (perhaps made so by stress or overwork). But we must remain wary that doing so could simply represent a defensive strategy that allows us to distance the miscreant from normal researchers and distinguishes their behaviour from normal practice. Braud and Wade (1982, p. 20) describe the genesis of fraud as often involving smaller steps, "those who falsify scientific data probably start and succeed with the much lesser crime of improving upon existing results. Minor and seemingly trivial instances of data manipulation — such as making results appear just a little crisper or more definitive than they really are, or selecting just the 'best' data for publication and ignoring those that don't fit the case — are probably far from unusual in science. But it is only a difference in degree between 'cooking' the data and inventing a whole experiment out of thin air". This resonates with Stapel's own account. In his biography, *Faking Science: A True Story of Academic Fraud*, he begins (p. iii): "I was doing fine, but then I became impatient, overambitious, reckless. I wanted to go faster and better and higher and smarter, all the time. I thought it would help if I just took this one tiny little shortcut, but then I found myself more and more often in completely the wrong lane, and in the end I wasn't even on the road at all. I left the road where I should have

Braud and Wade (1982, p. 20) describe the genesis of fraud as often involving smaller steps, "those who falsify scientific data probably start and succeed with the much lesser crime of improving upon existing results."

gone straight on, and made my own, spectacular, destructive, fatal accident."

Recent findings from psychology suggest that these more minor transgressions are relatively common. John, Loewenstein, and Prelec (2012) surveyed over 2,000 psychologists about their involvement in questionable research practices. Worryingly, they found that some practices occurred much more regularly, such as multiple analyses with selective reporting (78%) and optional stopping once significance has been reached (36%), but also practices that fall closer to the conscious fraud end of the continuum such as excluding data once their effects on the analysis are known (62%) and falsifying data (9%). Extending beyond psychology, Fanelli (2009) presented a meta-analysis of survey data on scientific misconduct that gave a pooled weighted mean of 1.97% who admitted to having ever fabricated, or falsified research data, but this figure rose to 14.12%

when asked if they had personal knowledge of a colleague who altered, fabricated or falsified research data, and increased further to 46.24% when misconduct was defined more comprehensively, for example as “experimental deficiencies, reporting deficiencies, misrepresentation of data, falsification of data” (p. 7). When the American Association for the Advancement of Science surveyed a random sample of its members, they found that 27% believed they had encountered or witnessed fabricated, falsified, or plagiarized research over the previous 10 years, with an average of 2.5 examples (Titus, Wells, & Rhoades 2008). Incidences are typically even higher when restricted to the biomedical sciences (Ranstam et al., 2000; Roberts & St John, 2014; Wells, 2008).

One indicator of fraud is the rate at which papers are subsequently retracted from journals. Steen, Casadevall, and Fang (2013) found 2,047 retracted articles indexed in PubMed, with the number having risen sharply in the last decade, both by reason of fraud and of “error” (which included plagiarism). This could not be attributed just to the increase in numbers of publications, but also represented a pro rata increase (though Gross, 2016, warns against over-interpreting these figures, since increased retraction rates might also reflect the greater attention being paid to malpractice or greater willingness to bring instances to the

attention of editors and publishers). Fang, Steen, and Casadevall (2012) followed up by consulting secondary sources to identify the reasons for retraction if none had been mentioned in the retraction notice (as was often the case). They found that 21.3% of retractions were attributable to error while 67.4% were attributable to misconduct, including fraud or suspected fraud (43.4%), duplicate publication (14.2%), and plagiarism (9.8%). It is not uncommon for retractions to be announced ambiguously and with little fanfare, so that they can go unnoticed. Retracted papers continue to be cited even after their retractions (Budd, 2013), including by 18% of the authors of retracted papers themselves, with less than 5% mentioning that the papers were retracted (Gross, 2016). Recently, internet sites such as Retraction Watch (<http://retractionwatch.com>) have been established with the goal of publicising unsound research. Their influence is difficult to gauge, but they may make it easier for dubious practices to be highlighted. Most detection of scientific misconduct is by laboratory colleagues of the transgressor, including their supervisors and students, who work at sufficiently close quarters to notice oddities in behaviour or data (Shamoo & Resnik, 2003). But the consequences for such whistleblowers can be severe. Lubalin, Ardini, and Matheson (1995) found that 47 of 68 complainants whom they surveyed suffered at least one negative consequence,

