CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

There'll also be an index of terms, I think, but it's not written yet.

NON-SCIENCE STUFF

IP Intellectual Property

FLOSS free/libre open source software

GPL GNU General Public License

GNU GNUs Not Unix

OS open science

FLOSS free/libre open source software

GPL GNU General Public License, the founding document of the Free Software Movement

OS open science

OA Open Access

APC Article Processing Charge

IF Journal Impact Factor

STEM Science, Technology, Engineering, and Math

DOI Digital Object Identifier

DIY Do It Yourself

HGP Human Genome Project

Tools

WOK *Web of Knowledge*

BLAST Basic Local Alignment Search Tool

SDSS Sloan Digital Sky Survey

Organizations

FSF Free Software Foundation

CIA Central Intelligence Agency

NCBI National Center for Biotechnology Information

CERN European Organization for Nuclear Research*

HEW Department of Health, Education, and Welfare

IRB Institutional Review Board

USPHS U.S. Public Health Service

PLOS Public Library of Science

OKF Open Knowledge Foundation

OECD Organisation for Economic Co-operation and Development

TREND Teaching and Research in (Neuro)science for Development in Africa

Places

FDA Food and Drug Administration

MIT the Massachusetts Instituite of Technology

NIH National Institute of Health

NYU New York University

OEAC Other Euro-American Countries, including Australia, NZ, and Canada

ROW "Rest of World", for what that's worth

BRICS Brazil, Russia, India, China, considered the rising economic and scientific powerhouse countries

ROW "Rest of World", for what that's worth

OECD Organisation for Economic Co-operation and Development

UB FOH University of Botswana, Faculties of Humanities

UOM FOS University of Mauritius, Faculties of Science

UCT COMM University of Cape Town, Commerce

UNAM FHSS University of Namibia, Humanities and Social Sciences

*derived from the french *Conseil Europeen pour la Recherche Nuclear*

SCI Science Citation Index

WOS Web of Science

LHC Large Hadron Collider

SCIENCE

FACS Fluorescence Activated Cell Sorting

GPCR G-protein coupled receptor

G-PROTEIN guanine nucleotide-binding proteins

GRK G-protein coupled receptor kinase

RA retinoic acid

5-HT 5-hydroxytryptamine

CNS central nervous system

GDP guanosine diphosphate

GTP guanosine triphosphate

LSD lysergic acid diethylamide

MESCALINE 3,4,5-trimethoxyphenethylamine

OCD obsessive-compulsive disorder

DMT *N*,*N*-Dimethyltryptamine

DOM 2,5-Dimethoxy-4-methylamphetamine

STP Serenity, Tranquility, Peace

PFC prefrontal cortex

LC locus coeruleus

PET positron emission tomography

HTR head twitch response

DEA Drug Enforcement Administration

PPI pre pulse inhibition

MAOI monoamine oxidase inhibitor

HEKS human embryonic kidney 293 cells

GWAS Genome Wide Assocation Study

PTSD PostTraumatic Stress Disorder

ADHD Attention Deficit Hyperactivity Disorder

PPV Positive Predictive Value

SIRNA small interfering RNA

TODO LIST

1 SCIENTIFIC STRUCTURES

In the history and sociology and feminism of science, we talk about science as gift economy, or a commons of knowledge. The scientific project "builds on the shoulders of giants"; the intense, methodological, and highly-specific technoscientific answers we seek in scientific inquiry are atoms, if you will, in the molecules of theories. To stand on those shoulders, we need communication – we collaborate within buildings and between labs, we maintain social ties that influence gifts of stem cell lines and antibodies, we go out to bars at conferences, we track the ever-cresting wave of published academic research. Science is yes, the experiment, but the experiment and information gathered by others goes hand in hand.

Moreover, scientific progress is essentially meaningless unless it's communicated ("success in conveying one's ideas or in evoking understanding in others.") In the philosophical tradition of Ludwik Fleck's work, facts only emerge from a community process of discussion and assessment.[18] In a more practical sense, there's no benefit in synthesizing the mythological cure for cancer unless you (a) get community credit for your and/or (b) cancer patients actually receive the magic drug.

Science has always been tied to the communication networks of researchers and amateurs, although the altruism of that communication is certainly questionable. The 1665 establishment of the first journal, the *Philosophical Transactions of the Royal Society of London*, "aimed at creating a **public record** of original contributions to knowledge,"[24] This "original" journal was a bi-annual periodical of "letter-excerpts, reviews and summaries of recently-published books, and accounts of observations and experiments from European natural philosophers", and remained a money-losing endeavor until the middle of the 20th century.[20]

establish that scientific communications

Today, the structures of scientific communication number over 25,000 journals, along with an ever-expanding and specializing network of conferences, blogs, and tweets, to name a few. The rate at which scientific information moves within communities and between scientists and the "public" has increased dramatically, in both speed and scope. Like most other institutions no longer limited to the mail system and the printing press, the advent of Internet and the increasing role of digital technologies has a transformative effect on communication, and thus science itself.

the politics of "life itself"

Scientific communication has always been limited, in both the scope and application, and underwritten by concerns about prestige rather than progress. Academic scientific work has been constrained by funding issues; an increasingly competitive academic environment where only 1 in 10 PhD recipients will find a stable (i.e. tenured) job in their academic field ; outside pressures that lead to data fudging; an increasing amount scientific spin on what an experiment has actually accomplished and a decreasing amount of purely exploratory research.

what kind *of progress?*

research phd and jobs

The criticism of production of knowledge is reaching a fever pitch of excitement about fraud, funding, reproducibility, access, and a flock of

SCIENTIFIC ACTIVITY	COMMUNICATION REQS.
idea discovery hypothesis generation	awareness lit. review informal discussion
funding/approval	lit review
conduct research	awareness
disseminate results	formal publication informal dissemination

Table 1: communication requirements in research cycle

other issues. The scientific ecosystem has been metaphorically thrust into the punishing light of the democratizing Internet, and it turns out science isn't the foolproof method our 7th grade biology teachers said it was. And they certainly don't follow the scientific method.

1.1 CYCLES OF SCIENCE

At every step of even the canonical research cycle, communication with and from other researchers plays a role (see Table 1). It is impossible to participate in academic-scientific discourse without both intellectual and financial access to the communication. The blocks along the way are numerous and multisituated.

rewrite transition

The research cycle itself consists of 5 steps, with interpolations and cross threads between each.

