Taking stock of the credibility revolution: Scientific reform 2011-now

Coordinated by Gilad Feldman (giladfel@gmail.com) at University of Hong Kong psychology department.

This is an open collaborative book. If you wish to help develop this book, visit: https://tinyurl.com/replicationcrisissummary2019 or contact the coordinator.

Team

The book was written by the PSYC2020 Fundamentals of Social Psychology course undergraduate students. Each of the chapters details contributors for that chapter.

Coordinator:

- Gilad Feldman giladfel@gmail.com, University of Hong Kong, http://giladfeldman.org, @giladfeldman

Advisors:

- Boley Cheng boleyc@hku.hk
- Myrte Louise Druyvesteyn mldruyv@hku.hk
- Bill Jiaxin Shi billshi@connect.hku.hk
- Mira Ziqing Yao ziqingy@connect.hku.hk

Reviewers:

- Yeung Siu Kit (Wilson) Email: u3517520@connect.hku.hk
- Xiao Qinyu Email: xqy1020@connect.hku.hk

Contents

1 - Realization of a problem

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Aluel	Ayok-Loewenberg	https://www.researchgate.net/profile/Aluel_Ayok-Loewenberg	osf.io/ef34c	u3565061@connect.hku.hk	aluelayok@gmail.com	3035650610
Sheleoni	Chung	https://www.researchgate.net/profile/Chung_Sheleoni	https://osf.io/ksdy7/	u3527004@connect.hku.hk	sheleoni@gmail.com	3035270044
Daniel	Li	https://www.researchgate.net/profile/Daniel_Li54	https://osf.io/h5zea/	u3565527@connect.hku.hk	dyl2@illinois.edu	3035655270
Helen	Cheuk	https://www.researchgate.net/profile/Yat_Sum_Cheuk	https://osf.io/re8jh/	yscheuk@connect.hku.hk	helencheuk01@gmail.com	3035496014
Akir	Hall	https://www.researchgate.net/profile/Akir_Hall	https://osf.io/2mcns /	u3565375@connect.hku.hk	akirhall@gmail.com	3035653753

Managerial Summary

With the rapid technological improvements made in the past decade, the results of psychological research done around the globe have been made readily available for scientists to build further research off of. However, any subsequent experiments would require scientists to have complete certainty in the research that they are building their foundations on in order to produce reliable results. With that being said, psychological research has an incredible amount of variables ranging from gender perspectives to cultural biases, so the ability to recreate the results of other scientists in similar conditions is vital in providing that certainty for researchers to conduct a more in depth exploration.

Where this has become a crisis is that there have been an alarming number of experiments revealed to singularities and unable to be reproduced in any scientifically acceptable manner. Since publications of scientific research generally mandate the documentation of all aspects of the experiment conducted, an inability for subsequent researchers to follow the same criteria and produce the same results creates a defective base that nothing can be built on. As with any scientific field, psychology has infinitely more aspects for scientists to research as our understanding is constantly being expanded and clarified, but this cannot be done with fraudulent and overblown results diluting veritable outcomes. This chapter is particularly concerned with the events that led to the realization that there is, in fact, a reproducibility crisis in psychological publications. Whether it be data fabritation or the lack of research validation and transparency, recent revelations have highlighted the significant need for replications in order to validate reported findings.

In Depth Report

Science as a Game

In an age where scientific breakthroughs have more and more common, the idea of what makes a good scientist differs depending on the eye of the beholder, like two sides of the same coin: those who seeks the truth, or those who make new discoveries? It can be argued that a scientist is (or a researcher) driven by a need to make new discoveries in academia. With the pressures of obtaining necessary research funding and grants, it comes as a natural consequence that a “good scientist” must constantly make new discoveries and provide new insights to stay afloat in academia. Replications of experiments which offer no new discovery or "significant findings", therefore offer little incentive for researchers to investigate. These factors have created an ideal environment for incorrect data to go by undetected, thus contributing to the replication crisis.

As early as the 1970s, researchers have suggested that sciences did in fact, resemble a game (Mahoney, 1976), and psychological sciences is no exception. According to Bakker, Dijk and Wicherts (2012), the game psychological sciences involves players (individual researchers), competing teams (hypotheses and paradigms), judges (editors and reviewers) that in pursuit of game points (publication) and trophies (awards, funding and professorships). Simply put, one wins in the arena of psychological sciences by producing as many new or "useful" results to contribute to academia. In most research, a heuristic to measure success is by obtaining an alpha level (or namely p-value) of less than .05 in null hypothesis significance testing (NHST).

Data Fabrication in Replication

Karen Ruggiero

While there is no exact point in time where the credibility in science became a widely recognized issue in academia, the deep rooted problem in data collection as well as malpractices in research has slowly surfaced in the past decade. These cases of manipulating research methods and results in an attempt to strengthen publications have deeply scarred the scientific community, and only serve to deface the journals and institutions that their peers rely on. For example, incidents of data fabrication have occurred even in most reputed universities with one of the most prominent cases being Karen Ruggiero's data fabrication at Harvard University in 2001. Karen M. Ruggiero, had been found to fabricate data of 600 subjects in the field of social psychology while she worked at Harvard. As she admits, the data she has fabricated had been reported in two papers. (Holden, 2017). According to other news sources, Ruggiero had also fabricated data that were then published in the world's most prestigious journals such as the *Journal of Personality and Social Psychology* (Witkowski, 2014).

Although she was banned from signing contracts with any institution funded by the US government for five years, after that her academic career seemed to have continued, as she has worked Editor-in-Chief of *Behavioral Health* journal in 2006. Fraudulent results were also appeared in the journal, and full-text versions of her fraudulent work was reportedly still available in both printed version and the electronic database, despite her articles were previously retracted (Witkowski, 2014).

Diederik Stapel

In 2011 an interim report was published by Tilburg University investigating the depth of misconduct by distinguished Dutch social Psychologist, Diederik Stapel. His misconduct involved the fabrication of whole experiments as well as the manipulation of data, and is thought to date back to 2004.

Prior to this it was trusted that contrived findings would be uncovered if, other researchers attempted replications of an original study failed. Due to the aforementioned incentives this method of defense was proven ineffective. As a result Stapel's repeated misconducts went undetected. In 2003 the American Psychological Association (APA) started using an 'electronic manuscript tracking system.' Since 2003, 40 manuscripts were submitted, with 24 accepted into APA journals (Crocker and Cooper, 2011).With surprising ease, he was able to get his fraudulent work to jump through various hoops necessary, and achieve publication undetected.

Not only did he compromise his reputable position and the credibility of 'science', his fraud affected a number of students working under his guidance, who unbeknownst were handling fabricated data. Suspicions were alerted after students noticed inconsistencies in Stapel's published work and brought the problem to department heads. A Nature article by Crocker et al (2011) cited 'power and prestige,' as a reason for why Stapel was able to further continue his misconduct. The interim report stated that concerns over misconduct had been raised previously, but no investigation commenced. As well as having sole control over data collection; students reported his intimidating demeanor, when students questioned their lack of involvement in collecting their data. Furthermore, there were a number of 'red flags' such as 'insufficient clarity in the manuscripts as to how data was collected' (Crocker et al 2011)' that went undetected.

