



Working Paper Series

SWPS 2015-24 (August)

Do Academics Doubt Their Own Research?

Lee Stapleton

US

University of Sussex

SPRU Working Paper Series (ISSN 2057-6668)

The SPRU Working Paper Series aims to accelerate the public availability of the research undertaken by SPRU-associated people of all categories, and exceptionally, other research that is of considerable interest within SPRU. It presents research results that in whole or part are suitable for submission to a refereed journal, to a sponsor, to a major conference or to the editor of a book. Our intention is to provide access to early copies of SPRU research.

Editors

Tommaso Ciarli
Daniele Rotolo

Contact

T.Ciarli@sussex.ac.uk
D.Rotolo@sussex.ac.uk

Associate Editors

Florian Kern

Area

Energy

F.Kern@sussex.ac.uk

Paul Nightingale,
Ben Martin, &
Ohid Yaqub

Science, & Technology Policy

P.Nightingale@sussex.ac.uk
B.Martin@sussex.ac.uk
O.Yaqub@sussex.ac.uk

Matias Ramirez

Development

Matias.Ramirez@sussex.ac.uk

Joe Tidd &
Carlos Sato

Technology Innovation
Management

J.Tidd@sussex.ac.uk
C.E.Y.Sato@sussex.ac.uk

Maria Savona &
Mariana Mazzucato

Economics of Technological Change

M.Savona@sussex.ac.uk
M.Mazzucato@sussex.ac.uk

Andrew Stirling

Transitions

A.C.Stirling@sussex.ac.uk

Caitriona McLeish

Civil military interface

C.A.McLeish@sussex.ac.uk

Disclaimer

The works available here are the responsibility of the individual author(s) and do not necessarily represent the views of other SPRU researchers. As matters of policy and practice, SPRU does not endorse individual research contributions.

Guidelines for authors

Papers shall be submitted in pdf or Word format. They should contain a title, an abstract, and keywords. Papers should be submitted to one of the Editors, who will process them and send them to the appropriate Associate Editor. Two members of SPRU will be asked to provide a short written review within three weeks. The revised versions of the paper, together with a reply to the reviewers, should be sent to the Associate Editor, who will propose to the Editors its publication on the series. When submitting the authors should indicate if the paper has already undergone peer-reviewing, in which case the Associate Editors may decide to skip internal review process.

Websites

SWPS: www.sussex.ac.uk/spru/research/swps

IDEAS: ideas.repec.org/s/sru/ssewps.html

Research Gate: www.researchgate.net/journal/2057-6668_SPRU_Working_Paper_Series

Do academics doubt their own research?

Lee Stapleton

SPRU (Science Policy Research Unit), School of Business, Management & Economics,

University of Sussex, Brighton, BN1 9SL, U.K.

E-mail l.stapleton@sussex.ac.uk Tel +44(0) 1273 872781

Abstract

When do experts doubt or question their own previously published research and why? An online survey was designed and distributed across academic staff and postgraduate research students at different universities in Great Britain. Respondents ($n = 202 - 244$) identified the likelihoods of six different (quasi) hypothetical occurrences causing them to doubt or question work they have published in peer reviewed journals. They are: two objective and two semi-objective citation based metrics, plus two semi-objective metrics based on verbalised reactions. Only limited support is found from this study to suggest that the authors of primary research would agree with any judgements made by others about their research based on these metrics. The occurrence most likely to cause respondents to doubt or question their previously published research was where the majority of citing studies suggested mistakes in their work. In a multivariate context, only age and nationality are significant determinants of doubt beyond average likelihoods. Understanding and acknowledging what makes authors of primary research doubt their own research could increase the validity of those who pass judgement.

Key words: citations; criticism; experts; meta-analysis; peers; systematic reviews

1. Introduction

1.1. To err is human

Using individual academics as the unit of analysis, studies on uncertainty have considered intersubjective differences of opinion - the extent to which different academics agree or disagree on a particular question or issue (e.g. Aspinall, 2010; Stirling, 2010). Differences of opinion do not just occur between people. Intrasubjectively, people change their minds all the time and it is not difficult to conceive of experts changing their minds. Bell and Morse (2008, p. 204) note that “[w]hen we read the books of, and listen to the lectures of, scientists, it may appear that they are splendidly confident creatures, comprehending and

understanding the world in terms of their science. Yet those of us that count scientists as our friends and know them personally know that, like the rest of us, they are often riddled with self-doubt and concern and anxiety about their work". Similarly, in a technological context MacKenzie (1998, p. 326) notes how those directly involved in the production of knowledge are relatively uncertain "about matters such as reliability, safety or predictability of the technology". In the context of "guidebooks and working manuals of all sorts" Fleck (1998, p. 157) notes how their status "is essentially provisional, as they are typically always being rewritten". However, the instances where academics are publicly frank about changing their mind, may not reflect the true scale of the phenomenon: "most of the individual learning which experienced researchers do in the course of their research remains private and not spoken about" (Brew, 2001, p. 12). This, at least partly, explains why changes of opinion amongst academics is understudied, approached only at the group or community level, not the individual (Section 1.2).

1.2. Insights from consensus forming methods

Generalising to experts, more broadly than academics, changes of opinion over *short* periods of time have long been studied in the context of group approaches to consensus forming. There are different methods available for catalysing the formation of a consensus or, less ambitiously, narrowing the range of views amongst a group of experts who disagree on a particular question or issue. These include face-to-face discussion, Delphi (Dalkey and Helmer, 1963) and Estimate-Talk-Estimate (Gustafson et al., 1973). The Delphi method, originally developed and applied in a military context, involves using questionnaires to elicit the opinion of multiple experts on one or more questions at two or more different points in time. Each expert is then provided with an anonymous answer summary and possibly the arguments put forward by the other participants justifying their answers. The provision of the answer set can catalyse a convergence of opinion in the subsequent round(s). In the words of its developers, Delphi is a methodology to "obtain the most reliable consensus of opinion of a group of experts [...] by a series of intense questionnaires interspersed with controlled opinion feedback" (Dalkey and Helmer, 1963, p. 458). Experts who take part in a Delphi exercise do not meet each other in order for any convergence of opinions to occur. Virtual interaction is arbitrated by a facilitator. Estimate-Talk-Estimate is similar to Delphi except that face-to-face discussion is used between elicitation rounds instead of anonymous feedback. The rationale for the interaction element of these methods (anonymous and virtual in the case of Delphi) assumes that more knowledgeable experts will be less likely to modify their answers compared to those who are less knowledgeable - the latter being more likely to adopt mimetic behaviour (Munier and Rondé, 2001). There are alternative methods to consensus forming which do not rely on experts changing their opinions after feedback from or discussion with other experts. The simplest is statistical averaging of individual responses. Another method involves using objective seed questions to differentially weight answers. Experts who fare better on the seed questions have their

opinions weighted greater than those who fare less well (Aspinall, 2010; Bamber and Aspinall, 2013; Cooke, 1991). Despite these alternatives, approaches for reaching consensus which make use of opinion changes in the context of (virtual or real) groups are still widely used. Short-term changes of opinion, facilitated using consensus forming methods, invite the question of longer-term changes of opinion not least because they are intuitively more likely if divergence of opinion is a function of time.

Beyond short-term group based consensus forming exercises, longer term changes of academic opinion have hitherto been approached using scientific communities as the unit of analysis (Collins, 1999; Keller and McInerney, 2008; Kolstad, 1996; Oppenheimer et al, 2008), not individual academics. Changes in the views of a scientific community over time could be a function of group dynamics, entrants and leavers to that community and contextual factors rather than reflecting changes in the views of individuals over time.