1. The idea; hypothesis generation, asking the question of scientific or funder-driven interest
    - In New Orleans, an opening for an asthma grant, so my PI applied to research asthma in the context of serotonin signaling
    - an idea, late at night (Otto Loewi, discovery of acetylcholine)
    - previously established work by a lab (routine X->Y->Z)
    - within the context of a greater project by the PI (re: translational neuroscience lab?)
    - a happenstance observation; (see Charney et al., serendipity generally)
2. Set up the preliminaries of funding and approval by the relevant boards; obtain laboratory space and/or a grad student
3. Collecting data, running a set of multiply-envisioned on-the-fly experiments. "Make things work" until there's a meaningful data set; in "softer" sciences, enough evidence for an effect ($p < 0.05$) against the null hypothesis.
4. Writing and rewriting of the paper – introduction, background, results, discussions, methods, further directions for research – and submit it to a journal of your choice.

5. Wait for the inevitable rejection from the Big Journal, re-submit. Eventually, publish.

The concerns around a scientific transformation travel up along all steps of the research cycle, starting with publishing and journals access.

1.2 ACCESS TO KNOWLEDGE

> "The conventional wisdom among public health authorities is that the Ebola virus, which killed at least 10,000 people in Liberia, Sierra Leone and Guinea, was a new phenomenon, not seen in West Africa before 2013.
>
> The conventional wisdom is wrong. We were stunned recently when we stumbled across an article by European researchers in *Annals of Virology*: "The results seem to indicate that Liberia has to be included in the Ebola virus endemic zone." The paper was published in 1982.
>
> There is an adage in public health: "The road to inaction is paved with research papers."
>
> Part of the problem is that none of these articles were co-written by a Liberian scientist. The investigators collected their samples, returned home and published the startling results in European medical journals. Few Liberians were then trained in laboratory or epidemiological methods. Even today, downloading one of the papers would cost a physician here $45, about half a week's salary."
>
> Bernice Dahn, Vera Mussah, and Cameron Nutt
> "Yes, We Were Warned About Ebola", *The New York Times*

Bernice Dahn is the chief medical officer of Liberia's Ministry of Health, where Vera Mussah is the director of county health services. Cameron Nutt is the Ebola response adviser to Dr. Paul Farmer at the nonprofit group Partners in Health.

6-10 pages of double-column size 10 text in *Nature*, *Science*, or *Cell* represent the cutting edge of scientific research. One such article is "The neurobiology of psychedelic drugs", written by Vollenweider and Kometer, researchers at a Swiss lab responsible for much of the past and present literature on psychedelic drugs. It was published fiver years ago, in the September 2010 issue of *Nature Reviews Neuroscience*. Although no longer cutting edge, it is a comprehensive review bearing directly on my research. Article access costs $18; a year-long subscription to just *Nature Reviews Neuroscience* is $265.

insert data on div!

Some journals I can access through institutional subscription; *Nature Reviews Neuroscience* is not one of them (although we do have access to *Nature*, the flagship journal). In this case, because the paper is five years out of date, it is now hosted on a variety of other sites. If it was a newer article, however, or a copyright claim *Nature* pursued more vigorously, my resources would be more or less exhausted.

Interlibrary Loans are also an option, but take more time than the digital era really requires, especially for reviews

Nature *is paywalled back to 1867*

Luckily for me, I know someone who knows someone affiliated with the Massachusetts Instituite of Technology (MIT). Very illegally but in the tradition of scholarship, gave me his MIT information, enabling me to access a significantly wider array of journals than otherwise possible.

insert data that I made!

Literature access for academic scientists is typically mediated through institutional, library-based journal subscriptions. Surveys of academic researchers, largely those at high-profile and wealthy minority world institutions, report their access to the literature is "generally good and improving."[8] However, 84% of librarians, responsible for overseeing fiscal access to scholarly resources, disagreed with "There is no access problem to scientific publications in Europe."[45] For the non-affiliated researcher or member of the public with Internet access, just 35% of the peer-reviewed literature is freely available, with this figure falling to less than 30% for recently published work.[46]

rewrite and when we mean accessible, we mean the PDF version, probably in English, maybe in an OCR format, certainly figure- and mB heavy

1.2.1 *The rights to knowledge*

At it's most moral, access to knowledge – often defined as synonymous with literature – is a human right. Scientific research is conducted for and funded by the public interest. Advances in medical knowledge can be literally life-saving; if it were for any reason other than profit, denying information about health interventions to nurses and doctors would have already been condemned as a human rights violation.[48] Developments in ecology can immediately effect ecosystem management and conservation policies; materials science has implications for building and design. Specifically, in the case of healthcare workers, barriers to accessing scientific literature:

1. Prevent healthcare workers from accessing information
2. Prevent policy makers prevented from accessing info on building better systems
3. Impede research capacity and sustainable country-specific scientific development
4. Clinicians, health policy-makers, and researchers are unable to participate as equals in global science
5. subscription- based medical journals have shown so little interest in raising the profile of health problems in the developing world is that, to remain profitable, these journals are forced to publish materials that will appeal to readers who can pay.

rewrite in own words

Literature is necessary for scientific progress, and scientists working in more resource-poor areas (i.e. not the G8. Western European nations, and Israel) are still working. Papers are crucial for scientists keeping up to date with new techniques, with writing backgrounds to grants and formulating new research questions. Without very good institutional affiliation, most of the supporting scientific history is out of reach. The lack of access to literature hinders future scientific work from progressing as rapidly and efficiently as it might; it means researchers may be working with outdated tools or concepts.

find ref for fleck on journal science

rewrite intro

Yet another of the conflicts surrounding literature access are due to the changing collaboration between scientific authors and publishers in the digital age. Essentially, the production of the manuscript is done for free; or rather, on grant money, but the publisher still profits fairly handsomely. A manuscript travels through several stages.

1. Authors/labs submit their manuscript (*gratis*)
2. Publishers/editors and editors review the paper.
   a) They can either reject outright, or submit for peer review
3. The journal coordinate the peer review of said manuscript, shipping it out to 1-3 other academics in the field.
4. Academics review the paper, give feedback and suggestions and send it back to the journal (*gratis*).

   a) During the months of the process, the submitting author can do nothing else with the manuscript.

5. Revisions by the original writer, and hopefully the journal agrees to publish it.

6. Submitting lab pays a per-figure and/or per-page Article Processing Charge (APC) and signs away their Intellectual Property (IP) rights to the journal.

7. Journal takes the final work as their own and redistribute, typically under restrictive licenses and behind pay-walls.