Frank L. Schmidt

Schmidt (2016) listed 6 common ways a statistically significant result can be fabricated. "(a) adding subjects one by one until the result is significant, then optional stopping; (b) dropping studies or measures that are not significant; (c) conducting multiple significance tests on a relation and reporting only those that show significance (cherry picking); (d) deciding whether to include data after looking to see the effect on statistical significance; (e) hypothesizing after the results are known (harking); and (f) running a lab experiment over until you get the "right" results." (Schmidt,2016).

Other than that, there are different ways in which researchers can manipulate their research design to increase their odds in "winning" at the game of obtaining a p-value < .05. An article by Bakker, Dijk & Wicherts (2012) describes the four strategies that researchers employ in research: (a) to perform a single large study with sufficient power and then publish it, (b) perform a large study and employ sequential testing until a significant result is obtained (for instance, by testing an extra dependent variable that correlates with the primary dependent variable, adding subjects, removing outliers and rerunning analysis), (c) to divide the total sample size N by 5 and in equal chunk and perform a maximum of five studies of N/5 in each chunk and publishing only the significant result in the expected direction and (d) to carry out a maximum of five small studies while applying QRPs described previously when needed, and reporting only the first study that yielded favorable results (Bakker et al., 2012, p.545).

In fact, these claims were supported by studies. John(2012) carried out a study investigating the prevalence of questionable research practice within the academic field of psychology. They surveyed more than 2000 psychologist, participants were asked to indicate anonymously if they had previously involved in these 10 questionable research practices, as illustrated in figure 3, such as falsifying data, rounding off p-value, selectively reporting studies that worked, selectively exclude relevant data....and more. Results showed that a high percentage of psychologists admitted their involvement in at least one of the questionable research practices. Moreover, it is also notable that a large percentage of participants had doubts on the research integrity of the academic field.

More specifically, nearly half of the researchers surveyed revealed that they only publish results that have worked, and 57% have admitted to sequential testing in their research. Sequential testing (or sequential analysis) means that the researcher is able to observe the data during data collection, and stop the data collected reaches a statistically significant result, or until maximum sample size has been exhausted (John, Loewenstein & Prelec, 2012).

<table>
<tr><td>1. Failing to report all of a study’s dependent measures in the publication</td><td rowspan="9">Significant Forms of Data Fabrication</td></tr>
<tr><td>2. Collecting more data after initial results to increase significance</td></tr>
<tr><td>3. Failing to report all of a study’s conditions in the publication</td></tr>
<tr><td>4. Prematurely ending data collection because target results had been achieved</td></tr>
<tr><td>5. Rounding off p values</td></tr>
<tr><td>6. Selectively reporting successful works in the publication</td></tr>
<tr><td>7. Optionally excluding data that could otherwise impact desired results</td></tr>
<tr><td>8. Reporting unexpected findings as predicted results in the publication</td></tr>
<tr><td>9. Knowingly downplaying or ignoring</td></tr>
</table>

demographic variables on results in the publication	
10. Falsifying data	

Figure 3- Examples of how publications have engaged in data fabrication. (John et al. ,2012)

P hacking/ScienceMedia

Peter Onneken

A group of German researchers devised an intentionally flawed study that set out to demonstrate the weight loss benefits of chocolate. The sole real part of the study was to be the participants and the clinical trial. The aim was to expose the failings of both the media and the nutritional science world.

Under the fake alias of Johannes Bohannon PhD, researcher of the institute of diet and health (a fictitious institution.) John Bohannon along with Peter Onneken and Diana Löbl recruited 16 participants (aged 19-67) through facebook and screened them for any underlying health or eating issues. Gunter Frank, a GP, conducted the trial. Researchers divided participants into 3 groups. Group one followed a low-carbohydrate diet. The second group followed the same low-carb diet as group one but also ate 1.5 oz. bar of dark chocolate each day (Bohannon, Koch, Homm & Driehaus, 2015). The remaining (control) group were instructed to continue eating as they had done before the study. Participants weighed themselves every morning for 3 weeks.

The study concluded with the administering of questionnaires and blood tests. The study concluded that the “consumption of chocolate with high cocoa content can significantly increase the effects of weight loss diets...long term weight-loss seems to occur easier and more successfully by adding chocolate.” (Bohannon, Koch, Homm & Driehaus, 2015). Furthermore the chocolate group lost weight at a rate of 10% faster than the low carb diet group.. A difference that was significantly different in addition to this group having better well being scores and cholesterol readings than the low carb diet group.

While this may appear to be natural results at face value, the authors deliberately used p-hacking by measuring 18 different things about this tiny sample they had a “60% chance of getting some“significant” result with $p < 0.05$.” In order to avoid getting caught at the peer review stage, they turned to fake publishers and submitted their article to 20 journals. From the responses they received they had chosen the international archives of medicine. It was accepted directly into their premier journal for 600 euros with no revisions whatsoever despite the journal having said that “All articles submitted to the journal are reviewed in a rigorous way, following the standards of ICMJE.”

After having constructed a press release, and shot 2 promotional music videos about the study, the article landed in the “*Daily Star*, the *Irish Examiner*, *Cosmopolitan*’s German website, the *Times of India*, both the German and Indian site of the Huffington Post, and TV news in Texas and an Australian morning talk show.” As well as sharp magazine whos fact checker merely confirmed the researchers name and a few sentences.

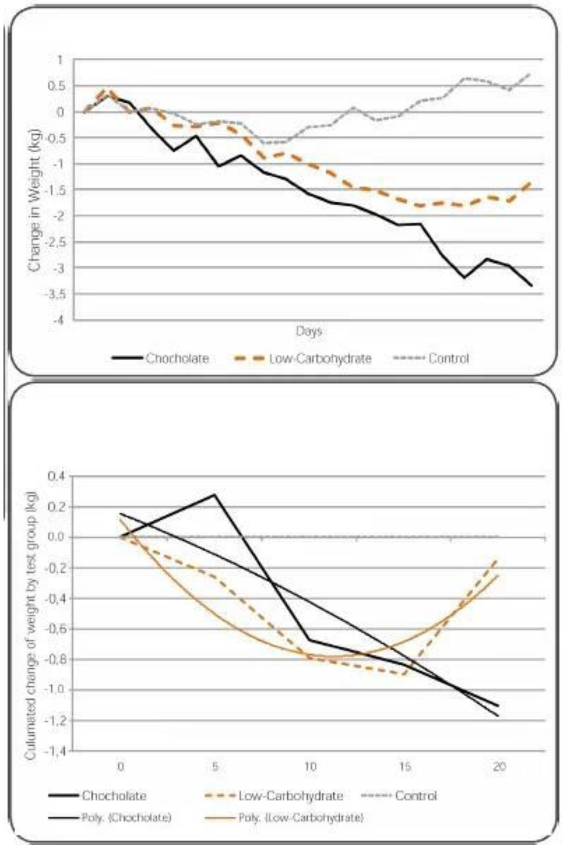


Fig.3 Statistical evidence showing that consumption over time of high cocoa concentration chocolate is effectively more capable of inducing weight loss than low-carb diets (Bohannon, Koch, Homm & Driehaus, 2015)

Mainstreaming Scientific Transparency

Daryl J. Bem

Esteemed Psychologist Daryl J. Bem's apparent proof for the existence of ESP in 2011, lead many researchers to reconsider the academic conventions of the time and reappraise their beliefs about what rigorous scientific practices actually are . As controversial as parapsychology can be, this study was even more disconcerting to mainstream psychology because of where they were published: The Journal of Personality and Social Psychology. A journal that is well respected and boasts both a high impact factor and a similarly high rate of submission rejections (87% in 2017).