1.3. Insights from research synthesis methods

Published studies are often taken at face value when they are cited failing to consider that the cited authors may not agree with the (positive, negative or ambivalent) context of the citation. Similarly, the opinions of primary authors are neglected by research synthesists (producers of narrative reviews, systematic reviews and meta-analyses). Good practice guidelines for research synthesis draw attention to the need to think carefully about the period over which studies are included e.g. “the basis for this timeframe should be theoretical rather than arbitrary [...] problems may evolve in ways that reduce the meaningfulness of older research to the present” (Wilson, 2009, pp 162-163). *Meaningfulness* is established on the basis of current dominant trends in the literature by research synthesists, not the authors of the original primary research. Another guideline states that studies should be included “based on the suitability of the methods for studying the synthesis question” (Cooper and Hedges, 2009a, p. 9). *Methodological suitability* is again established on the basis of current dominant trends in the literature by research synthesists, not the original authors.

Beyond these eligibility criteria, the process of synthesising research may involve weighting research dependent on its quality. The research synthesis literature is equivocal on the relationship between study quality and outcomes. Some argue that problematic methods may not necessarily lead to biased results (e.g. Cooper, 1989; Hunter and Schmidt, 2004; Woodward and Wui, 2001). Others suggest relationships between study quality and study outcomes (e.g. Chalmers et al., 1983; Moher et al., 1998; Nurmohamed et al., 1992; Schulz et al., 1995; Stanley, 2001). To what extent are these conclusions invalidated because of difficulties in assessing quality? It is an onerous task for research synthesists (and most citers of primary research) to establish meaningfulness, suitability and quality of research *which has been carried out by others*. To do this effectively presumes, for example, that methods and results in primary studies are sufficiently documented. A lack of

documentation (on, for example, methodological assumptions) is not the same thing as a lack of adherence (to those assumptions). The original authors may or may not subscribe to the conclusions drawn by research synthesists with respect to these issues. Further, there is no guarantee that research synthesists will follow good practice in terms of identification, analysis and reporting (Hopewell et al., 2013; Moher et al., 1999; Moher et al., 2009; Moja et al., 2005; Stanley et al., 2013) which begs the question of the relative validity of their judgements compared to the authors of primary research whose work they are dependent upon.

1.4. Insights from bibliometrics

Partly driven by pressure from governments and research funding organisations, universities are increasingly fixated on metrics in order to assess the productivity of individual academic staff and, in the aggregate, their competitiveness relative to other institutions (Wilsdon et al., 2015). Perhaps the simplest and most obvious way of assessing individual productivity is in terms of output volume. Although crude in its most basic form, outputs can be constrained, for example, to publications in high quality journals where quality could be a function of the differential citations received by articles across journals e.g. journal impact factor (Garfield, 2006). Indeed, basic citation counts, on their own, are a widely used metric for establishing the worthiness of academics relative to their peers (Wilsdon et al., 2005). However, quality of research is not always a good predictor for the number of citations compared to: the reputation (Callahan et al., 2002; Nieminen et al., 2006) and language (Bornmann et al., 2012) of the publishing journal; the reputation of the publishing authors (Bornmann et al., 2012; Leimu and Koricheva, 2005); the magnitude of interest in the phenomena under study (Ioannidis, 2005a; Nieminen et al., 2006). Articles positioned earlier in a journal issue may get cited more frequently compared to those which appear later, regardless of their respective quality (Ayres and Vars, 2000; Hudson, 2007). The rejection of null hypotheses (positive results) may also serve to increase citations independent of whether such studies are of higher quality compared to those which accept the null (Fanelli, 2013). Number of citations may not even be a good predictor of the extent to which published work is actually used (MacRoberts and MacRoberts, 2010) and, perhaps most worryingly, they are vulnerable to gaming (Wilsdon et al., 2015). Understanding when and why academic staff might agree, or otherwise with citation-based judgements about their work would be a useful complement to the extant bibliometrics literature.

1.5. Summary and research questions

Opinion change amongst academics is somewhat of a black-box: under-studied and under-understood beyond the private ruminations of the subjected academics (Section 1.1). However, the history and development of consensus forming methods confirms that people's opinions are not always fixed – they can and do change e.g. when presented with high quality counter-arguments or a majority of dissenters (Section 1.1.2). Judgement is

regularly passed against the research of others e.g. via quality weightings assigned to primary studies by research synthesists. These judgements are passed without considering or validating whether those who carried out the original primary research would adhere to those judgements (Section 1.4). Finally, the worth-iness of academics is increasingly being quantified by metrics without considering or caring about which metrics, if any, those academics find worth-y (Section 1.4). These observations lead to the following two research questions which are addressed in this exploratory study:

- Under what circumstances might academics come to doubt or question their own previously published research?
- How does doubt co-vary with socio-demographic and employment characteristics?

2. Methodology

2.2. Survey design

An online survey consisting of 12 questions was designed which begins with contextual questions (Questions 1 – 9) covering: demographics, seniority, contract type, subject area, number of peer-reviewed journal publications and year of first publication. The focal question (Question 10) elicited the subjective likelihoods of six occurrences inducing respondents to doubt or question work they had published in peer reviewed journals (Table 1).

Table 1. (Quasi) hypothetical doubt-inducing occurrences

Category	Occurrence	Abbreviation
Objective citation metric	Substantially fewer citations compared to your other work	CIT_YOU
Objective citation metric	Substantially fewer citations compared to work by others on the same topic	CIT_OTH
Semi-objective citation metric	The majority of citing studies suggest mistakes in your work	CIT_MIS
Semi-objective citation metric	The majority of citing studies suggest alternative ways of doing things	CIT_ALT
Semi-objective verbal metric	In the workplace, the majority of verbalised peer reactions to work you have published are adverse, not positive	VER_WRK
Semi-objective verbal metric	In other work contexts (e.g. at meetings or conferences), the majority of verbalised peer reactions to work you have published are adverse, not positive.	VER_MTG

Each occurrence includes a quantity anchor – either *majority* (an interval anchor) or *substantially* (an ordinal anchor). Majority covers a wide but defined interval, 50.1-100%. However, people understand the concept of majority subjectively. Some may not register a proportion as a majority if it is below 70%. There is no equivalent defined interval for *substantially*. It is an orderable (ordinal) concept, larger than other orderable concepts such as *moderately*, which in turn is larger than *slightly*. Both majority and substantially could be replaced by specific proportions (ratio anchors) such as ‘70%’, ‘four fifths’ or ‘60-100%’ but increasing specificities given the sensitive subject matter could serve to increase mis-response and non-response biases. Being more specific could also give rise to other problems, for example creating uncertainty about the range over which results are applicable. If 80% of citing studies suggesting mistakes in your work (CIT_MIS) is an occurrence which is very likely to make you doubt your work, would this still be the case if the proportion fell to 70%? What about 60%?

No occurrence is explicit about quality. The admittance of doubt is arguably logical when, for example, exposed to robust, high quality arguments against some aspect of your research. Quality issues could be approached partially, to avoid truisms. CIT_MIS, for example, could be modified to focus on mistakes which affect salient findings or mistakes suggested by the authors of articles published in high impact journals or journals which the respondent has tended to revere. In short, the 6 hypothetical occurrences could be complemented by alternatives.