And then the scientist pays to access similar articles

reasonably...financial incentives to publish would only further distort communication practices

The author, after *doing* all the science, receives no monetary kickback On the other hand, commercial publishers (which comprise around 50% of publishers[44]), post enormous profit margins, between 24-54%, to the tune of $1.1 billion dollars profit in 2010 for Elsevier-Reed, the largest scholarly publisher.[34] There is a clear mismatch between those doing the bulk of the effort and the bulk of the profit.

Profits for the publishers also derive not just from scientific labor, but from scientific funding. Half of all basic research in the U.S.[5]* are at least in part publicly funded. The taxpayer-funded National Institute of Health (NIH) issues sustaining multi-year R01 grants to most researchers; these grants pay for laboratory equipment, publication, and salaries of lab heads and those working under them. The results of this publicly-funded research are then made inaccessible behind pay walls – interested readers will have to pay the for-profit publishers to access the research funded by a public institution.

The NIH does require funded researchers to deposit a free version of the article after a twelve month embargo.[[NIH Public Access Policy]]

1.2.2 *The Role of Licensing*

Access is also about more than money, extending to licensing issues in the digital age.

Journal publications are a continuously and overwhelmingly increasing resource. One group estimated 50 million articles by the end of 2008[29]; a different method estimated 114 million documents on the web in 2013.The growth rate in the past decade has been between 8-9% on average; dependent on the field.While this is consistent with the increases in raw numbers of both journals and publishing authors, that doesn't mean it's any easier to keep up with. While researchers report reading a steadily increasing amount of literature[45], this precipitous increase in the number of scientific publications has led to the "impossibility of being expert"[19].

find ref for Madian Khabsa, C. Lee Giles, 2014

find ref for Bornmann

Fraser and Dunstan estimated that for a single recruit to "become expert" in the subspecialalty echocardiology, i.e. having read all the relevant papers, "would take 11 years and 124 days, by which time at least 82,142 more papers would have been added, accounting for another eight years and 78 days." To finally catch up, it would take a net total of 40 years and 295 days. And that's if the interested party could read five papers an hour (one every 10 minutes, followed by a break of 10 minutes) for eight hours a day, five days a week, and 50 weeks a year, for a capacity of 10,000 papers in one year.[19]

*And most high-profile research is done in these countries. For more on this, see **??** on page ??

The advent of machine-readable text and linked networks of scholarly information opens the possibility of managing the literature with text-mining, annotations, and new forms of communication and processing, perhaps making it easier for researchers to stay on top of recent developments. Unfortunately, the pay walled and copy-written scientific literature is both fiscally and legally out of reach for massive semantic processing; scientists cannot employ technology to efficiently manage their reading work flow.

Thus, the scientific literature remains piecemeal – single papers demonstrating single results, where over 100,000 possibly-relevant papers (granted, not all of high quality) are available and thus, perhaps, overwhelming.

1.3 MEANINGFUL RESEARCH

> "Does the pressure to publish in prestigious, high-ranking journals contribute to the unreliability of science?"
> Brembs, Button, and Munafò

Citation is the metaphorical currency of science, a way of paying homage to previous work and making sure scientists receive their proper due. It also functions, in a globalized/linked/connected world, as convertible and real currency. Citation counts now function, however, as a proxy for scientific success, first by those outside the academy, and now as a selection metric for researchers engaged in the "literature deluge".[17]

Nominally, choosing papers to read and base future work on is based on relevance and applicability of the literature. There is, of course, a deluge of literature, and it's nigh-impossible to keep up with the flow of information. One technique employed by scientific audiences in deciding where to invest their reading time is the prestige rank of a journal; similarly, when choosing which journal will be the most beneficial venue for their work, those same rankings come into play.

> "Most researchers acknowledge an intrinsic hierarchy in the scholarly journals ("journal rank") that they submit their work to, adjusting not only their submission but their reading strategies accordingly."[3]

> "The citation game has created distinct hierarchical relationships among journals in different fields."[49]

Qualitatively, the top of the biology journal hierarchy is the *Cell*, *Nature*, and *Science* triumvirate; journals that are instantaneously recognizable and eminently reputable. To get a *Cell* paper is to be immediately taken a little more seriously. †

Presumably physics and plants as well, but it seems like Nature *mostly does bio? I have literally no idea*

The Impact Factor Quantitatively, the journal hierarchy is represented by the Journal Impact Factor (IF), calculated for journals by publishing house? mass-media company Thomson-Reuters, specifically their *Web of Knowledge* (WoK) citation network. The IF was originally proposed as one metric of many to track scientific productivity: a simple mathematical formula reflecting the number of citations of a journal's

†Sources are: my life, everybody's life, a lot of blog posts, general atmosphere. Like, if I see a *Nature* headliner, I'm more likely to be excited and impressed.

material divided by the number of citable materials published by that same journal.‡[30]

> "The original intention for the use of the impact factor was to allow comparison between the citation rates of journals... This has proven invaluable for researchers and librarians in the selection and management of journals."[30]

$$\text{Impact Factor} = \frac{\text{number of citations}}{\#\,\text{citable materials published}} \tag{1.1}$$

That's all well and good, but like with many metrics, it's applied with a widening and undiscriminating brush. IFs have evolved from one metrics of citation rates to one approximating journal quality overall, on the premise that a higher citation rate of papers indicates higher quality papers.[30,40]From there, the journal IF sometimes serves as a marker of quality on individual papers and researchers.[39]Eugene Garfield, the creator of the IF of considers these applications an abuse of a simple equation. it nonetheless has serious consequences for the scientific ecosystem of research, hiring, grantsmanship, and publishing.

Metrics immediately lead to gaming the system.

find ref for David adam, ciation analysis

Journals Journal editors and publishing administrators shape what we see and pay attention to; peer reviewers have to maintain impartiality when reviewing the make-or-break publication of their competitors. Since journal reputations depend on the popularity of their article, editors make estimates of likely citations for submitted articles to gauge their interest in publication. Thus, higher impact journals select for sexy, exciting, and likely-to-be-cited literature.[49]

Who thought this would go well?