The proof came in the form of 9 experiments conducted over the course of a decade, and involving in excess of 1000 participants, that seemed to indicate that precognition ''feeling the future'' was a legitimate ability. The experiments investigated 4 standard psychological effects: affective priming, habituation, approach/avoidance and facilitation of recall. The critical difference was that these effects were time reversed so that the important manipulation came after the participant had completed the task. Experiment 8 stemmed from the idea that the rehearsal of words from a word list improves recall of the practiced words. Participants were first presented with 48 common nouns from 4 categories in a standardised order and were asked to visualise what each word represented. Following this, the participants were given a free recall test and asked to write down as many words as they could in any order. Afterwards came the rehearsal section of the task. 24 of the original words were presented to the participants and they had to write down 6 words (which turned red when clicked) into slots on the computer screen. This was repeated for each category. The prediction that Bem made (and was illustrated by his results) was that recall would be higher for the words that were to be practiced in the future- an effect he called: 'retroactive facilitation of recall' (Bem, 2011)

Many scholars pointed out various issues with Bem's approach to experimentation and methodology that, in truth, reflected many of the standard research practices of the time. The only difference was that most other research findings were benign and not concerned with the supernatural. Thus the results- reliable or unreliable, valid or invalid, did not cause any uproar within the scientific community like Bem's did. An important criticism from Wagenmakers et al was the idea of exploration vs confirmation. They quoted Bem himself in pointing out that the research does not often follow the standard procedure of: hypothesis- theory- design- test- analyse (Wagenmakers, Wetzels, Borsboom, & Van Der Maas , 2011) and proposed a 2 step method. The first step, when one does not have a strong theory, involves exploring the data until an interesting hypothesis presents itself. The second step is where such a hypothesis is deliberately tested against new data in order to confirm or disconfirm the predictions. This would lead to statistical tests being adjusted accordingly and also prevent authors from portraying exploratory studies as ones that had been confirmatory all along.

Psychologists at the universities of Edinburgh, goldsmiths and hertfordshire accepted Bem's invitation for replications. These 3 replications of experiment 9 failed despite being remarkably close to the original study e.g. having the same number of participants and benefiting greatly from Bem's transparency, using the software and materials that he provided. Ironically, the researchers struggled to publish their pre-registered replications in JPSP, Science Brevia and Psychological science. When the paper was finally submitted for peer review by The British Journal of psychology, one of the peer reviewers was Bem himself and thus the paper was subsequently rejected by the editor. The replication was eventually published in PLoS ONE, an open access journal with a high impact factor.

While, amongst the converted, this study was (and still is) seen as an important milestone in the field of ESP research. This is because it is viewed as giving the field an opportunity to be taken seriously by mainstream science. To the rest of psychology however, Bem's paper represents a moment where all that was wrong with the current system was crystalised in a single publication and the realisation that anyone could have been Bem.

Table 7

Psi Performance in All Nine Experiments: Probability Levels (p), Effect Sizes (d), and Correlations (r) With Stimulus Seeking (SS)

Phenomenon tested and experiment	p full sample	d full sample	Correlation with SS	p high SS	d high SS	p low SS	d low SS
Precognitive approach/avoidance							
1. Detection of Erotic Stimuli	.01	0.25	.18*	.00002	0.71	.524	−0.01
2. Avoidance of Negative Stimuli[a]	.009	0.20	.17**	.001	0.45	.215	0.08
Retroactive priming							
3. Retroactive Priming I[a]	.007	0.26	−.05	.148	0.17	.036	0.24
4. Retroactive Priming II[a]	.014	0.23	−.07	.059	0.27	.035	0.23
Retroactive habituation							
5. Retroactive Habituation I							
Negative trials[b]	.014	0.22					
6. Retroactive Habituation II							
Negative trials[b]	.037	0.15					
Erotic trials	.039	0.14	0.24***	.002	0.57	.219	−0.09
7. Retroactive Induction of Boredom[a]	.096	0.09	.16**	.018	0.22	.483	0.00
Retroactive facilitation of recall							
8. Facilitation of Recall I	.029	0.19	.22**	.0003	0.57	.525	−0.08
9. Facilitation of Recall II	.002	0.42	−.10	.049	0.44	.013	0.40
Mean effect size (d)		0.22			0.43		0.10

[a] Probabilities and effect sizes in this row are based on the mean of the t values across the variations of the data analysis. [b] The Stimulus-Seeking Scale was not administered in this experiment.
* $p < .05$. ** $p < .02$. *** $p < .01$.

Fig4- Original study data illustrating effect known as 'retroactive facilitation of recall' . (Bem, 2011)

Modifying and Omitting Data

Dirk Smeester

One year after Diederik Stapel's fraud was exposed, Uri Simonsohn accused another psychologist of data fabrication. In 2011, Dirk Smeesters, a psychology professor at the Rotterdam School of Management was accused of modifying data after running the experiments, omitting some negative data in order to produce a significant result in favor of his hypotheses. Omitting important data without mentioning in the paper is a serious academic offense. The accusation started when Uri Simonsohn discovered unusual patterns in the data of one of the Smeesters published papers, he emailed Smeesters and demanded the raw data of his published study for verification. Smeesters claimed that the data were lost. He also denied violating academic integrity and defended himself by claiming that he was just "massaging the data" . He also claimed that data massaging was a common practice among psychology researchers and it was a part of psychology research culture. Massaging data means eliminating unnecessary information by extracting important data, it is clearing not what Smeesters did. Using statistical techniques, a scientific integrity committee discovered that the data in two of his published articles were statistically highly unlikely. Given the nature of the experiment the positive results were too abundant compared to negative results. As a result, his deception was exposed. This resulted in Smeesters's resignation and the end of his academic career.

False Positives

A false positive maintains that an effect is present when it is not. In the context of psychological literature, this would be incorrectly rejecting the null hypothesis. An article by Simmons, Nelson, and Simonsohn (2011), suggests that a p value of $p < .05$, changes in data collection and analysis can increase the likelihood of a Type 1 error . They propose that it is much easier to find a 'false effect' than to find valid evidence suggesting that an effect is not present. As aforementioned there is an evident motivation to neglect compliance with scientific principles in favour of 'novelty' (Rahal, R 2015). It is scarce for reputable journals to publish identical replications or null findings (Simmons et al, 2011). Factors such as limitations as a result of regulation, that permit only certain research designs and what is reported (Armstrong and Green 2017) all serve as motivations for this. This can result in publication bias**.**

When researchers incorrectly reject the null hypothesis resulting in a false positive or Type 1 error the implications can be severe. False positives have the potential to result in great expenditure in research programs that are built on incorrect data. This in turn can lead to inefficient policy changes.

When collecting data, the researcher must consider sample size, how and which conditions must be compared amongst other factors (Simmons et al, 2011). Simmons et al (2011) suggest that researchers favour an ‘easier’ approach. When considering a number of analytic options to substitute in, they opt to use an analytic tests that are likely to result in ‘statistical significance,’ and will choose to just report these findings. A consequence of this is that the likelihood of committing a type 1 error becomes greater than 5%.