such as being pressured to withdraw their allegation, being ostracized by colleagues, or suffering a reduction in research support, and 56% believed that whistleblowing stigmatises the complainant. Retraction sites provide an opportunity for concerns to be raised anonymously, thus protecting the complainant, and can ensure that allegations about particular researchers are more widely discussed and acted upon rather than swept under the carpet. To illustrate, the blog site sciencefraud.org published 274 anonymous emails in the period July to December 2012 from bio-scientists claiming research misconduct in studies that had been published. In January 2013, legal threats forced the closure of this site, but a further 233 anonymous emails were submitted that could not be publicised. The fate of the papers referred to in these emails was analysed by Brookes (2014). He found no initial differences in the characteristics of the “public” and “private” cases, but the number of retractions was 650% higher for public cases, and the rate of publishing corrections or other errata were 770% higher. Overall, some kind of corrective action was taken on 23% of the publicly discussed papers, but this was only 3.1% for the unpublished papers.

Conclusions

In conclusion, then, although parapsychology does have two very well substantiated cases of fraud, they seem very typical

for the sector rather than signalling something distinct about the subject area (Levy is very reminiscent of frauds like the Cornell cancer researcher, Mark Spector, who was being groomed to take over as lab director, and Soal's case seems very similar to that of Cyril Burt). Given the recent revelations about misconduct across a range of disciplines that I have touched on here, the incidence of fraud in parapsychology does not seem to be very high when compared with the sector generally (though a per capita comparison could possibly suggest otherwise). It seems to me completely unwarranted, then, for commentators to imply that parapsychology has a particular problem with experimenter misconduct.

Indeed, there are grounds for thinking that parapsychology is actually much *less* susceptible to fraud than other research areas. Although there are as yet no systematic empirical studies of the characteristics of perpetrators of scientific misconduct, Gross (2016) has described the modal fraudulent scientist as “a bright and ambitious young man working in an elite institution in a rapidly moving and highly competitive branch of modern biology or medicine, where results have important theoretical, clinical, or financial implications. He has been mentored and supported by a senior and respected establishment figure who is often the co-author of many of his papers but may have not been closely involved in the research.”

And, based on their extensive primary research, Braud and Wade (1982, p 86) conclude that “the crime rate in science is influenced by three principal factors: the rewards, the perceived chances of getting caught, and the personal ethical standards of the scientist”. (To this I would add the expected consequences of getting caught, since there are a number of instances in which those who have admitted misconduct have been allowed by an institution to leave discreetly so as to not tarnish their reputation or embroil them in lengthy and costly legal and administrative proceedings). In other words, fraud will be more common where it is likely to be lucrative, where one's research can pass relatively unscrutinised (or even unnoticed), and where if discovered one's actions are more likely to be dealt with quietly rather than publicly. I would argue that none of these conditions pertain to parapsychology, which is highly under-resourced, is subject to very high levels of scrutiny whenever findings are positive, and which has a track record of public exposure when fraud is discovered.

This is not grounds to be complacent. Fraud is very difficult to detect without access to raw data. Kennedy (2014) notes that those perpetrating fraud are often reluctant to share raw data for reanalysis (which certainly was the case with Soal), and this may be more so now that the stochastic properties of “real” data are

better understood (for example Benford's Law, which describes the frequency distribution of leading digits in many naturally occurring data sets — see Benford, 1938, Miller, 2016). The case seems to me compelling for the establishment of a data repository so that the evidence on which research claims are made can be scrutinised by anyone who has an interest to, and this seems a natural extension to the design registries and repositories for unpublished research that are already available or are being developed.

References

- Anonymous (1972). Peas on Earth. *Horticultural Science*, 72, 438. Reproduced at http://rowdysites.msudenver.edu/~churchcy/BIO3600/bio3600_misc/Peas%20on%20Earth.htm
- Beloff, J. (1993). *Parapsychology: A concise history*. London, UK: Athlone Press.
- Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78, 551–572.
- Bhattacharjeethe, Y. (2013). Mind of a con man. *New York Times*, available at <http://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html?pagewanted=all>
- Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science*. London, UK: Century Publishing.
- Brookes, P. S. (2014). Internet publicity of data problems in the bioscience literature correlates with enhanced corrective action. *PeerJ* 2, e313; doi:10.7717/peerj.313
- Budd, J. M. (2013). The Stapel case: An object lesson in research integrity and its lapses. *Synesis: A Journal of Sci-*

ence, Technology, Ethics, and Policy, G47-G53.