Outputs in peer-reviewed journals as opposed to outputs more generally (e.g. books, conferences, reports to funders, working papers) were specified to avoid confounding effects. A limitation being that disciplines differ in their propensities to disseminate research in journals (Butler and Visser, 2006; Sorzano et al., 2014). The likelihood of each occurrence inducing respondents to doubt or question their previously published research was answerable using a 5-point Likert item ranging from *Very unlikely* to *Very likely*. 3 additional answer options were made available in each case: *Rather not say*; *Don't know*; *Other* (RDO). To militate against mis-response and non-response bias these occurrences could be regarded as hypotheticals, rather than actuals. Respondents are not asked to identify that each has occurred, they are asked to identify the likelihood of a reaction (doubt) *if* they occurred. Respondents are also asked (Question 11) if they doubt or question any of the peer reviewed journal articles they have published. This question goes beyond the hypothetical, directly eliciting whether doubt has occurred. Recognising the possibility of response biases because of its sensitivity, this question is: (a) *crude*: answerable using a nominal yes / no scale (with RDO options) requiring no specific information; (b) *contextual*: only included to understand if and how answers are related to answers from other questions, principally Question 10; (c) *penultimate*: positioned towards the end of the survey, behind only an optional feedback / comments question. The survey was pre-tested ($n = 5$) and piloted ($n = 7$) in October and November 2014 respectively with respondents

selected using convenience sampling. These procedures resulted in the removal of one question and minor amendments to the wording of other questions compared to the final version available in Supplementary Material A.

2.3. Survey distribution

Ethical approval for the study, including the sampling strategy, was granted in December 2014. The final version of the survey was distributed using a combination of targeted, convenience and random sampling between January – March 2015. Targeted and random sampling were more effective in terms of generating responses compared to convenience sampling. 4 universities in Great Britain with accessible ‘global e-lists’ to instantaneously e-mail all academic staff and students were targeted. Gatekeepers (research ethics officers) were contacted at each institution to obtain permission to access e-lists. Permission granted and survey distributed to all staff and students via global e-list at 1 institution. In terms of random sampling, 10 British universities were randomly selected. Gatekeepers were contacted to obtain permission to e-mail Heads of Schools (or their equivalent) at each institution. Permission granted by 2 institutions to e-mail survey to staff and postgraduate students via Heads of Schools. 5 (6) Heads at institution 1 (institution 2) confirmed distribution. Convenience sampling exploited contacts and networks at 7 institutions. This includes the author’s institution but excludes distribution amongst colleagues and students in the author’s School.

2.4. Survey analysis

Standard univariate descriptive statistics are followed by bivariate tests of differences and relationships, according to data types: Pearson’s chi-square (nominal by nominal), Mann-Whitney U (nominal by ordinal with 2 independent samples), Kruksal-Wallis chi-square (nominal by ordinal with >2 independent samples), independent t-test (nominal by ratio) and Spearman’s rho (ordinal by ordinal; ordinal by ratio). Multivariate static ordinary least squares (OLS) regression is used to simultaneously test for multiple determinants of doubt. The bivariate and multivariate tests use simple equally-weighted composite measures (Nardo et al., 2005) of doubt for each respondent (Equation 1) to allow for analysis which is easier to interpret compared to referring to the disaggregated doubt indicators. Potential disadvantages of composites include hiding critical, unusual trends in one or more of the constituent (Jüni et al., 1999; Saisana and Tarantola, 2002).

$$Doubt_i = \sum_{q=1}^Q w_q d_{qi}$$

With: $\frac{1}{Q} = w_q, \sum_q w_q = 1, 0 \leq w \leq 1$, for all $q = 1 \dots Q, i = 1 \dots N$. Where: w = weight; d = likelihood q = occurrence, Q = occurrences

Equation 1

The multivariate tests also use the non-aggregated doubt indicators. Ordinary least squares (OLS) is used in preference to quantile (Koenker and Bassett, 1978) and ordinal (Aitchison and Silvey, 1957; McCullagh, 1980) regression methodologies. The former invoke non-unique solutions in half of the tests on the non-aggregated doubt indicators. The latter cannot accommodate composite non-integer ordinal dependent variables (doubt; Equation 1) and generate coefficients which are not easy to interpret. Three models are developed using composite doubt as the dependent variable, differing in terms of specification: non-log, semi-log, double-log (Equation 2)

$$Doubt_i^1 = \beta_0^1 + \beta_1^1 G_i + \beta_2^1 N_i + \beta_3^1 A_i + \beta_4^1 S_i + \beta_7^1 P_i + u_i^1$$

$$\ln Doubt_i^2 = \beta_0^2 + \beta_1^2 G_i + \beta_2^2 N_i + \beta_3^2 A_i + \beta_4^2 S_i + \beta_7^2 P_i + u_i^2$$

$$\ln Doubt_i^3 = \beta_0^3 + \beta_1^3 G_i + \beta_2^3 N_i + \beta_3^3 \ln A_i + \beta_4^3 S_i + \beta_7^3 \ln P_i + u_i^3$$

Where: $Doubt_i$ = equally-weighted composite measure of doubt; G_i = dummy variable for gender (male =0; female =1); N_i = dummy variable for nationality (British =0; other =1); A_i = Age; S_i = dummy variable for seniority (non-professors =0; professors =1); P_i = productivity; u_i = idiosyncratic error

Equation 2

Coefficients from the optimal model are reported and discussed. The optimal model is determined based on relative performance with respect to scores against 7 diagnostic criteria (Table 2). The specification of this model – non-log, semi-log or double-logged – is then adopted for regression analysis on each of the six disaggregated doubt indicators.

Table 2. Scoring rules against diagnostic criteria

No.	Test	Description	Score for 'yes'
1	Coefficient signs	Do all statistically significant coefficients (p<0.05) have the expected signs?	2
2	Coefficient magnitudes	Do all statistically significant coefficients have plausible magnitudes?	2
3	Heteroscedasticity	Lagrange multiplier	1
4	Normality	Lagrange multiplier	1
5	Multicollinearity	Centred variance inflation factors	1
6	R ²	Goodness of fit. Involves comparing and ranking the non log, semi log and	Max of 2

		double-log model variants. Score 2 for highest, 1 for 2 nd , 0 for 3 rd	
7	RESET	Regression specification error test with 1 fitted term	2

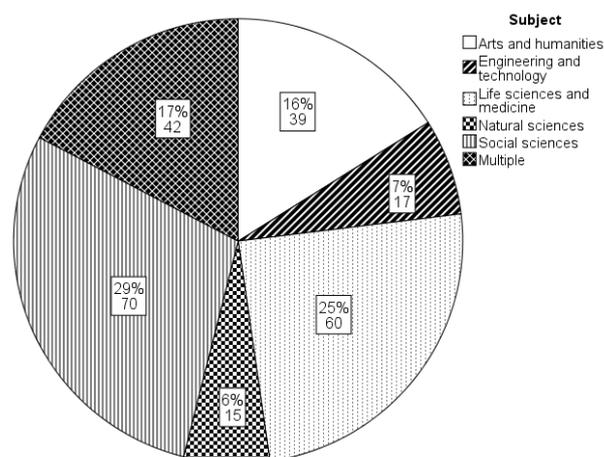
3. Results

3.2. Descriptive statistics

A slim majority (125 or 51.2%) of the 244 respondents were female. Ages ranged from 24-78 (mean = 44.04). Most (164 or 67.2%) are UK nationals and all identified the UK as the country in which they work or study. Respondents varied in terms of seniority - postgraduate students (28 or 11.6%), research assistants and research officers (5 or 2.1%) at the junior end of the spectrum through to Full Professors (41 or 17.0%).