Moreover, Simmons et al (2011) maintain that a compelling amount of literature, ‘are self-serving in their interpretation of ambiguous information.’ They used the analysing of ‘reaction times: how to treat outliers’ as an example to show this. Upon inspection of 30 ‘Psychological Science’ articles they found considerable disparity in researcher’s decisions. Most omitted the outliers for being ‘too fast,’ but there was a large difference in what was considered ‘too fast’. Some used the fastest 2.5%, others suggested that ‘fastest’ was quicker than 200 or 300 ms. Likewise there were inconsistencies in what researchers considered ‘too slow’. Although this does not necessarily suggest that these findings are incorrect, this in essence makes the conclusions drawn justifiable (Simmons et al, 2011). This can be further illustrated by Silberzahn, Uhlmann, Martin, Aust, Autry and Carlsson (2018), who invited 29 groups of researchers to analyse the same research question: ‘whether soccer referees are more likely to give red cards to dark-skin-toned players than to light-skin-toned players.’ Despite analysing the same dataset, researchers ran different tests. The estimated effect size varied from ‘0.89 to 2.93 in odds-ratio units’. Results showed 69% of researchers had a significant finding. These findings show how subjective analytic decisions can lead to a researcher arriving at a particular result.

Simmons et al (2011) used computer simulations to assess how four common researcher degrees of freedom can affect the likelihood of a false positive outcome. Figure 1 illustrates this clearly. They also suggested a series of ‘requirements for authors’ and ‘guidelines for reviewers (See figure 2).’ The instructions attempt to reduce the likelihood of false positives.

Table 3. Study 2: Original Report (in Bolded Text) and the Requirement-Compliant Report (With Addition of Gray Text)

Using the same method as in Study 1, we asked 20 34 **University of Pennsylvania undergraduates to listen** only **to either "When I'm Sixty-Four" by The Beatles or "Kalimba"** or "Hot Potato" by the Wiggles. We conducted our analyses after every session of approximately 10 participants; we did not decide in advance when to terminate data collection. **Then, in an ostensibly unrelated task, they indicated** only **their birth date (mm/dd/yyyy) and** how old they felt, how much they would enjoy eating at a diner, the square root of 100, their agreement with "computers are complicated machines," **their father's age**, their mother's age, whether they would take advantage of an early-bird special, their political orientation, which of four Canadian quarterbacks they believed won an award, how often they refer to the past as "the good old days," and their gender. **We used father's age to control for variation in baseline age across participants.**

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to "When I'm Sixty-Four" (adjusted M = 20.1 years) rather than to "Kalimba" (adjusted M = 21.5 years), $F(1, 17) = 4.92$, $p = .040$. Without controlling for father's age, the age difference was smaller and did not reach significance (Ms = 20.3 and 21.2, respectively), $F(1, 18) = 1.01$, $p = .33$.

Figure 1- 'Study 2: Original Report and the Requirement-Compliant Report ' (Simmons et al, 2011)

Table 2. Simple Solution to the Problem of False-Positive Publications

Requirements for authors

1. Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
2. Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
3. Authors must list all variables collected in a study.
4. Authors must report all experimental conditions, including failed manipulations.
5. If observations are eliminated, authors must also report what the statistical results are if those observations are included.
6. If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

Guidelines for reviewers

1. Reviewers should ensure that authors follow the requirements.
2. Reviewers should be more tolerant of imperfections in results.
3. Reviewers should require authors to demonstrate that their results do not hinge on arbitrary analytic decisions.
4. If justifications of data collection or analysis are not compelling, reviewers should require the authors to conduct an exact replication.

Figure 2- 'Simple Solution to the Problem of False-Positive Publications' (Simmons et al, 2011)

Questionable Research Validity in Psychology

According to Schmidt(2016), many psychology studies have a high rate of replication failure due to various reasons including a low statistical power and sampling error. Statistical power is the probability that the test rejects the null hypothesis when it is false. It ranges from 0 to 1, the higher the power results in a low probability for type 2 error (false negative) to occur. According to Cohen (1962) , the average statistical power in most psychology studies are between 0.40 to 0.50. The average power has not increased over the last 50 years (Shen, 2011). A low power showed that there is a high probability of retaining a false null hypothesis. While in the academic field of psychology, the null hypothesis is rarely true. Lipsey and Wilson (1993) carried out a study examining more than 300 academic papers, less than 1% of the null hypotheses are true. Therefore suggesting that type 2 error in psychology is highly probable.

Moreover, some studies produce unreliable results due to sampling error, as sampling error can occur easily when the sample size is small. Researchers believed that random sampling compensates the possible inaccuracy of studies with a small size, they believed that a small sample drawn randomly from the population is still representative to the general population. However, a study showed otherwise, data from a study with more than 1000 participants was randomly broken down into smaller non-overlapping studies, the studies with small sample size produced results with large range of varying r coefficients, the results produced were far from the r coefficient obtained with more than 1000 data (Schmidt,1985). Therefore, research with small sample size cannot be generalized. Other than that, a recent pre registered survey of more than 7,000 psychological researchers in the USA regarding their own (and colleagues) use of QRPs, it was estimated that 18% have used at least one QRP in the past year, and roughly 24% reported to knowing people in their social network who have used QRPs. Still, QRPs remains in the grey area of scientific practice, which makes it difficult for us to investigate the severity of QRP use (Fox, Honeycutt & Jussim 2018).

As a result of the prevalence of questionable research practices, scientific progress in psychology has been stalled. With studies producing false results and unreplicable data the psychological field has become embroiled in a web of deception that has become a reproducibility crisis.

Revelation of Mass Replication Studies

Similarly, an increasing number of researchers have begun to grasp just how widespread the lack of reproducibility has become among the academic community with studies deemed false or incorrect as a result. Where a replication is found unsuccessful the direction of the effect could be found to change or a smaller effect is found or there can be failure to find an effect at all. Nature conducted a survey of 1,576 researchers giving them a brief questionnaire regarding reproducibility in research The data revealed contradictory findings, whilst 52% acknowledged there was a significant ‘crisis’ (see figure 6), the majority maintain they trust published literature (Bakwe, 2016).

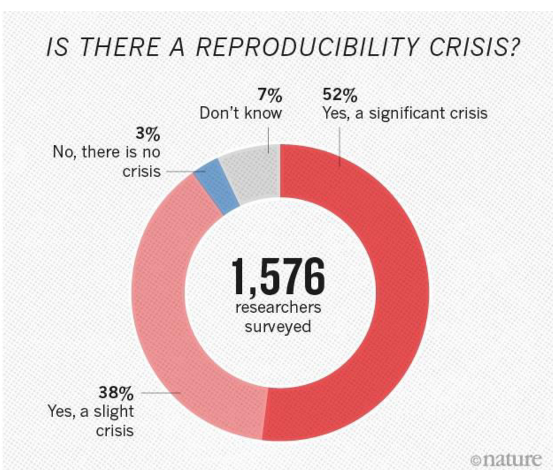


Figure 6- Figure showing researcher’s opinions regarding the replication crisis (Baker, 2016)

Additionally, a study conducted in 2012 revealed a shocking 1.07% overall replication rate of psychological research publications derived from five hundred randomly selected publications (Makel et al, 2012). This percentage was the result of comparing the category of replication that sample publications fell under as well as their success, which was found to be heavily correlated with the author of each publication. Comparing between direct (i.e. cloned experimental and sampling operations) and conceptual (i.e. adapted hypothesis based processes), no statistical significance was found to indicate that either method was more successful than the other ($p = .09$) despite a notable majority of published replications being conceptual rather than direct. In light of the large percentage of conceptual replications in the sample group (81.9%), their statistical indifference in success was considered to be "somewhat counterintuitive" but accounted for by the idea that the majority of failed conceptual replications were more likely to be refused by publishers or left unsubmitted altogether.