Enserink, M. (2012a). *Rotterdam marketing psychologist resigns after university investigates his data*. June 25, 2012, Available at <http://www.sciencemag.org/news/2012/06/rotterdam-marketing-psychologist-resigns-after-university-investigates-his-data>

Enserink, M. (2012b). *Diederik Stapel under investigation by Dutch prosecutors*. Oct. 2, 2012, Available at <http://www.sciencemag.org/news/2012/10/diederik-stapel-under-investigation-dutch-prosecutors>

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* 4(5), e5738. doi:10.1371/journal.pone.0005738

Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *PNAS* 109, 17028-33

Goodstein, D. (2000). In defence of Robert Andrews Millikan. *Engineering & Science*, 4, 30-38. Available at <http://calteches.library.caltech.edu/4014/1/Millikan.pdf>

Grant, J. (2007). *Corrupted science: Fraud, ideology and politics in science*. Wisley, UK: AAPPL Artists' and Photographers' Press.

Gross, R. (2010). *Psychology: The science of mind and behaviour*. London, UK: Hodder Education.

Gross, C. (2016). Scientific misconduct. *Annual Review of Psychology*, 67, 693-711

Judson, H. F. (2004). *The great betrayal: Fraud in science*. London, UK: Harcourt, Inc.

Kennedy, J. E. (1975). *Summary of rat implementation work*. Unpublished document, available at <http://jeksite.org/psi/frat.pdf>.

Kennedy, J. E. (2014). *Experimenter misconduct in parapsychology: Analysis manipulation and fraud*. Published at

<http://jeksite.org/psi/misconduct.pdf> and <http://jeksite.org/psi/misconduct.htm>

Kohn, A. (1986). *False prophets: Fraud and error in science and medicine*. Oxford, UK: Basil Blackwell.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524-532.

Lubalin, S., Ardini, M. E., & Matheson, J. L. (1995). *Consequences of whistleblowing for the whistleblower in misconduct in science cases*. Washington, DC: Research Triangle Institute. Available at <https://ori.hhs.gov/sites/default/files/final.pdf>

Markwick, B. (1978). The Soal-Goldney experiments with Basil Shackleton: New evidence of data manipulation. *Proceedings of the Society for Psychical Research*, 56, 250-277.

Markwick, B., & West, D. J. (unpub.). *Dr Soal: A psychic enigma*. Unpublished manuscript.

Miller, S. J., (2016, Ed.). *Benford's Law: Theory and applications*. Princeton, NJ: Princeton University Press.

Newton, R. R. (1977). *The crime of Claudius Ptolemy*. Baltimore, MD: Johns Hopkins University Press.

Price, M. (2010). Sins against science. *Monitor on Psychology*, July/August 2010. Available at <http://www.apa.org/monitor/2010/07-08/misconduct.aspx>

Ranstam, J., et al. (2000). Fraud in medical research: An international survey of biostatisticians. *Controlled Clinical Trials*, 21, 415-27.

Rhine, J. B. (1974a). Security versus deception in parapsychology. *Journal of Parapsychology*, 38, 99-121.

Rhine, J. B. (1974b). A new case of experimenter unreliability. *Journal of Parapsychology*, 38, 215-225.

Rhine, J. B. (1975). A second report on a case of experimenter fraud. *Journal of Parapsychology*, 38, 306-325.

Roberts, D. L., & St John, F. A. (2014). Estimating the prevalence of researcher misconduct: A study of UK academics within biological sciences. *PeerJ* 2:e562; DOI 10.7717/peerj.562

Shamoo, A. E., & Resnik, D. B. (2003). *Responsible conduct of research*. New York, NY: Oxford University Press

Stapel, D. (2012). *Faking science: A true story of academic fraud* (trans. Nicholas J. L. Brown), available under a creative commons licence from <http://nick.brown.free.fr/stapel/FakingScience-20141214.pdf>

Stapel, D. A., & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, 332(6026), 251-253.

Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why has the number of scientific retractions increased? *PLoS ONE* 8, e68397. doi:10.1371/journal.pone.0068397

Stokes, D. (2015). The case against psi. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.) *Parapsychology: A handbook for the 21st century* (pp. 42-48). Jefferson, NC: McFarland.

Titus, S. L., Wells, J. A., & Rhoades, L. J. (2008). Repairing research integrity. *Nature*, 19, 980-982.

Tucker, W. H. (1997). Re-reconsidering Burt: Beyond a reasonable doubt. *Journal of the History of the Behavioral Sciences*, 33(2), 145-162.

Wade, N. (2010, August 27). Harvard researcher may have fabricated data. *New York Times*. Available at <http://www.nytimes.com/2010/08/28/science/28harvard.html>

Wells, J. A. (2008). *Final report: Observing and reporting suspected misconduct in biomedical research*. Washington, DC: Gallup Org. http://ori.hhs.gov/sites/default/files/gallup_finalreport.pdf

Westfall, R. S. (1973). Newton and the fudge factor. *Science*, 179, 751-758.