In terms of contract type, excluding postgraduate students, the majority of respondents were permanently employed (134 or 54.9%) compared to fixed term (51 or 20.9%) or open-ended (18 or 7.4%) alternatives. In identifying the broad subject area in which peer-reviewed journal articles are published, the modal response was *Social sciences* with *Natural sciences* being least represented in the sample (Figure 1).

Figure 1. The subject area of peer-reviewed journal publications identified by respondents.



When asked about the number of peer reviewed journal articles published to date, answers ranged from 1 - 400 (mean = 29.1). The earliest respondent first published in 1964, the most recent in 2015 (mean = 2001). Productivity, formed by dividing number of publications by active publishing years, ranged from 0.1 - 13.9 (mean = 1.5). Taking the doubt-inducing occurrences (Table 3; Figure 2a-e) and excluding RDO answers to ordinalise the data, *'The majority of citing studies suggest mistakes in your work - CIT_MIS'* was the occurrence most

likely to cause respondents to doubt or question their published work (median = *Very likely*; mode = *Very likely*). ‘*Substantially fewer citations compared to your other work – CIT_YOU*’ was the occurrence least likely to induce doubt (mean = 2.51; median = *Quite unlikely*; mode = *Quite unlikely*). ‘*Substantially fewer citations compared to work by others on the same topic – CIT_OTH*’ was similarly unlikely to induce doubt (mean = 2.74; median = *Neither likely or unlikely*; mode = *Quite unlikely*).

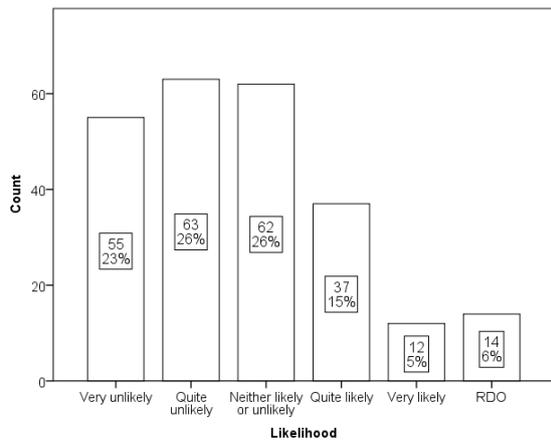
Table 3. Mean, median and modal likelihoods

	Doubt inducing occurrence					
	CIT_YOU	CIT_OTH	CIT_MIS	CIT_ALT	VER_WRK	VER_MTG
Mean likelihood	2.51	2.74	4.28	3.15	3.73	3.84
Median likelihood	<i>Quite unlikely</i>	<i>Neither likely or unlikely</i>	<i>Very likely</i>	<i>Neither likely or unlikely</i>	<i>Quite likely</i>	<i>Quite likely</i>
Modal likelihood	<i>Quite unlikely</i>	<i>Quite unlikely</i>	<i>Very likely</i>	<i>Neither likely or unlikely</i>	<i>Quite likely</i>	<i>Very likely</i>
N	229	231	231	229	225	229

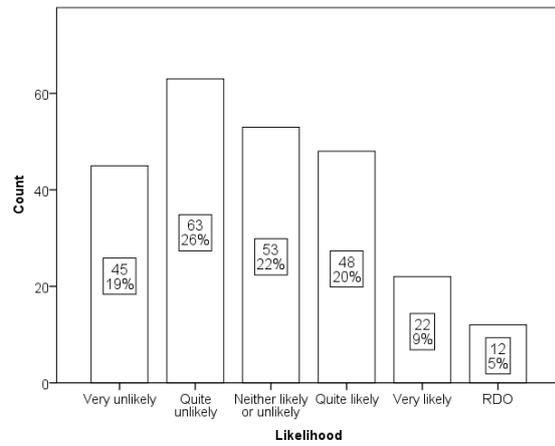
Non-integer mean likelihoods interpreted using: 1 = very unlikely; 2 = quite unlikely; 3 = neither likely or unlikely; 4 = quite unlikely; 5 = very likely

Figure 2. Distribution of likelihoods.

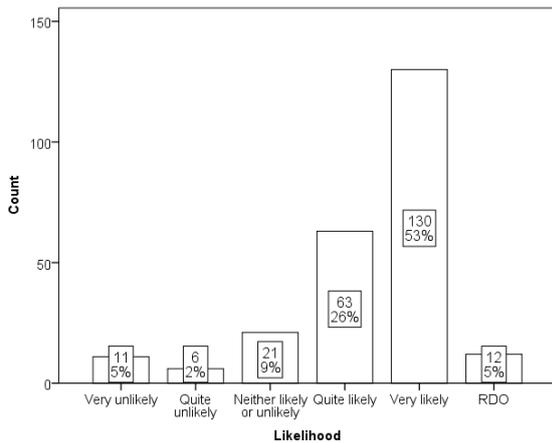
(a) substantially fewer citations compared to your other work



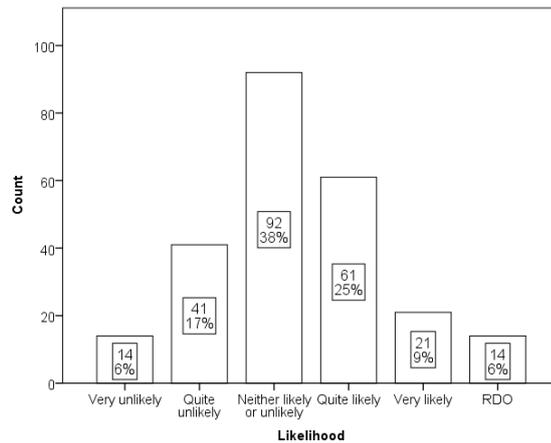
(b) substantially fewer citations compared to work by others on the same topic



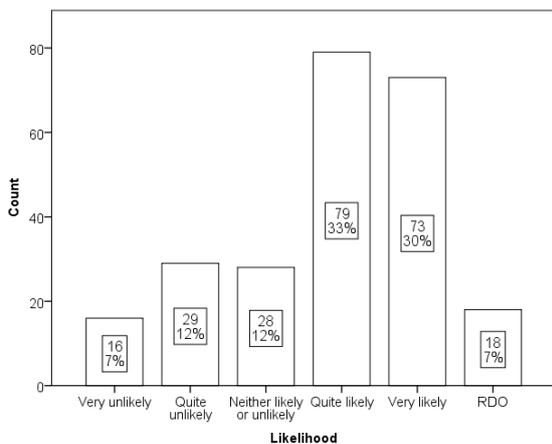
(c) The majority of citing studies suggest mistakes in your work



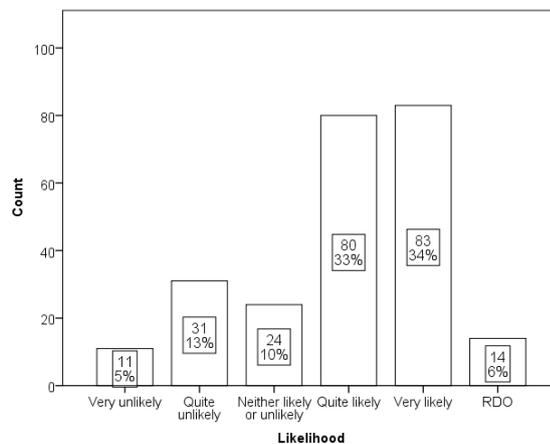
(d) The majority of citing studies suggest alternative ways of doing things