Furthermore, the overlap between authors of replications and original publications revealed a dramatic influence in the success of sample publications. As shown in the study, with at least one third of the original authors on a replication team, the success rate skyrocketed to a staggering 98.2% and had a p value of $< .01$ compared to publications made without any overlapping authors (Makel et al, 2012). However, the effects of such overly positive results in overlapping authorship could suggest the presence of an effect known as "the file-drawer problem" where researchers prefer not to publish their failed replications. Taken into account with the resulting statistical indifference of research procedures, it would be logical to assume that key factors of the replication crisis can be attributed to the fact that researchers are less incentivized to pursue replications if they do not have an overlapping author or that leave their failed attempts unpublished entirely.

Quiz

Write 10 easy quiz 4 multiple choice questions to ensure reading and understanding of this chapter:

1. What did Dirk Smeesters do resulting in his resignation?
    a. <u>Data fabrication through omitting negative data</u>
    b. Plagiarism (copying someone else's work without referencing)
    c. Stealing other people's data
    d. Fabricating data without carrying out the studies.
2. According to Schmidt(2016), what makes sampling error likely to occur in research studies?
    a. <u>A small sample size</u>
    b. When participants are not recruited through random sampling
    c. Possible experimenter bias in the research methods
    d. When there is no control group in the study
3. In the study by Silberzahn et al (2018) it was found that _____ % of researchers found a significant result in 'whether soccer referees are more likely to give red cards to dark-skin-toned players than to light-skin-toned players.'
    a. 31%
    b. 52%
    c. 69%
    d. 76%

4. What ability did Daryl J. Bem's studies claim to prove?

    a. Telepathy
    b. Psychokinesis
    c. <u>Precognition</u>
    d. Astral Projection

5. According to Bohannon, using P-hacking, what was their chance of finding a significant result among the 18 things they measured?

    a. 95%
    b. <u>60%</u>
    c. 33%
    d. 5%
6. What is sequential testing (or sequential analysis)?
    a. To collect data from subjects in the ascending order of age
    b. <u>To observe data while collecting data</u>
    c. To re-run data analysis subsequent to publishing the results
    d. A process to check whether the research has followed the proper sequence to prevent p-hacking.
7. How many psychological researchers, according to John, Lowewenstein and Prelec (2012), have admitted to using tactics of sequential testing?

   a. Less than 20%
   b. 42%
   c. <u>57%</u>
   d. 85%
8. What are the two categories of replications?
   a. <u>Direct/conceptual</u>
   b. Indirect/direct
   c. Imitation/alteration
   d. APA/MLA
9. Which of the following is a theorized effect where researchers choose not to publish failed replication attempts?
   a. Folding chair problem
   b. <u>File drawer problem</u>
   c. Pride maintenance problem
   d. Safe drawer problem
10. Which factor was largely correlated with the potential success of replications?
   a. Replication attempts
   b. Procedural similarity
   c. IQ of researchers
   d. <u>Overlapping authors</u>

References

Armstrong, J. S., & Green, K. C. (2017). Guidelines for science: Evidence and checklists. Scholarly Commons, 1-24.

Baker, M. (2016). Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the 'crisis rocking science and what they think will help. Nature, 533(7604), 452-455.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*(6), 543-554.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. doi: 10.1037/a0021524

Bohannon, J. (2015, December 16). I Fooled Millions Into Thinking Chocolate Helps Weight Loss. Here's How. Retrieved from https://io9.gizmodo.com/i-fooled-millions-into-thinking-chocolate-helps-weight-1707251800.

Chocolate causes weight loss. (n.d.). Retrieved from https://www.scribd.com/doc/266969860/Chocolate-causes-weight-loss.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, *65*, 145–153.

Data and Documents. (2019, November 14). Retrieved from https://science.sciencemag.org/content/suppl/2013/10/03/342.6154.60.DC1.

Engber, D. (2017, June 7). Daryl Bem Proved ESP Is Real. Which Means Science Is Broken. Retrieved from https://slate.com/health-and-science/2017/06/daryl-bem-proved-esp-is-real-showed-science-is-broken.html.

Enserink, M. (2012, June 25). Rotterdam Marketing Psychologist Resigns After University Investigates His Data. Retrieved from https://www.sciencemag.org/news/2012/06/rotterdam-marketing-psychologist-resigns-after-university-investigates-his-data.

Fox, N., Honeycutt, N., & Jussim, L. (2018). How Many Psychologists Use Questionable Research Practices? Estimating the Population Size of Current QRP Users.

French, C. (2012, March 15). Precognition studies and the curse of the failed replications | Professor Chris French. Retrieved from https://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failed-replications.

Gonzalez, M. M. (n.d.). Relaunching A Mega Journal – International Archives of Medicine. Retrieved from http://www.brainblogger.com/2015/02/17/relaunching-a-mega-journal-international-archives-of-medicine/.

Holden, C. (2017, December 10). Psychologist Made Up Sex Bias Results. Retrieved from https://www.scieOpen Science Collaboration, & Rahal, R. M. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), [aac4716]. https://doi.org/10.1126/science.aac4716

Ioannidis, J. P. (2005). Why most published research findings are false. PLoS medicine, 2(8), e124.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. Psychological science, 23(5), 524-532

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. American Psychologist, *48*, 1181–1209.

Mahoney, M. J. (1976). Scientist as subject: The psychological imperative.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research. *Perspectives on Psychological Science*, *7*(6), 537–542. doi: 10.1177/1745691612460688

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the Future: Three Unsuccessful Attempts to Replicate Bems 'Retroactive Facilitation of Recall' Effect. *PLoS ONE*, *7*(3). doi: 10.1371/journal.pone.0033423

Schmidt, F. L., & Oh, I.-S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology, 4*(1), 32-37.

Schmidt, F. L., Ocasio, B. P., Hillery, J. M., & Hunter, J. E. (1985). Further within-setting empirical tests of the situational specificity hypothesis in personnel selection. *Personnel Psychology*, *38*, 509–524.

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. Journal of Applied Psychology, *96*, 1055–1064.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... & Carlsson, R. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. Advances in Methods and Practices in Psychological Science, 1(3), 337-356.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological science, 22(11), 1359-1366.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... & Carlsson, R. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. Advances in Methods and Practices in Psychological Science, 1(3), 337-356.

Summary Report of Journal Operations, 2017. (2018). *American Psychologist*, *73*(5), 683–684. doi: 10.1037/amp0000347

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Han L. J. Van Der Maas. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. doi: 10.1037/a0022790

Witkowski, T. (2014, June 4). From the Archives of Scientific Fraud – Karen M. Ruggiero. Retrieved from https://forbiddenpsychology.wordpress.com/2014/05/22/from-the-archives-of-scientific-fraud-karren-m-rugiero/.