The IF depends on what article types are deemed citable – the fewer, the better (i.e. lower denominator, higher impact). Since reviews are included, this means highly-cited reviews or just a few very highly cited publications can skew impact factorsThe PLoS Medicine Editors 2006[40]

Scientists, especially high-profile and competitive ones, choose carefully which journals they'll submit to. It's a game of saying is this research trendy *and* of high enough quality *and* an original idea *enough* to make it in this high-impact journal or another? A manuscript submission takes months and in the intervening time, the manuscript can't be sent out anywhere else. This means choosing a too high impact journal is a loss of months of publication time, not to mention that while you're waiting on a decision, somebody might publish similar results firsts; but, publishing in a less-cited journal can have serious consequences on tenure decisions, grant applications, and other administrative gambols.§ It also sets up a choice: should one publish in a high-profile, non-specific journal or a lower-profile, but more relevant sub-disciplinary journals?

find ref for IF in scientific assessment

Generally, impact factor trumps audience: while a field-specific journal might make your research more visible to people who could use it, it won't make the same immediate impression as a *Cell* publication.

find ref for decisions about choosing a journal

1.4 REPRODUCIBILITY

The principle of the elusive scientific method is reproducibility. Researchers document their methods and results to such an extent that

‡insert time (2 years) and other data

§There's a number of sources denying that IFs are *specifically* counted in any of these, especially in the U.S. But they're certainly powerful tokens in the scientific imaginary, from which reviewers of any kind are hardly exempt.

any other researcher is able to replicate their data independently, or, more ideally, build experiments atop the results of their colleagues. With reproducibility, two different labs should produce similar types of results (e.g. protein-protein interactions) with different methods (e.g. protein inhibitors, mutagenesis studies).

Like the access crisis, however, there is a perception of a reproducibility crisis in science. Certainly the most high-profile beginning to the discussion was a 2005 article authored by John Ioannidis, provocatively titled *Why Most Published Research Findings Are False*. The paper proposes a statistical argument that the likely false positive rate in the published bio-medical literature clocks in at >50%.[26] The responses were many and furious,[22,27]but in the end most contributors concluded, based purely on statistical considerations, one can expect a percentage of the medical literature to be a false positive (50% according to Ioannidis' calculations, 14% in Jager and Leek's calculations).[27]

find ref for more citations for battle post-Ionnidis publication

More than just statistical theorizing, researchers regularly encounter similar hurdles. In a survey, ~50% of respondents responded that they had difficulty reproducing at least one set of published results[33] One academic lab repeated the exact same genome-scale small interfering RNA (siRNA) gene screening 5 months apart. Reproducibility ranged from 39% to 49% in terms of target hits; this is due in part to the stochastic nature of biology, but also partly to the variety of ways two different data sets responded to the same methodologies.[1]

Biotech and pharmaceutical companies trying to monetize those discoveries also demonstrate the practical scope of the problem. The rate of effective translation from basic research into clinical drug treatments has always been low, but the increasing costs of drug development and renewed focus on research reliability has prompted drug companies to join the conversation with data. The in-house target validation studies run by most bio-medical companies provide a unique data set on the reproducibility of bio-medical, cancer-focused research. Amgen's researchers attempted to reproduce the results of 53 high-profile 'landmark' cancer studies over 10 years; results could only be recapitulated in 6 of the cases, an 11% success rate.[2] Researchers at Bayer tracked the fate of 67 validation projects; in 2/3 of the cases, the validation data was so inconsistent with the published literature that the project was significantly delayed, or more commonly, entirely terminated.[35]. At Bayer, this did not correlate with fields, experimental conditions, model systems, or journal impact – it's just that most research could not be reproduced at industrial labs.

lol at pharma and cultures of molecules/reductionsm

with regard to increasing Food and Drug Administration (FDA) restrictions

To be fair, most researchers know how, well, temperamental science can be. Scientific work draws on the "tacit knowledge" of researchers, specific to times and places, and failed replicaiton is not usually due to malicious case of fraud.[21] In one case, a Berkeley, CA lab and Boston, MA lab, collaborating on a Fluorescence Activated Cell Sorting (FACS) project, discovered that even with identical protocols they had consistently different results. In a laborious year-long process, the two labs (or rather, the primary researchers in both labs) isolated the difference in their experimental results to the speed of agitation at an early step of the organ isolation process. While the results initially *seemed* irreproducible and incommensurable, their end conclusion was principally that biological research is at such a level of complexity that even minor changes from lab to lab in the micro-environment can "break" an experiment.[25] It's not completeley clear when unreplicated

This gets back to local biologies? But in the sense of localized non-controllable environments turns to unreplicable; as one researcher put it, "most failures to replicate exhibit incompetence."[21]

Thus, the "crisis" in reproducibility may not be new to science, and may be in many ways attributable to both chance and social pressures.

rewrite conclusion to reproducibility issues

1.5 STATISTICAL FAILURES: UNDERPOWERED AND BADLY ANALYZED DATA SETS

Since researchers must publish to succeed, and statistical significant and clean results are more likely to be published, researchers have incentives to "engage in research practices that make their findings publishable quickly".[4]

One of these is low statistically powered studies – studies with low sample size, small effects, or both – increase the likelihood that a nominally statistically significant finding reflects a true effect.

Three main problems contribute to unreliable findings in low-powered studies, even when all other practices are ideal:

1. low probability of finding true effects
   a) low power: the chance of discovering genuinely true effects is low
   b) a "power of 20%" = if there are 100 genuine non-null effects to be discovered in that field, these studies are expected to discover only 20 of them
2. low Positive Predictive Value (PPV)
   a) probability that a 'positive' research result reflects a true positive; depends on prior probability of it being true, the statistical power, and level of significance
   b) the lower the power, the lower the PPV
3. exaggerated estimate of magnitude

   paralleled by the decline effect

   a) "winner's curse" - likely when claims of discovery are based on statistical significance
   b) worst for small, low-powered studies, which only detect effects which happen to be large
   c) "If, for example, the true effect is medium-sized, only those small studies that, by chance, overestimate the magnitude of the effect will pass the threshold for discovery."
   d) by performing replication, we eventually bring it down to the *actual* odds ratio; if we raise the power of studies, which is...unusual

low power and other biases

1. low-powered studies are more likely to provide a wide range of estimates
2. publication bias, selective data analysis and selective reporting of outcomes are more likely to affect low-powered studies

a) A ‘negative’ result in a high-powered study can- not be explained away as being due to low power 19,20 , and thus reviewers and editors may be more willing to pub- lish it, whereas they more easily reject a small ‘negative’ study as being inconclusive or uninformative

3. may be of lower quality in other aspects

> Our results indicate that the median statistical power in neuroscience is 21%. We also applied a test for an excess of statistical significance. . . The test revealed that the actual number (349) of nominally significant studies in our analysis was significantly higher than the number expected (254; $p < 0.0001$).

> **what to do** f the intended analyses produce null findings and you move on to explore your data in other ways, say so. Null findings locked in file drawers bias the literature, whereas exploratory analyses are only useful and valid if you acknowledge the caveats and limitations.

>> Pre-register your study protocol and analysis plan Preregistration clarifies whether analyses are confirmatory or exploratory, encourages well-powered studies and reduces opportunities for non-transparent data mining and selective reporting. Various mechanisms for this exist (for example, the Open Science Framework).