(e) In the workplace, the majority of verbalised peer reactions to work you have published are adverse, not positive

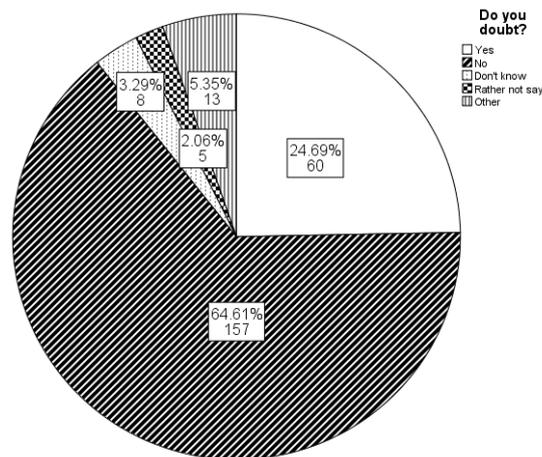


(f) In other work contexts (e.g. at meetings or conferences), the majority of verbalised peer reactions to work you have published are adverse, not positive



In terms of the three occurrences most likely to induce doubt, significantly more than half of respondents identified themselves as being *Quite likely* or *Very likely* to doubt their work if the majority of verbalised peer reactions were adverse not positive (both in the work place and in other work contexts such as conferences). Over three quarters of respondents self-identified as being *Quite likely* or *Very likely* to doubt their work if the majority of citing studies suggested mistakes in their work. By contrast, only a quarter of respondents claimed to actually doubt or question any of the peer reviewed journal articles they have published (Figure 3). This should be treated as a lower-bound because of possible mis-response (selecting No when you mean Yes) and non-response (if people who did not complete the survey are more likely to be doubters than the average respondent) biases.

Figure 3. Do you doubt or question any of the peer reviewed journal articles you have published?



3.3. Bivariate statistics

Intrasubjectively the likelihoods are reasonably correlated (Cronbach's $\alpha = 0.77$). In other words, respondents who are relatively unlikely (likely) to doubt their work when faced with one occurrence would be similarly unlikely (likely) to doubt their work when faced with another occurrence. Women are significantly more likely to doubt compared to men (median doubt = 3.46 versus 3.25; Mann-Whitney U = 4564.5; P = 0.036 < 0.05). UK citizens are more likely to doubt (median doubt = 3.47) compared to nationals of other countries (median doubt = 3.17) and this difference is also statistically significant (Mann-Whitney U = 3796.5; P = 0.018 < 0.05). Age is weak-inversely, and significantly, related to doubt with younger (older) people being more (less) likely to doubt or question their work (Spearman's rho = -0.198; P = 0.002 < 0.05). Similarly, and collinear with age, seniority is weak-inversely, and significantly related to doubt: postgraduate students and early career researchers tend to doubt and question their research more than their senior counterparts (Spearman's rho = -0.132; P = 0.030 < 0.05). Total number of publications and year of first publication are both

collinear with age and seniority, so unsurprisingly also exhibit relationships with doubt: the more you have published (Spearman's rho = -0.196; $P=0.002<0.05$) and the earlier you started publishing (Spearman's rho = 0.251, $P=0.000<0.05$) the less you doubt. A significant relationship also exists when considering productivity: the higher your mean annual output of peer-reviewed journal articles, the lower your propensity to doubt (Spearman's rho = -0.139; $P = 0.023>0.05$). Neither the subject area in which respondents tend to publish (Kruskal-Wallis chi-sq = 6.94; $P = 0.225>0.05$) nor the contract type governing the employment of respondents (Kruskal-Wallis chi-sq = 2.56; $P = 0.465>0.05$) were significant factors affecting doubt.

Perhaps surprisingly, respondents who claim they doubt or question one or more of their publications were not more likely to doubt when presented with the hypothetical occurrences compared to those who stated they do not doubt (median doubt for both groups = 3.33). Although women (men) were more (less) likely to doubt their published research when presented with the hypothetical occurrences, this is (nominally but not significantly) reversed when considering actual doubt. 31% of men and 24% of women claim to doubt or question one or more articles (Pearson's chi-sq = 1.452; $P = 0.228>0.05$). As per hypothetical doubt, UK nationals are significantly more likely to actually doubt their work compared to nationals of other countries. 32% of UK nationals and 19% of other respondents admit doubt (Pearson's chi-sq = 3.907; $P = 0.048<0.05$). Younger people were also more likely to actually doubt previously published research (independent-t = 2.043; $P=0.044<0.05$). With seniority the pattern is not so straightforward. Although this was inversely and significantly related to hypothetical doubt, there is no such linear trend with actual doubt. Instead both full professors and junior research staff are significantly more likely to doubt compared to other staff and postgraduate students¹ (Pearson's chi-sq = 11.64; $P = 0.020<0.05$). It is reasonable to hypothesise that as the number of articles published by respondents increases, *ceteris paribus*, so does the probability of actually doubting one or more of those publications across the *entire* portfolio. This is not inconsistent with the observation that localised hypothetical doubt induced with respect to *specific* publications decreases as the number of publications increases. The mean number of publications associated with respondents who admit doubt is significantly higher compared to those who claim not to doubt (46.6 versus 21.2; independent t = 2.720; $P = 0.008<0.05$). Similarly, respondents who admit doubt tended to start publishing significantly earlier compared to those who claim not to doubt (1997 versus 2002; independent t = -2.499; $P = 0.014<0.05$). The mean publication intensity of doubters is also significantly greater than non-doubters (2.08 versus 1.29; independent t = 2.34, $P = 0.022<0.05$). As per hypothetical doubt, the subject area respondents tend to publish in is not a significant factor affecting actual doubt (Pearson's chi-sq = 0.904; $P = 0.924>0.05$). Neither does

¹ Conflation (from 7 to 5 categories) used here and below (subject – 6 to 5 categories; contract type – 4 to 2 categories) to meet chi-square's minimum expected cell size assumption.

contract type appear to play a role in influencing actual doubt (Pearson's chi-sq = 1.775; P = 0.183>0.05).

3.4. Multivariate statistics

The diagnostic performance of the non-logged model was higher compared to the semi-log and double-log alternatives (Table 4).

Table 4. Model diagnostic performance

No.	Test	Non-log	Semi-log	Double-log
1	Coefficient signs	2	2	2
2	Coefficient magnitudes	2	2	2
3	Heteroscedasticity	1	0	0
4	Normality	0	0	0
5	Multicollinearity	1	1	1
6	R ²	2	1	0
7	RESET	2	2	2
Diagnostic performance		91%	73%	64%

Taking the non-logged model forward, only age and nationality are significant predictors of the extent to which the six hypothetical occurrences would, overall, induce respondents to doubt or question their previously published research (Table 5). The mean doubt likelihood across all respondents in the regression analysis is 3.377. The ceteris paribus likelihood of non UK nationals doubting their work is, on average, 0.333 units less than UK nationals (P=0.004<0.05). Whereas, a 1 (or 10) year increase in the age of respondents is associated with a 0.014 unit (or 0.14) reduction in doubt likelihood (P=0.004<0.05).