2 - Identifying the problems

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Yan in	Wong	https://www.researchgate.net/profile/Yan_Wong16	https://osf.io/c5fnx	u3546936@connect.hku.hk	stefaniwong99@gmail.com	3035469360
Kwan Yi	Mui	https://www.researchgate.net/profile/Kwan_Yi_Mui	osf.io/23f7w	ashleym@connect.hku.hk	muiky0410@gmail.com	3035566766
Chi Nok	Lam	https://www.researchgate.net/profile/Chi_Nok_Lam	osf.io/rq4se	david720@connect.hku.hk	chinoklam@gmail.com	3035566508
Yi Chen Angela	Wu	https://www.researchgate.net/profile/Angela_Wu11	osf.io/j9r6a	angela25@connect.hku.hk	angelawu25@gmail.com	3035558783
Tsz Yan	Ho	https://www.researchgate.net/profile/Tsz_Yan_Ho2	http://osf.io/gswvy	u3554912@connect.hku.hk	htybobo@gmail.com	3035549122
Lok Hang Desmond	Lee	https://www.researchgate.net/profile/Desmond_Lee6	https://osf.io/g8xjs/	u3554931@connect.hku.hk	desmondlee8023@gmail.com	3035549316

Managerial summary

In this chapter, you will get more insight on understanding different problems that researchers and psychologists face when analysing data. There are a number of misunderstandings of statistics by them as well as the public, including the psycholgy's aversion of null results, the misinterpretation of significance and undisclosed flexibility in data collection and analysis. The above misunderstandings have blocked amounts of useful and valuable reports so far in different fields of psychology.

The problem in publication will also be examined in the chapter and how the publication system that exists traditionally possibly hinders the scientific developments by creating the "drawer files" phenomenon, preventing scientific knowledge to be shared across the science community. In addition, this chapter will shed light on the ongoing credibility crisis in scientific publishing due to a lack of transparency in experimental research process.

In depth report

Misunderstanding of statistics

Statistics are very important in psychological research, however, statistical bias and misunderstanding of statistics remain controversial in psychology. Fail in reporting replication reports with null results has blocked the publications of related reports and indeed, they are still valuable to be extinguished by researchers.

In replication reports that are published in recent years, there is a controversial statistical issue in the psychological field, namely, the psychology's aversion of null results and which researchers failed to publish them. (Ferguson & Heene, 2012) Null findings were being misinterpreted considering that reports that reject the null hypothesis were tended to be more reliable, indeed it wasn't the aim of the null hypothesis significance testing(NHST). And this can be risky because publication bias then arises and becomes a problem in many subfields in psychology. According to Rosenthal (1979), publication bias is the propensity for statistically significant results to be disclosed over nonsignificant data. People, mainly the publishers, may consider reports with statistically nonsignificant results are due to Type II error and even accuse that scholars are not working hard to find significant results. While the publication bias against null results arises, scholars tend to convert unsuccessful reports to a successful one by clearing some data and rerun them until the results are in favour of researchers' hypothesis. They increase the sample size until the result is statistical significance without the consideration for the triviality of the resultant effect size. (Ferguson & Heene, 2012) The reason behind such a phenomenon is that psychologists have so much knowledge in understanding the limitations of null-hypothesis statistical testing. The frame of the theoretical model has limited scholars in deciding publishing replication reports. Indeed, the argument of claiming the results which fall below the arbitrary $\alpha = .05$ line are not meaningful may not be true. When the assumptions of parametric statistics are not included in the involved databases.

Apart from this, the misinterpretation of significance among researchers and the public also cause bias in psychological reports. According to a research report by Haller and Krauss (2002), most researchers, particularly psychology students, may not understand the meaning of conducting a significance test. The Null Hypothesis Significant Test (NHST) has become a major method in testing the power of data for a long period. According to the overview of students' misinterpretation of NHST by Haller and Krauss(2002), there are two classes of people: the first class of students claimed that the NHST means that the measure lies 5% above the random-percentage. Haller and Krauss described it as meaningless interpretation. The second class is that the NHST can assess the probability of their hypothesis. From seeing the above misinterpretation of NHST by students as well as psychologists, it is clear to see that people have limited insight into the significant test. (Falk & Greenbaum, 1995 ; Oakes, 1986)The rationale behind such a misunderstanding of NHST is that statistical textbook may be the source of causing such a problem,ac Statistics are very important in psychological research, however, statistical bias and misunderstanding of statistics remain controversial in psychology. Fail in reporting replication reports with null results has blocked the publications of related reports and indeed, they are still valuable to be extinguished by researchers.

Furthermore, undisclosed flexibility in data collection and analysis affect one's understanding of data. Since False positive is the most costly(Simmons, Nelson & Simonsohn et al., 2011), researchers often decide when to stop the collection of data according to the basis of interim data analysis. False-positive means that the inaccurate elimination of the null hypothesis. (Simmons, Nelson & Simonsohn et al., 2011) According to a recent survey, around 70% of researchers admitted that they did the same thing during the data analysis process. (John, Loewenstein & Prelec et al, 2011) Nevertheless, it is found that there is an error by doing so. If an effect with small sample size is significant, it does not necessarily mean that it is also significant with larger sample size. Simmons, Nelson &Simonsohn (2011) researched this claim. In the research, they obtained data from a researcher who has already collected 10 or 20 observations within each of two conditions. After that, they test the significance of every 1,5,10 or 20 per-condition. The researcher then stopped the collection of data when either one statistical significant is obtained or when the amount of observations of each condition reaches 50. In figure 1, it shows the false positive rate of the researcher's data. Also, the figure shows that the significant effect is 22% after every new pre-condition observation.

According to Haller and Krauss (2002)'s doubts. The statistical textbooks often focus on the execution of the formal procedures rather than the meaning of the results. Although a significance test in the data is not related to the probability of H0 or H1, it is not included in textbooks.

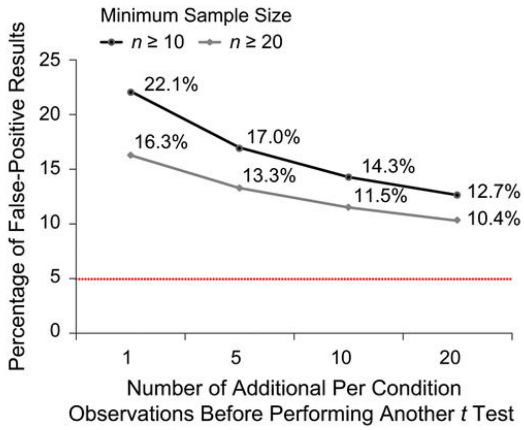


Figure 1 Likelihood of obtaining a false-positive result when data collection ends upon obtaining significance (*p* ≤ .05, highlighted by the dotted line). The figure depicts likelihoods for two minimum sample sizes, as a function of the frequency with which significance tests are performed.

In figure 2, it describes a data analysis of continuing sampling until the number of pre-condition observations reaches 70. The p-value in t-test was conducted after each pair of observations. The dotted line refers to the conventional significance criterion of p ≤ .05.In figure 2, it illustrates that the claim by Simmons, Nelson and Simonsohn(2011) that the difference in sample size may contradict to the statistical significance.