>> **Make study materials and data available** Making research materials available will improve the quality of studies aimed at replicating and extending research findings. Making raw data available will enhance opportunities for data

Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition (originally *Voodoo Correlations in Social Neuroscience*) – Vul et al. [43])

*Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Result*s – Wicherts et al. [47])

1.5.1 *Excess Significance*

> Publication bias occurs when results of published studies are systematically different from results of unpublished studies. . . Empirical research consistently suggests that published work is more likely to be positive or statistically significant (P,0.05) than unpublished research.

9,11

What happens when something doesn’t work? When do we keep pushing, and how do we say “*This* disproves our hypothesis” and that does not? An experiment that is *not* working which *should* work (with regards to the current literature) lays its blame on reagants, on the technical skill of those involved, on the time of year.[¶] In many ways, a negative results presents no difference; and without the contrast between background and signal, the signal gets lost.

[¶]Of course, the time of year *is* a meaningful factor, re: Otto Loewi and acetylcholine in frogs

A 1991 *Lancet* issue is typically cited as the first publication to directly point at the issue. Easterbrook, Gopalan, Berlin, and Matthews reviewed a set of clinical research trials, concluding:[10]

> "Studies with statistically significant results were more likely to be published than those finding no difference between the study groups... Studies with significant results were also more likely to lead to a greater number of publications and presentations and to be published in journals with a high citation impact factor. An increased likelihood of publication was also associated with a high rating by the investigator of the importance of the study results, and with increasing sample size."

The tendency is usually referred to as the "positive publication bias", but it's might be more accurately described as bias towards *interesting* results: a high-profile study refuting another high-profile study in a positive-negative loop is just as exciting as the first positive results were.[32]

Between 1990-2007, the proportion of papers with positive results has steadily increased, starting at 70.2% in 1990 and by 2007, ending at 85.9% in 2007.[14] The positive publication bias is not limited to clinical trials or the eternal scapegoat of psychology, although it clusters there. "Harder" sciences, specifically ecology[28], and animal studies[38,41], as well as the physical sciences[14] show a similar (if decreasingly prevalent) bias.[13]

find ref for miller2014, understanding publicaiton bias in re-introduction biology

Papers with negative (non-significant) results are less likely to be published and receive, overall, less citations.[12] This is in large part due to scientific reticence, with authors unwilling to put in the necessary and extended effort in attempting to publish a paper with negative results,instead focusing on "wonderful" results.While journal editors show no bias in their acceptance of negative versus positive results papers[?] , this is perhaps more attributable to the fact that submitted negative papers tend to be of higher quality.Peer reviewers are more likely to recommend a positive-results paper be published, award positive-results papers better methodological scores, and are more critical of and detect more errors in papers with non-significant results.[11,36] As measured by citation rates, readers are less interested in negative results papers, and for-profit funders of larger-scale trials have a vested interest in specific (positive) results, as in the case of industry-funded nicotine and cigarette trials in the 1990s.

find ref for song et al papers on publicatino bias

find ref for calnan 2006

find ref for all citations from Senn2012, misunderstanding publication bias

find ref for senn2012, misunderstanding publication bias

File Drawer Effect Many studies in a given research area may be conducted, but are never reported because of the publication bias. This creates a set of journal articles wholly unrepresentative of the actual conclusions of conducted research. In the extreme case, if the null hypothesis is "true" (i.e. the relationship being studied does not exist), but the 5% of studies that by *chance* show statistical significance are published, the rest of abandoned data stays hidden in a researcher's file drawer.[37] This biases the state of the field, and in Ioannidis's words, "claimed research findings may often be simply accurate measures of the prevailing bias".[26] More realistically, this has serious consequences for synthesizing the results of a given field, leading to an overestimation of effect sizes[4]and the continued wasted efforts based on papers that failed to replicate, but never had the failed replication published.

find ref for AnneSophie 2013, citation bias favoring statistically significant studies

find ref for Turner and Spilich, 1997

find ref for levine et al 2009; jennings and van horn 2012

Other forms of dissemination

In addition to journal papers, study results may be presented at relevant conferences, submitted to medicine regulatory authorities, and in clinical study reports to trial sponsors. 3 Results presented at conferences are usually only available in abstract form. Clinical study reports contain much more detailed information on methodology and a wider selection of outcomes – fuller results. Data submitted to regulatory authorities are also important sources of trial results. However, clinical study reports and data submit- ted to regulatory authorities have limited accessibility, and their use for making evidence-based decisions may not be as straightforward as results published in peer-reviewed journals. Therefore, results of trials that are not published in peer-reviewed journals are conventionally considered as being "unpublished."

rewrite other forms of gray publishing and introduced bias

find ref for Song2013 - publication bias: what is it?

Whatever the reason the *de facto* dissemination system for scholarly production, is not representative of the research that's actually happening, which naturally has consequences for the theories we test going forwards.

1.6 RETRACTION

On the other hand, positive results are sometimes not as positive as they should be. Processes of peer-review and editorial jurisdiction are supposed to ensure that post-review articles are scientifically sound. The retraction of those articles represents an enormous reputation cost to both journal and author; retraction often leads to firing (or mandatory stepping down) of the editor, and in the longer-term, a drop in citation rates to the author's prior work.[31] [||] Retraction is not always bad (for reasons discussed momentarily), but it is notable. In 2008, Cokol, Ozbay, and Rodriguez-Esteban concluded that

12.5% over five years post retraction

> "...in the period between 1990 and 2006, we found a significant increase (r = 0.55, p = 0.02) [in retractions]...retraction rates are still on the rise.
>
> Cokol et al.[7]

Not only is the overall rate of retractions increasing, but they're increasing in the journals we respect the most. In a 2011 *Infection and Immunity* publication, Fang and Casadevall found a "strikingly robust" correlation between a journal's "retraction index" and its impact factor (Figure 1). [16] This is, of course, concerning for those of

Some researchers propose the pressure to publish, increasingly important to funding and hiring decisions, is enough of push for scientists to produce flawed manuscripts at a higher rate.[12] The other interpretation is merely that flawed manuscripts are being identified more successfully, in which case science is perhaps fulfilling its self-correcting ideals of awareness and responsiveness to false effects.[7,15]

Grieneisen and Zhang examined not just numbers of retractions, but *why* articles were being retracted. Of the 4,449 retracted articles found in 42 of the largest bibliographic databases from 1928-2011, the reasons behind retraction were usually relatively benign.