Table 5. Coefficient estimates from the non-log model (n=202)

Variable	Coefficient	Std. Error	t-Statistic	P
Gender	0.174	0.111	1.564	0.119
Nationality	-0.333	0.116	-2.872	0.004
Age	-0.014	0.005	-2.909	0.004
Professors	0.167	0.169	0.988	0.324
Productivity	-0.033	0.038	-0.857	0.392
C	4.055	0.248	16.341	0.000

R-squared	0.093	Mean dependent var	3.377
-----------	-------	--------------------	-------

Using this non-logged model specification but replacing the (composite doubt) dependent variable with the six individual indicators in turn, there is limited variability in the estimates for the nationality and age coefficients (Table 6). In the three regressions where nationality is significant (CIT_OTH, CIT_MIS and VER_WRK) the mean of the dependent variable for non-UK nationals is at least 0.310 units lower (CIT_MIS). The biggest decrease is associated with CIT_OTH where the mean of the dependent variable decreases by 0.396 units from 2.758 to 2.362. In the four regressions where age is significant (CIT_OTH, CIT_MIS, VER_WRK and VER_MTG) a 1 (or 10) year increase in age is associated with at least a 0.016 (or 0.16) unit decrease in doubt (CIT_MIS). The biggest age effect is associated with VER_WRK where the mean of the dependent variable decreases by 0.023 (or 0.23) units for every 1 (or 10) year increase in the age of respondents.

Table 6. Nationality and age coefficients associated with the disaggregated doubt indicators

Coefficient / standard error / significance	Doubt inducing occurrence					
	CIT_YOU	CIT_OTH	CIT_MIS	CIT_ALT	VER_WRK	VER_MTG
Nationality	-0.266	-0.396	-0.310	-0.180	-0.385	-0.361
Std. Error	0.171	0.181	0.153	0.151	0.179	0.162
P	0.120	0.030	0.045	0.235	0.033	0.064
Age	-0.014	-0.020	-0.016	-0.008	-0.023	-0.020
Std. Error	0.007	0.007	0.007	0.006	0.008	0.007
P	0.058	0.010	0.015	0.241	0.003	0.008
Mean dep. var.	2.508	2.758	4.269	3.151	3.756	3.864
N	221	223	223	218	217	221

4. Discussion

The nature and extent to which the four citation based occurrences differ across individual authors is estimable in public (Google Scholar) and proprietary domains (e.g. Scopus, ISI Web of Knowledge). These metrics are used by research synthesists to develop eligibility criteria and quality weightings for the inclusion of studies (Section 1.3). They are also used

by researchers, research groups, students and policy makers to different extents, tacitly or explicitly, in the formation of judgements about primary research. There is only limited support from this study to suggest that the authors of primary research would agree with any judgements made by others about their research based on these metrics. Where the majority of citing studies suggest mistakes in your work (CIT_MIS), the median and modal respondents were *Very likely* to doubt or question their work as a consequence. To the extent to which research synthesists and others form judgements using this metric, a high level of support for these judgements could exist from authors of the original primary research. Acknowledgment of this could serve to increase the professional and public acceptability of judgements and decisions taken about research based on this metric. Where the majority of citing studies suggest alternative ways of doing things (CIT_ALT), for example different methodologies, concepts or theoretical frameworks, median and modal respondents were *Neither likely or unlikely* to doubt or question their work as a consequence. Explicit or tacit decisions taken by research synthesists and others based on this metric may therefore tend to be regarded neutrally or ambivalently by the authors of the original primary research. By contrast, respondents tended to be relatively intransigent where their research received substantially less citations compared to their other work (CIT_YOU; median and modal doubt – quite unlikely). This also applies, to a lesser degree, where their research received substantially less citations compared to work by others on the same topic (CIT_OTH; median doubt – neither likely or unlikely; modal doubt – quite unlikely). Given the tenuous relationship between these metrics – CIT_YOU and CIT_OTH - and research quality (Section 1.4), their relative inability to induce doubt is not surprising. Any use made of these metrics by research synthesists, or others, could be modified or restricted accordingly if we accept the value of acknowledging the opinions of those who produce the research outputs which others make use of formally, informally, publicly or privately.

In contrast to the four citation based metrics, it is practically impossible to chronicle and analyse how the nature and extent of adverse verbalised reactions – VER_WRK; VER_MTG - varies across large cohorts of individuals or populations of researchers. These two occurrences tend to be highly localised in terms of their experience, acknowledgment and documentation. The occurrence of both tended to induce median and modal respondents to be *Quite likely* to doubt or question their research; going further, the modal response to whether the latter would induce doubt (VER_MTG) was *Very likely*. The small group nature of these occurrences, largely absent from public record limits their potential for informing any decisions taken by wider users with regards to the validity of primary research. Using objective survey derived data (which itself is available to different degrees in the public domain), only nationality and age were found to be significant determinants of doubt in a multivariate context. The doubt averages just discussed are higher (lower) where respondents are UK nationals (nationals of other countries) and inversely related to the age of authors.

Understanding how doubt varies across authors depending upon different occurrences suggests two questions. (1) *Are instances which result in doubt the same thing as instances justifying doubt?* It may seem easy to conceive of situations where doubt is justified when faced with the articulation of, what appear to be, objective or near objective truths e.g. universally accepted errors in a dataset or errors in the implementation of a methodology. It is reasonable to hypothesise that these occurrences would be more likely to stimulate multiple doubt inducing occurrences (e.g. citing studies suggest mistakes – CIT_MIS - coupled with adverse verbal reactions – VRB_WRK and VRB_MTG) compared to situations where doubt is induced by more subjective reasoning. (2) *Are instances which do not result in doubt the same thing as instances not justifying doubt?* Non-occurrence of doubt does not mean that doubt should not occur. Non-occurrence of doubt could instead be a function of irreducible ignorance, indeterminacy or total ignorance² (Walker, 2003). More subjectively, and obviously, it could also simply be something which is withheld from the public domain – authors doubting, but only in private.

In summary, in passing judgment on primary research, users of that research have hitherto neglected when and where the original authors might agree, or otherwise, with the judgements passed against their research. Understanding and acknowledging which occurrences are most likely to make the authors of primary research doubt their own research could increase the validity of these judgements. There is significant scope beyond this exploratory study for a more nuanced investigation into whether, when and why academics doubt their own research. Implicit in this study was an operationalisation of doubt which accords with the lay understanding of this term: to doubt means to lack conviction about something, or to be uncertain about something. Doubt, therefore, is not the preserve of some disciplinary subset of academics. Indeed, the subject area in which respondents tend to publish was not a significant factor affecting doubt in this study (Section 3.3). Furthermore, even if doubt was the preserve of (ostensibly) positivistic 'truth' seeking disciplines, how do those disciplines actually differ from the non-positivistic? In a biomedical context, Ioannidis' (2005b) seminal paper suggested that the majority of published research is, in fact, false whilst paradoxically admitting that we can never be 100% sure of what the truth is. Arguably, truth is (at best) elusive, doubt is ubiquitous. Nevertheless, doubt is multifaceted beyond the crude aggregate explored in this study and may co-vary with characteristics, such as values, beyond the easily measurable objective variables captured here.

² "Reducible ignorance may be resolved by conducting further research [...] Irreducible ignorance applies when neither research nor development can provide sufficient knowledge [...] Total ignorance is the other extreme from determinism on the scale of uncertainty, which implies a deep level of uncertainty, to the extent that we do not even know that we do not know" (Walker, 2003, p. 9).