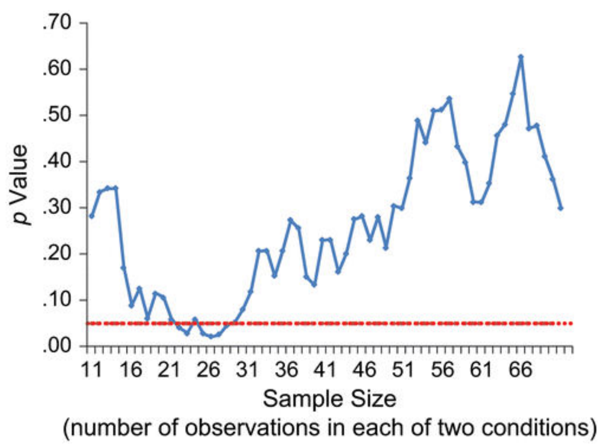


Figure 2 Illustrative simulation of *p* values obtained by a researcher who continuously adds an observation to each of two conditions, conducting a *t-test* after each addition. The dotted line highlights the conventional significance criterion of *p* ≤ .05.

Misuse of statistics: Questionable Research Practices

Statistics is the foundation of knowledge in science and has always been widely viewed with the notion of objectivity. The credibility of scientific research has always been highly recognized, and people assume scientific research as the best methodology for producing knowledge. However researchers can consciously or unconsciously use questionable research practices to manipulate results and shape what some readers might consider as truth. The conventional consensus in statistical research is that if the p-value is smaller than 0.05, it is statistically significant, meaning that if the null hypothesis (the hypothesis of no difference and no relationship between variables) is true, there is only a 5% chance for having the sample results, and it is unlikely that the results are due to a sampling or experimental error. As objective as it is, the p-value of a statistical analysis can be manipulated by the experimental design and how the data are used, and it can be manipulated to under 0.05, the cutoff for statistical significance, and this is known as p-hacking. One way of p-hacking is when there are many dependent variables and very few participants, there is a huge chance of obtaining a statistically significant result, and the variables that did not give a significant result will not be reported. When there are less participants, the fluctuation in their dependent variables and outliers has more influence on the overall data,knowing that p-values are sensitive to sample size.For example if we are investigating whether Chinese adolescents are more intelligent than American adolescents, one exceptional genius in the Chinese group can skew the overall IQ level of Chinese adolescents more if there are only 10 Chinese adolescents than if there are a thousand Chinese adolescents. In some other cases, the researcher may have a strong predisposition to their hypotheses that when the results did not give a significant p-value, they unconsciously are convinced that they made a sampling error, thus repeating the experiment until they have a statistical significance result or changing the flexibility of the outliers so that only the data that supports the hypothesis are used. The key problem is that the trials that did not give a statistical significant results are not published, because they are quite literally insignificant to our scientific knowledge, even though it is extremely important to acknowledge the context of the results to interpret it.

There are many problems with p-hacking, because the p-value was not created to be a definitive test for the accuracy of data, because it only assumes the null hypothesis. Statistics has its limitations and we should put more emphasis on the hypothesis and experimental limitations to determine whether the evidence should be taken as facts.

Credibility: Open Science On The Rise

Open science is extremely crucial to the path forward in achieving credible findings in the field of psychology, especially in the age of data sharing like never before. An issue that significantly reduces the credibility of empirical studies is the lack of transparency in the experimental research process; also known as ‘closed-science’. Closed science occurs when there is a lack of open research published and communicated -- done unconsciously or deliberately -- that is accessible for other researchers to view and replicate. Inadequate open disclosure of the experimental details and process poses the risk that the researchers may not have strictly followed the methodology and process stated, thus making the data noncredible. In addition, when there is a lack of sharing of research materials and data, it is easy for researchers to cherry-pick reported results or fabricate their experimental aim, hypothesis, and even methods upon conducting the study to match results or preconceptions. These biases can bring about significant effects towards conclusions drawn. Evaluations of data, experimental methods, and conclusions cannot be made if other researchers cannot access or understand how a researcher conducted their study in full details (i.e. experimental stimuli, number of participants, mean age, gender, race, distinct criterias, recruitment method). In addition, the lack of open-science creates a substantial area for malpractice to occur. Other researchers who aim to replicate the published research study in hopes to test for reliability may not be able to generate corresponding results and conclusions due to such lack of information. This is especially important in the field of social psychology as many factors can affect the responses and results.

To reduce credibility crisis, psychologists should aim to make all published research data material as openly available and transparent to other researchers and most general public. This can be achieved through utilising emerging tool platforms such as the Open Science Framework (OSF) which aims to create a more open and rigorous centralized workflows by capturing most if not all aspects of a research from the development of the topic idea, the designing of methodology, collection and analysis of data, writing of results and conclusions, to the final publishing of the report paper. Moreover, another method to increase transparency is the use of open notebook journals which similarly records the entire primary research project, including personal notes, thoughts, processed data, associated laboratory materials of the research should also be enforced. Qualitative researchers should always use the strategy of reflexivity to note how his or her personal subjectivity may have affected the findings, as data are often interpreted through the researcher's individual lens. Last but not least, triangulation, which refers to the use of various research methods (method triangulation) and researcher (researcher triangulation) based on the assumption that through comparing the different data collected from different methods and researchers can possibly overcome potential biases from the use of a single method or researcher, can also be utilised. Thus, the aforementioned methods may establish or maximize credibility in modern empirical research.

Publication system: aversion to null findings, no replications, wow findings, closed networks, closed review system, no accountability, huge file-drawer

The publication system that exists in the science community currently prevents most research and papers from being published. P- values have the job of interpreting whether the effect exists by seeing whether it falls below or above the threshold of $p<0.05$ and represents the probability of the data given that the null hypothesis is true. So if the P-value is lower, then it is safe to assume that the null hypothesis could be rejected. If a paper has a p-value smaller than 0.05, then it could be concluded that such an effect exists, and the value of the paper would be deemed higher, as it shows the phenomenon exists and brings something new to the science community, but if a paper has p-value larger than 0.05, then the paper would be deemed a lower value. This is because in science, $p<0.05$ has a meaning that it is statistically significant, and means that what is found in the study could possibly be generated to the public. Setting the standard in publication creates the phenomenon of "drawer files" - studies and papers that don't get published because their results deemed not significant by the publication system. Sometimes, the researcher may also censor themselves, adding on the "drawer files" phenomenon. This contributes to the problem of "closed science" - because the results are not published, others cannot understand and learn from the work. The science community is oriented towards "wow" and novel results, but that doesn't mean we cannot learn from other studies and papers. The drawer files prevent the science community to grow and learn from others - the researchers would be the only people that can get their hands to the studies, but if they can also publish their results, then other can also understand what is going on in the research. All research that has good questions, hypothesis and robust methods has its value. Sharing of data between academics allows for collaboration of work and sharing scientific knowledge would be easier.

Replications of the existing studies are also important because it increases the reliability of the results, but the value of replication has not been widely recognized. A replication that has failed would eventually become one of the drawer files, would be difficult for the public to understand more about the phenomenon. Some studies might be outdated, and even if one effect existed at the past does not mean they can be generated to the population now, but the closed network prevents science to develop.