rewrite transition here

clarify these definitions of retractions

47% alleged publishing misconduct

[||] This is only in cases of non-self-reported retractions; if an author voluntarily steps forward, their citation rate typically briefly increases

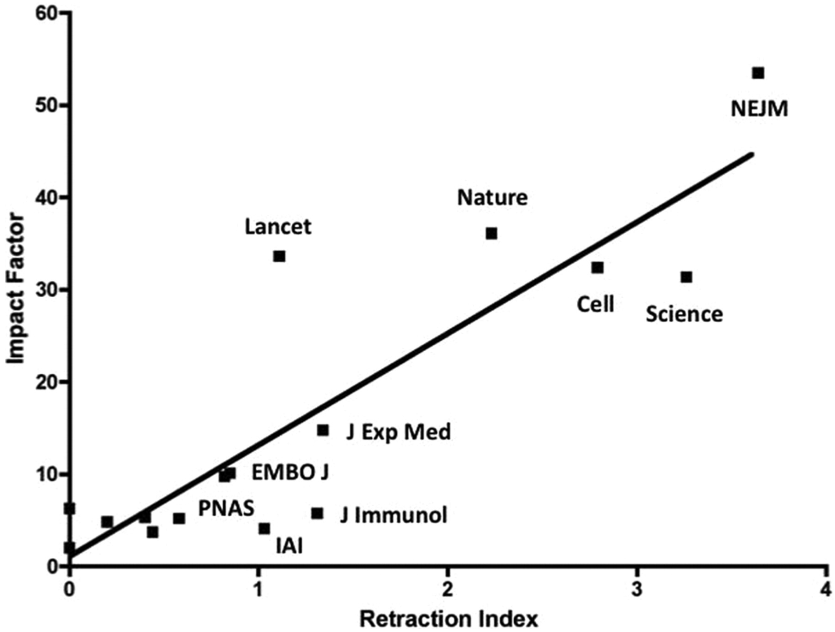

Figure 1: "Correlation between impact factor and retraction index. The 2010 journal impact factor...is plotted against the retraction index as a measure of the frequency of retracted articles from 2001 to 2010."[16]

20% alleged research misconduct

42% the usage of questionable data or interpretations

stack to >100% because more than one reason is often cited

Of the alleged research misconduct, fifteen individuals accounted for >50% of the retractions. While even inGrieneisen and Zhang's data there is a growth in retractions by a factor of 11.36 (excluding repeat offenders, adjusting for literature growth), they assert that research misconduct of the fraud variety underlie most retractions; i.e. the retraction rate is likely due to social factors, not changes in rigor or intentional deception.[23]

retractions conclusion; also include section on fraud, specifically, re: fanelli 2009

1.6.1 *The Remains of Retractions*

An additional problem remains; once an article has been retracted, that doesn't mean it will cease being cited. In the case of highly controversial papers**, retractions are usually due to high-profile fraud. In less obvious cases, however, retracted papers continue to be cited long after their nominal removal.

find ref for longevity of retracted papers

1.7 SUMMARY

So the *validity* of scientific communications is in questions, largely due to the unit of published literature and the pressures on scientists to create certain specific kinds of knowledge. There is issues of access to the literature; but even that literature does not fill all the knowledge gaps. It is subject to funding whims, human biases in selection criteria, concerns about prestige, often statistical bunk, occasionally outright

**STAP, everybody!

fraud that isn't always caught on. Publishing is the outlet for most of these, but the problems run much deeper, into the very beginnings of scientific projects. What kinds of systems could re-create a scientific ecosystem that was not subject to these things, or at least remained flexible enough to continually re-prioritize the accuracy of scientific work and the, ahem, pursuit of truth?

[1] Nicholas J. Barrows, Caroline Le Sommer, Mariano A. Garcia-Blanco, and James L. Pearson. Factors affecting reproducibility between genome-scale siRNA-based screens. *Journal of Biomolecular Screening*, 15(7):735–747, August 2010. ISSN 1087-0571, 1552-454X. doi: 10.1177/1087057110374994. URL http://jbx.sagepub.com/content/15/7/735. (Cited on page 13.)

[2] C. Glenn Begley and Lee M. Ellis. Drug development: Raise standards for preclinical cancer research. *Nature*, 483 (7391):531–533, March 2012. ISSN 0028-0836. doi: 10.1038/483531a. URL http://www.nature.com/nature/journal/v483/n7391/full/483531a.html. (Cited on page 13.)

[3] Björn Brembs, Katherine Button, and Marcus Munafò. Deep impact: unintended consequences of journal rank. *Frontiers in Human Neuroscience*, 7, 2013. ISSN 1662-5161. doi: 10.3389/fnhum.2013.00291. URL http://journal.frontiersin.org/Journal/10.3389/fnhum.2013.00291/full. (Cited on page 11.)

[4] Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, April 2013. ISSN 1471-003X, 1471-0048. doi: 10.1038/nrn3475. URL http://www.nature.com/doifinder/10.1038/nrn3475. (Cited on pages 14 and 16.)

[5] Sheila Marie Campbell. Federal support for research and development. Congress of the US, Congressional Budget Office, 2007. URL http://books.google.com/books?hl=en&lr=&id=D_Lqv-i2qwgC&oi=fnd&pg=PR7&dq=%22the+Federal+Government%E2%80%99s+Role+in+Research%22+%22Federal+Outlays+for+Research+and%22+%22and+Development+as+a+Percentage%22+%22of+Defense+and+Other+Federal%22+%22International+Research+and+Development%22+&ots=SLKWrc3qdS&sig=uNBfHOh6biF-IKd3FtJ3jmzu1so. (Cited on page 10.)

[6] Noah D. Charney, John J. Castorino, Megan J. Dobro, and Sarah L. Steely. Embryo development inside female salamander (ambystoma jeffersonianum-laterale) prior to egg laying. *PLoS ONE*, 9(3):e91919, March 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0091919. URL http://dx.plos.org/10.1371/journal.pone.0091919. (Cited on page 7.)

[7] Murat Cokol, Fatih Ozbay, and Raul Rodriguez-Esteban. Retraction rates are on the rise. *EMBO Reports*, 9(1):2, January 2008. ISSN 1469-221X. doi: 10.1038/sj.embor.7401143. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2246630/. (Cited on page 17.)

[8] Philip M Davis and William H Walters. The impact of free access to the scientific literature: a review of recent research. *Journal of the*

Medical Library Association : JMLA, 99(3):208–217, July 2011. ISSN 1536-5050. doi: 10.3163/1536-5050.99.3.008. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3133904/. (Cited on page 8.)