Acknowledgments

This research was funded by a Research Development Fund (RDF) grant to the author from the School of Business, Management & Economics (BMEc) at The University of Sussex. I am very grateful for feedback provided by two Editors and two anonymous reviewers from the Science Policy Research Unit Working Paper Series (SWPS) at The University of Sussex. I am similarly grateful to the following individuals at Sussex for their feedback: Ohid Yaqub, Andy Stirling, Noam Bergman, Michael Hopkins and Ben Martin.

References

Aitchison J. and Silvey S.D. (1957) The generalization of probit analysis to the case of multiple responses. *Biometrika* 44, 131–140.

Aspinall W (2010) A route to more tractable expert advice. *Nature* 463: 294-295.

Ayres I and Vars FE (2000) Determinants of citations to articles in elite law reviews. *The Journal of Legal Studies* 29(S1), 427-450.

Bamber J. L. and Aspinall W.P. (2013). An expert judgement assessment of future sea level rise from the ice sheets. *Nature Climate Change*, 424-427.

Bell S and Morse S (2008) *Sustainability Indicators: Measuring the Immeasurable? 2nd Edition*. London: Earthscan.

Bornmann L Schier H Marx W and Daniel HD (2012) What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics* 6, 11-18.

Brew A (2001) *The Nature of Research: Inquiry in Academic Contexts*. London: Routledge-Falmer.

Butler L and Visser MS (2006) Extending citation analysis to non-source items. *Scientometrics* 66, 327–343.

Callaham M Wears RL and Weber E (2002) Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Jama*, 287, 2847-2850.

Chalmers TC Celano P Sacks HS and Smith H. 1983. Bias in treatment assignment in controlled clinical trials. *New England Journal of Medicine* 309: 1358-1361.

Collins HM (1999) Tantalus and the aliens: publications, audiences and the search for gravitational waves. *Social Studies of Science* 29: 163-197.

Cooper H (1989) *Integrating Research: a Guide for Literature Reviews*, 2nd Edition. Sage: Newbury Park, California.

Cooper H and Hedges LV (2009a) Research synthesis as a scientific process. In: Cooper H Hedges LV and Valentine JC (eds) *Handbook of Research Synthesis and Meta-Analysis*. 2nd Edition. New York, NY: Russell Sage Foundation, 3-16.

Cooper H and Hedges LV (2009b) Potentials and limitations. In: Cooper H Hedges LV and Valentine JC (eds) *Handbook of Research Synthesis and Meta-Analysis*. 2nd Edition. New York, NY: Russell Sage Foundation, 561-572.

Dalkey N and Helmer O (1963) An experimental application of the Delphi method to the use of experts. *Management Science* 9: 458-467.

Fanelli D (2013) Positive results receive more citations, but only in some disciplines. *Scientometrics* 94, 701-709.

Fleck J (1998). Expertise: Knowledge, Power and Tradeability. In: Williams R Faulker W and Fleck J. (eds) *Exploring Expertise*. London: Macmillan, 143-171.

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association* 295, 90-93.

Gustafson DH Shukla RK Delbecq A and Walster GW (1973) A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. *Organizational Behavior and Human Performance* 9, 280-291.

Hopewell S, Boutron I, Altman DG, Ravaud P (2013) Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study. *British Medical Journal open* 3: e003342.

Hudson J (2007) Be known by the company you keep: citations - quality or chance? *Scientometrics* 71, 231-238.

Hunter JE and Schmidt FL. 2004. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 2nd Edition. Sage: California.

Ioannidis JP (2005a) Contradicted and initially stronger effects in highly cited clinical research. *Jama*, 294, 218-228.

Ioannidis JP (2005b). Why most published research findings are false. *PLoS Medicine* 2, e124.
Jüni P, Witschi A, Bloch R, Egger, M. 1999. The hazards of scoring the quality of clinical trials for meta-analysis. *The Journal of the American Medical Association* 282, 1054-1060.

Keller K and McInerney D (2008) The dynamics of learning about a climate threshold. *Climate Dynamics* 30, 321-332.

Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica* 46, 33-50.

Kolstad CD (1996) Learning and stock effects in environmental regulation: the case of greenhouse gas emissions. *Journal of Environmental Economics & Management* 31, 1-18.

Leimu, R and Koricheva J (2005) What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution* 20, 28-32.

MacKenzie D (1998). *The Certainty Trough* In: *Exploring Expertise*. In: Williams R Faulker W and Fleck J. (eds) *Exploring Expertise*. London: Macmillan, 325-329.

MacRoberts MH and MacRoberts BR (2010) Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology* 61, 1-12.

McCullagh P (1980) Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society B* 42, 109-142.

Moher D Pham B Jones A Cook DJ Jadad AR Moher M Tugwell P and Klassen TP. (1998) Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *The Lancet* 352, 609-613.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *The Lancet* 354: 1896-1900.

Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine* 151: 264-269.

Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati, A. 2005. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *British Medical Journal* 330: 1053.

Munier F and Ronde P (2001) The role of knowledge codification in the emergence of consensus under uncertainty: empirical analysis and policy implications. *Research Policy* 30, 1537-1551.

Nardo M Saisana M Saltelli A Tarantola S Hoffman A and Giovannini E. (2005) Handbook on constructing composite indicators. OECD Statistics Working Papers, 2005/03, OECD Publishing.

Nieminen P Carpenter J Rucker G and Schumacher M (2006) The relationship between quality of research and citation frequency. *BMC Medical Research Methodology* 6(1), 42.

Nurmohamed MT, Buller HR, Dekker E, Hommes DW, Rosendaal FR, Briet E, Rosendaal FR, Vandenbroucke JP. (1992) Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *The Lancet* 340, 152-156.

Oppenheimer M O'Neill BC and Webster M (2008) Negative learning. *Climatic Change* 89, 155-172.

Rowe G and Wright G (1999) The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting* 15, 353-375.

Saisana M, Tarantola S. 2002. State-of-the-art report on current methodologies and practices for composite indicator development, EUR 20408 EN, European Commission-JRC: Italy.

Schulz KF Chalmers I Hayes RJ and Altman DG (1995) Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *The Journal of the American Medical Association* 273, 408–12.

Sorzano C Vargas J Caffarena-Fernández G and Iriarte A (2014) Comparing scientific performance among equals. *Scientometrics* 101, 1731–1745.

Stanley TD (2001) Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives* 15, 131-150.

Stanley TD, Doucouliagos H, Giles M, Heckemeyer JH, Johnston RJ, Laroche P, Nelson JP, Paldam M, Poot J, Pugh G, Rosenberger RS, Rost K. (2013) Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys* 27, 390-394.

Stirling A (2010) Keep it complex. *Nature* 468, 1029-1031.

Walker WE Harremoës P Rotman J Van Der Sluus JP Van Asselt MBA Janssen P Kraye Von Krauss MP (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment* 4, 5-17.

Wilsdon, J., Allen, E., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. and Johnson, B. (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Higher Education Funding Council for England (HEFCE), Bristol, UK.

Wilson DB (2009) Systematic coding. In: Cooper H Hedges LV and Valentine JC (eds) *Handbook of Research Synthesis and Meta-Analysis*. 2nd Edition. New York, NY: Russell Sage Foundation, 159-176.

Woodward RT and Wui YS. 2001. The economic value of wetland services: a meta-analysis. *Ecological Economics* 37: 257-270.



PARTICIPANT INFORMATION

STUDY TITLE

Uncertainty in research: are adverse peer reactions important?

INVITATION

Members of academic staff and postgraduate students in higher education institutions anywhere in the world are invited to complete this online survey provided they have authored or co-authored at least one peer reviewed journal article. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information.