[9] Kerry Dwan, Douglas G. Altman, Juan A. Arnaiz, Jill Bloom, An-Wen Chan, Eugenia Cronin, Evelyne Decullier, Philippa J. Easterbrook, Erik Von Elm, Carrol Gamble, Davina Ghersi, John P. A. Ioannidis, John Simes, and Paula R. Williamson. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE*, 3(8):e3081, August 2008. doi: 10.1371/journal.pone.0003081. URL http://dx.doi.org/10.1371/journal.pone.0003081. (Cited on page 15.)

[10] Phillipa J. Easterbrook, R Gopalan, Jesse A. Berlin, and David R. Matthews. Publication bias in clinical research. *The Lancet*, 337(8746):867–872, April 1991. ISSN 0140-6736. doi: 10.1016/0140-6736(91)90201-Y. URL http://www.sciencedirect.com/science/article/pii/014067369190201Y. (Cited on page 16.)

[11] Gwendolyn B. Emerson, Winston J. Warme, Fredric M. Wolf, James D. Heckman, Richard A. Brand, and Seth S. Leopold. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Archives of internal medicine*, 170 (21):1934–1939, 2010. URL http://archpsyc.jamanetwork.com/article.aspx?articleid=226270. (Cited on pages 15 and 16.)

[12] Daniele Fanelli. Do pressures to publish increase scientists' bias? an empirical support from US states data. *PloS one*, 5(4):e10271, 2010. URL http://dx.plos.org/10.1371/journal.pone.0010271. (Cited on pages 16 and 17.)

[13] Daniele Fanelli. "positive" results increase down the hierarchy of the sciences. *PLoS ONE*, 5(4):e10068, April 2010. doi: 10.1371/journal.pone.0010068. URL http://dx.doi.org/10.1371/journal.pone.0010068. (Cited on page 16.)

[14] Daniele Fanelli. Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3):891–904, March 2012. ISSN 0138-9130, 1588-2861. doi: 10.1007/s11192-011-0494-7. URL http://link.springer.com/article/10.1007/s11192-011-0494-7. (Cited on page 16.)

[15] Daniele Fanelli. Why growing retractions are (mostly) a good sign. *PLoS medicine*, 10(12):e1001563, 2013. URL http://dx.plos.org/10.1371/journal.pmed.1001563. (Cited on page 17.)

[16] Ferric C. Fang and Arturo Casadevall. Retracted science and the retraction index. *Infection and Immunity*, 79(10):3855–3859, October 2011. ISSN 0019-9567, 1098-5522. doi: 10.1128/IAI.05661-11. URL http://iai.asm.org/content/79/10/3855. (Cited on pages 2, 17, and 18.)

[17] Alessandro Figà-Talamanca. Strengths and weaknesses of citation indices and impact factors. In Alessandro Cavalli, editor, *Quality assessment in higher education*, pages 83–88. Portland Press, London, 2007. ISBN 9781855781719. URL http://vu.portlandpress.com/pp/books/online/QAHEE/001/0083/0010083.pdf. chapter: 8. (Cited on page 11.)

[18] Ludwik Fleck. *Genesis and Development of a Scientific Fact*. University of Chicago Press, Chicago, August 1981. ISBN 9780226253251. (Cited on page 6.)

[19] Alan G. Fraser and Frank D. Dunstan. On the impossibility of being expert. *BMJ*, 341:c6815, December 2010. ISSN 0959-8138, 1468-5833. doi: 10.1136/bmj.c6815. URL http://www.bmj.com/content/341/bmj.c6815. (Cited on page 10.)

[20] Aileen Fyfe, Julie McDougall-Waters, and Noah Joseph Moxham. Philosophical transactions: 350 years of publishing at the royal society (1665–2015). Technical report, The Royal Society, London, 2014. URL http://research-repository.st-andrews.ac.uk/handle/10023/6058. (Cited on page 6.)

[21] Jim Giles. The trouble with replication. *Nature*, 442(7101): 344–347, July 2006. ISSN 0028-0836. doi: 10.1038/442344a. URL http://www.nature.com.silk.library.umass.edu/nature/journal/v442/n7101/full/442344a.html. (Cited on pages 13 and 14.)

[22] Steven Goodman and Sander Greenland. Assessing the unreliability of the medical literature: a response to "why most published research findings are false. *Johns Hopkins University, Dept. of Biostatistics Working Papers*, February 2007. URL http://biostats.bepress.com/jhubiostat/paper135. (Cited on page 13.)

[23] Michael L. Grieneisen and Minghua Zhang. A comprehensive survey of retracted articles from the scholarly literature. *PLoS One*, 7(10):e44118, 2012. URL http://dx.plos.org/10.1371/journal.pone.0044118. (Cited on pages 17 and 18.)

[24] Jean-Claude Guédon. *In Oldenburg's long shadow: librarians, research scientists, publishers, and the control of scientific publishing*. Association of Research Libraries, Washington, D.C., 2001. ISBN 0918006813 9780918006813. (Cited on page 6.)

[25] William C. Hines, Ying Su, Irene Kuhn, Kornelia Polyak, and Mina J. Bissell. Sorting out the FACS: A devil in the details. *Cell Reports*, 6(5):779–781, March 2014. ISSN 2211-1247. doi: 10.1016/j.celrep.2014.02.021. URL http://www.cell.com/article/S2211124714001211/abstract. (Cited on page 13.)

[26] John P. A. Ioannidis. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, 294(2):218–228, July 2005. ISSN 1538-3598. doi: 10.1001/jama.294.2.218. (Cited on pages 13 and 16.)

[27] Leah R. Jager and Jeffrey T. Leek. An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15(1):1–12, January 2014. ISSN 1465-4644, 1468-4357. doi: 10.1093/biostatistics/kxt007. URL http://biostatistics.oxfordjournals.org/content/15/1/1. (Cited on page 13.)

[28] Michael D. Jennions and ANDERS P. MOeLLER. Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill'method. *Biological Reviews of the Cambridge Philosophical*

Society, 77(02):211–222, 2002. URL http://journals.cambridge.org/abstract_S1464793101005875. (Cited on page 16.)

[29] Arif E. Jinha. Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publishing*, 23(3):258–263, July 2010. ISSN 09531513, 17414857. doi: 10.1087/20100308. URL http://openurl.ingenta.com/content/xref?genre=article&issn=0953-1513&volume=23&issue=3&spage=258. (Cited on page 10.)

[30] Andrew P. Kurmis. Understanding the limitations of the journal impact factor. *The Journal of Bone & Joint Surgery*, 85(12): 2449–2454, December 2003. ISSN 0021-9355, 1535-1386. URL http://jbjs.org/content/85/12/2449. The impact factor, a simple mathematical formula reflecting the number of citations of a journal's material divided by the number of citable materials published by that same journal, has evolved to become one of the most influential tools in modern research and academia. The impact factor can be influenced and biased (intentionally or otherwise) by many factors. Extension of the impact factor to the assessment of journal quality or individual authors is inappropriate. Extension of the impact factor to cross-discipline journal comparison is also inappropriate. Those who choose to use the impact factor as a comparative tool should be aware of the nature and premise of its derivation and also of its inherent flaws and practical limitations. (Cited on page 12.)

[31] Susan Feng Lu, Ginger Zhe Jin, Brian Uzzi, and Benjamin Jones. The retraction penalty: Evidence from the web of science. *Scientific Reports*, 3, November 2013. doi: 10.1038/srep03146. URL http://www.nature.com/srep/2013/131106/srep03146/full/srep03146.html. (Cited on page 17.)

[32] Hendrika J. Luijendijk and Xander Koolman. The incentive to publish negative studies: how beta-blockers and depression got stuck in the publication cycle. *Journal of Clinical Epidemiology*, 65(5):488–492, May 2012. ISSN 0895-4356. doi: 10.1016/j.jclinepi.2011.06.022. URL http://www.jclinepi.com/article/S0895435611003258/abstract. (Cited on page 16.)

[33] Aaron Mobley, Suzanne K. Linder, Russell Braeuer, Lee M. Ellis, and Leonard Zwelling. A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. *PLoS ONE*, 8(5):e63221, May 2013. doi: 10.1371/journal.pone.0063221. URL http://dx.doi.org/10.1371/journal.pone.0063221. (Cited on page 13.)

[34] Heather Grace Morrison. *Freedom for scholarship in the Internet age*. PhD thesis, Communication, Art & Technology: School of Communication, 2012. URL http://summit.sfu.ca/item/12537. (Cited on page 10.)

[35] Florian Prinz, Thomas Schlange, and Khusru Asadullah. Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9):712–712, September 2011. ISSN 1474-1776. doi: 10.1038/nrd3439-c1. URL http://www.nature.com/nrd/journal/v10/n9/full/nrd3439-c1.html. (Cited on page 13.)

[36] Sara Rockwell, Bruce F. Kimler, and John E. Moulder. Publishing negative results: The problem of publication bias. *Radiation Research*, 165(6):623–625, June 2006. ISSN 0033-7587. doi: 10.1667/RR3573.1. URL http://www.bioone.org.libproxy.mit.edu/doi/full/10.1667/RR3573.1. (Cited on page 16.)

[37] Jeffrey D. Scargle. Publication bias: The "file-drawer problem" in scientific inference. *The Journal of Scientific Exploration*, pages 91–106, 2000. (Cited on page 16.)

[38] Emily S. Sena, H. Bart van der Worp, Philip M. W. Bath, David W. Howells, and Malcolm R. Macleod. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol*, 8(3):e1000344, March 2010. doi: 10.1371/journal.pbio.1000344. URL http://dx.doi.org/10.1371/journal.pbio.1000344. (Cited on page 16.)

[39] Richard Smith. Commentary: The power of the unrelenting impact factor—is it a force for good or harm? *International Journal of Epidemiology*, 35(5):1129–1130, October 2006. ISSN 0300-5771, 1464-3685. doi: 10.1093/ije/dyl191. URL http://ije.oxfordjournals.org/content/35/5/1129. (Cited on page 12.)

[40] The PLoS Medicine Editors. The impact factor game. *PLoS Med*, 3(6):e291, June 2006. doi: 10.1371/journal.pmed.0030291. URL http://dx.doi.org/10.1371/journal.pmed.0030291. (Cited on page 12.)

[41] Konstantinos K. Tsilidis, Orestis A. Panagiotou, Emily S. Sena, Eleni Aretouli, Evangelos Evangelou, David W. Howells, Rustam Al-Shahi Salman, Malcolm R. Macleod, and John P. A. Ioannidis. Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS Biol*, 11(7):e1001609, July 2013. doi: 10.1371/journal.pbio.1001609. URL http://dx.doi.org/10.1371/journal.pbio.1001609. (Cited on page 16.)

[42] Franz X. Vollenweider and Michael Kometer. The neurobiology of psychedelic drugs: implications for the treatment of mood disorders. *Nature Reviews Neuroscience*, 11(9):642–651, 2010. URL http://www.nature.com/nrn/journal/v11/n9/abs/nrn2884.html. (Cited on page 8.)

[43] Edward Vul, Christine Harris, Piotr Winkielman, and Harold Pashler. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4(3):274–290, May 2009. ISSN 1745-6916, 1745-6924. doi: 10.1111/j.1745-6924.2009.01125.x. URL http://pps.sagepub.com/content/4/3/274. (Cited on page 15.)

[44] Mark Ware and Michael Mabe. The STM report: An overview of scientific and scholarly journal publishing. Technical report, International Association of Scientific, Technical and Medical Publishers, 2012. URL http://connect.stm-assoc.org/2012_12_11_STM_Report_2012.pdf. (Cited on page 10.)

[45] Mark Ware and Michael Mabe. The STM report: An overview of scientific and scholarly journal publishing. Technical Report 4th edition, International Association of Scientific, Technical and

Medical Publishers, Prins Willem Alexanderhof 5, The Hague, 2595BE, The Netherlands, March 2015. (Cited on pages 8 and 10.)

[46] Bruce White. Total availability of journal articles to internet users. *Library Review*, 63(4/5):295–304, July 2014. ISSN 0024-2535. doi: 10.1108/LR-01-2014-0006. URL `http://www.emeraldinsight.com/doi/abs/10.1108/LR-01-2014-0006`. (Cited on page 9.)

[47] Jelte M. Wicherts, Marjan Bakker, and Dylan Molenaar. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, 6 (11):e26828, November 2011. doi: 10.1371/journal.pone.0026828. URL `http://dx.doi.org/10.1371/journal.pone.0026828`. (Cited on page 15.)

[48] Gavin Yamey. Excluding the poor from accessing biomedical literature: a rights violation that impedes global health. *health and human rights*, pages 21–42, 2008. URL `http://www.jstor.org/stable/20460085`. (Cited on page 9.)

[49] Neal S Young, John P. A Ioannidis, and Omar Al-Ubaydli. Why current publication practices may distort science. *PLoS Med*, 5 (10):e201, October 2008. doi: 10.1371/journal.pmed.0050201. URL `http://dx.doi.org/10.1371/journal.pmed.0050201`. (Cited on pages 11 and 12.)