WHAT IS THE PURPOSE OF THE STUDY?

The purpose of this study is to understand the extent to which adverse peer reactions could make you doubt or question work that you have had published in peer reviewed journals.

WHY HAVE I BEEN INVITED TO PARTICIPATE?

This invitation to participate is open to academic staff and postgraduate students who work or study at higher education institutions anywhere in the world.

DO I HAVE TO TAKE PART?

It is up to you to decide whether or not to complete the survey. If you decide to take part you are free to stop completing it at any point without giving a reason and without sharing your identity.

WHAT WILL HAPPEN TO ME IF I TAKE PART?

You will simply provide answers to 12 quick online survey questions. It is estimated that the survey will only take you about 5 minutes to complete.

WHAT ARE THE POSSIBLE BENEFITS OF TAKING PART?

By taking part you could help further our understanding of when and why people might change their minds about their previously published research.

WILL MY INFORMATION IN THIS STUDY BE KEPT CONFIDENTIAL?

Confidentiality, privacy and anonymity will be ensured because at no point will you be asked to reveal your identity.

WHAT WILL HAPPEN TO THE RESULTS OF THE RESEARCH STUDY?

The anonymous answers you provide will be analysed alongside those from other respondents using various statistical techniques. This analysis of the anonymous answers you provide may be

disseminated at conferences, meetings and in the peer-reviewed literature. Copies of any written material produced using the results of this survey can be obtained from the contact address below.

WHO IS ORGANISING AND FUNDING THE RESEARCH?

The person conducting this research is a member of staff in the Science Policy Research Unit (SPRU) within the School of Business, Management & Economics (BMEc) at the University of Sussex in the U.K. This study has been made possible following the award of a research grant from BMEc to the person conducting this research.

WHO HAS APPROVED THIS STUDY?

This research has been approved by the University of Sussex' Social Sciences Cross-school Research Ethics Committee (C-REC).

CONTACT FOR FURTHER INFORMATION

Name: Dr Lee Stapleton

Address: SPRU - Science Policy Research Unit, School of Business, Management & Economics, Jubilee Building, University of Sussex, Falmer, Brighton, BN1 9SL, United Kingdom.

E-mail: l.stapleton@sussex.ac.uk.

Telephone: +44(0)1273 872781

Staff web profile: <http://www.sussex.ac.uk/profiles/219203>

If you have any concerns about the way in which the study has been conducted, you should contact Lee Stapleton in the first instance.

The University of Sussex has insurance in place to cover its legal liabilities in respect of this study.

THANK YOU!

If you complete the following survey it is assumed that you have read, understood (and where applicable) consent to the information provided above. Thank you for your time, it is greatly appreciated!

DATE

January 2015

SURVEY

1. What is your gender?

- a. Female
- b. Male
- c. Other

If you selected other, please specify

2. How old are you?

[Free form answer]

3. What is your nationality?

[Free form answer]

4. In which country do you currently work or study?

[Free form answer]

5. What is your current position?

- a. Professor
- b. Reader, Associate Professor or Senior Lecturer
- c. Senior Research Fellow or Senior Teaching Fellow
- c. Assistant Professor or Lecturer
- d. Research Fellow, Research Associate or Teaching Fellow
- e. Research Officer or Research Assistant
- f. Postgraduate student
- g. Other

If you selected other, please specify

6. Are you employed on a fixed term, open ended or permanent basis? If you are a postgraduate student skip this question.

- a. Fixed term
- b. Open ended
- c. Permanent
- c. Other

If you selected other, please specify

7. Which of the following best identifies the broad subject area(s) of your peer reviewed journal publications to date? Multiple subject areas can be selected if appropriate.

- a. Arts and humanities
- b. Engineering and technology
- c. Life sciences and medicine
- d. Natural sciences
- e. Social sciences
- f. Other

If you selected other, please specify

8. How many peer reviewed journal articles have you published to date? If you are not entirely sure, please provide a best estimate. It does not matter for the purpose of this question whether you were the lead author, sole author or a co-author.

[Free form answer]

9. In what year was your first peer reviewed journal article published? If you are not entirely sure, please provide a best estimate. It does not matter for the purpose of this question whether you were the lead author, sole author or a co-author.

[Free form answer]

10. How likely is it that each of the following would make you doubt or question work that you have had published in peer reviewed journals? You will need to select an answer on all 6 rows for the survey to be submitted.

	Very unlikely	Quite unlikely	Neither likely or unlikely	Quite likely	Very likely	Rather not say	Don't know	Other
Substantially fewer citations compared to your other work								
Substantially fewer citations compared to work by others on the same topic								

The majority of citing studies suggest mistakes in your work								
The majority of citing studies suggest alternative ways of doing things								
In the workplace, the majority of verbalised peer reactions to work you have published are adverse, not positive								
In other work contexts (e.g. at meetings or conferences), the majority of verbalised peer reactions to work you have published are adverse, not positive								

If you selected other, please specify and rate likelihood using the same scale if appropriate

11. Do you doubt or question any of the peer reviewed journal articles you have published?

Yes	No	Rather Not Say	Don't know	Other

If you selected other, please specify

12. Do you have any comments or feedback that you would like to provide?

Recent papers in the SPRU Working Paper Series:

SWPS 2015-15. Roman Jurowetzki. May 2015. Unpacking Big Systems – Natural Language Processing Meets Network Analysis. A Study of Smart Grid Development in Denmark.

SWPS 2015-16. Davide Consoli, Giovanni Marin, Alberto Marzucchi, Francesco Vona. May 2015. Do Green Jobs Differ from Non-Green Jobs in Terms of Skills and Human Capital?

SWPS 2015-17. Anders Bornhäll, Sven-Olov Daunfeldt, Niklas Rudholm. May 2015. Employment Protection Legislation and Firm Growth: Evidence from a Natural Experiment.

SWPS 2015-18. Phil Johnstone, Andy Stirling. June 2015. Comparing Nuclear Power Trajectories in Germany And the UK: From ‘Regimes’ to ‘Democracies’ in Sociotechnical Transitions and Discontinuities.

SWPS 2015-19. Maria Savona. July 2015. Global Structural Change And Value Chains In Services. A Reappraisal.

SWPS 2015-20. Javier Lopez Gonzalez, Valentina Meliciani, Maria Savona. July 2015. When Linder Meets Hirschman: Inter-Industry Linkages and Global Value Chains in Business Services.

SWPS 2015-21. J. Sylvan Katz. July 2015. What is a Complex Innovation System?

SWPS 2015-22. Friedemann Polzin, Paschen von Flotow, Laurens Klerkx. August 2015. Accelerating the Cleantech Revolution: Exploring the Financial Mobilisation Functions of Institutional Innovation Intermediaries.

SWPS 2012-23. Tommaso Ciarli, Alex Coad, Ismael Rafols. August 2015. Quantitative Analysis of Technology Futures: A review of Techniques, Uses and Characteristics.

Suggested citation:

Lee Stapleton (2015). Do Academics Doubt Their Own Research? *SPRU Working Paper Series (SWPS)*, 2015-24: 1-26. ISSN 2057-6668.

Available at www.sussex.ac.uk/spru/swps2015-24

SPRU – Science Policy Research Unit

University of Sussex

Falmer, Brighton, BN1 9SL, United Kingdom

www.sussex.ac.uk/spru

Twitter: @SPRU

SWPS: www.sussex.ac.uk/spru/research/swps