In order for the results to be significant and the papers to be published, researchers might adopt other strategies so that their results would pass the threshold of $p<0.05$, such as "P-Hacking". "P-hacking" is the manipulation of data, like the removal of outliers that exists in the data, or removal of experimental conditions, i.e conditions that did not have a significant effect in order to have the desired p-value (Nuzzo, 2014). Another way of "p-hacking" would be measuring other unplanned variables, or variables unintended to measure but have seemed to cause the effect. Researchers can also adopt multiple measures, but only report the results that satisfy the $p<0.05$. The over-reliance of "significant results" drives researchers to generate figures that fit the significance level as it increases the chances of papers being published. The publication system promotes academic misconduct, in which researchers might turn to "harking" - meaning developing the hypothesis after getting the results, or decreasing the sample size so that significant results could be obtained more easily. The problems demonstrate how problematic the publication system is, and how it hinders scientific developments. This calls for a need for open science.

Bad incentives system: number of papers and impact factor/rankings, rather than solid science.

Apart from the problems illustrated previously, it should be noted that a bad reward system has also played its role in worsening the problematic credibility crisis in psychology.

It is common for psychology journals to solely publish works that achieve a statistically significant result in experiments, discarding and screening out works with experimental results that do not fulfil certain effect size and $p<.05$. This typical way for psychology journals to select works for publication has made severe implications to the credibility problem confronting psychology in the following ways.

Psychologists and university professors are usually rewarded with further research funding and a higher salary and status based on the following two indicators that are number of publications as well as impact factor respectively. In order to get promoted or receive funding for further psychological research, the use of p-hacking which manipulates experimental design to have a higher probability of achieving statistically significant results become more common. These practices lower the legitimacy and credibility of the conclusions drawn upon the experiments.

Even with ethical and procedural concerns being safeguarded by the strict restrictions listed by Open Science Framework with the risk of data and procedural manipulations barred or minimized, credibility problems do still exist under the bad reward system.

As previously mentioned, we can recognize publication of experimental works in academic journal in itself is a reward to the psychologists and university professors. However, the way of selection of experimental work in publication might not be a way that serves to reward hard-work, creativity and diligence.

For instance, let us look at a hypothetical example in explaining the arguments mentioned in the previous paragraph, we are having 10 professors doing the same experiment, by sampling probabilities as suggested by the p-value, maybe one professor successfully obtained results with $p<.05$ and concluded his hypothesis are to be accepted whilst 9 professors failed in obtaining results to reject the null hypothesis. Professors who fail to reject the null, may not submit their work to the academic journals when they are not doing a replication work, then the one whom successfully achieve the significant results will be rewarded with his work published. However, as we all know, the conclusion derived does not work in samples for most of the time then a type 1 error is likely to occur. This happens with the theory of ego depletion, a once widely-accepted and recognized theory happened not to be accepted by the psychology community nowadays.

Insofar as, impact factor , the number of citations that articles in journals received over-time is always seen as an indicator that determines the quality of an experimental work. People often thinks that the more the work is cited, the more recognizable and persuasive is the results of the work. However, this might not be the case, people often cite others work because they need to have a basis for themselves to generate a hypothesis instead of other reasons.

These demonstrates how the bad incentives system have drawbacks on the credibility of psychological work.

Reputation and Prestige

For scientists, establishing trust is crucial to communicating credibility(Fiske and Dupree, 2014). Reputation is an important indicator for research community, based largely upon the quality and quality of a researcher's publications. Even when a complete systemic information is absent, reputation allows informed quality assessments of both publications and scientists(Petersen *et al*. 2014). Readers expect higher expertise and trustworthiness when the researcher or a paper has a high reputation. Research misconduct, irreproducible results, conflicts of interest could decay a scientist's reputation.

With the number of journals grows 3.5% annually (Guide.lib, 2019), a reputation may matter more than ever, as people tends to select readings with good reputation/ prestige. Publishing in a prestige journal would guarantees high visibility of the paper, possibly with Open Access and other strong social media presence. For example, people may find prestige journals reliable because they have precise evaluation on articles, conduct appropriate peer reviews, and equipped with editorial policies.(Björn Brembs, 2013) A study surveyed 338 faculty members from the U.S. and Canadian institutions and from a variety of disciplines, including science, result shows that the highest priority for researchers to decide where to publish their manuscripts was “journal readership”.(Langin, 2019). For many researchers, a prestige journal ensure high publicity, and promote one’s career.

It matches another finding about citation lifecycle, that a paper’s citation count heavily relies on the author’s reputation in the early citation lifecycle, and the influence of authors’s reputation diminish drastically after a tipping point (Petersen *et al*. 2014). Both findings suggest that informed reputation has a spread effect (Woolston, 2015) that help researchers to gain review, and has become crucial factors in sharing scientific manuscript and results. Other ways to attain reputation like young researchers trying to collaborate with eminent authors (coat-tail effect), and gaining PhD at institutional prestige university (reputation by proxy). After all, “reputation aggregates,” Petersen suggested.

If doing the research and delivering its result to an appropriate audience are fundamental objectives researchers should prioritize. Yet, when a ‘big name’ could very possibly boost citations and readership regardless the paper’s scientific value and merits, not many researchers can find their value on other aspects of publishing beyond journal impact factor, which may prompt to some misconduct of researchers.

Under the increasingly competitive scientific environment, money funded in the field is insufficient to support every research projects, due to the current squeeze on resources, researchers often found themselves in a dilemma between readerships and reputation/ prestige, and between the benefits of science field or an individual researcher.

Considering this tradeoff may reveal the scientific culture, the relationship between replication and reputation may illustrate the problem. Replication is vital for increasing precision and accuracy of scientific claims. However, the present culture in science provides strong incentives for innovation and discourage incentives for certainty and reproducibility. (Nosek B et al. 2015), in another words, researchers would be more willing to generate exciting and innovative results regardlessly. Moreover, in a sense, researchers may treat their findings as possessions(Abelson R, 1986), results may be manipulated to align with certain direction, or with individual researcher’s benefits, to avoid dishonor failure. Discouraging adherence to scientific norms and values.

To promote integrity in scientific publication, Transparency and Openness Promotion (TOP) guidelines and more instructions, were designed to promote an open research culture, by facilitating review, replication, and sharing data, and reducing bias, (Nosek et al., 2015). Holly Bik, a biologist promotes her latest publications on her own social media by tweet and blog posts, to increase visibility in a fair and accurate manner.

Conclusion

"Reputation affects all areas of science,"(Petersen 2014) it is true that publication with prestige journals or producing exciting findings could gain concern, but it is also true that papers published in high-impact journals are more likely to be retracted (Björn 2013), reputation is a two sided sword. Moreover, in a study, it is found that scientists who value reproducibility and produces boring but certain findings are rated better and more ethical scientists than those who generate exciting but uncertain results by a wide margin. (Charles 2016). (see fig 1.)

In a nutshell, reputation and prestige absorbed from others could only provide short term conduciveness. Researches should not despair if their work is not valued in their early career. In a long run, it is firmly believed that the scientific content of a paper is the main and fundamental part, that build a scientist's own reputation. "Quality work is still the best statement you can make." (Santo Fortunato). Researchers should be committed themselves to integrity, and attention to their responsibility, and produce trustworthy and certain work.

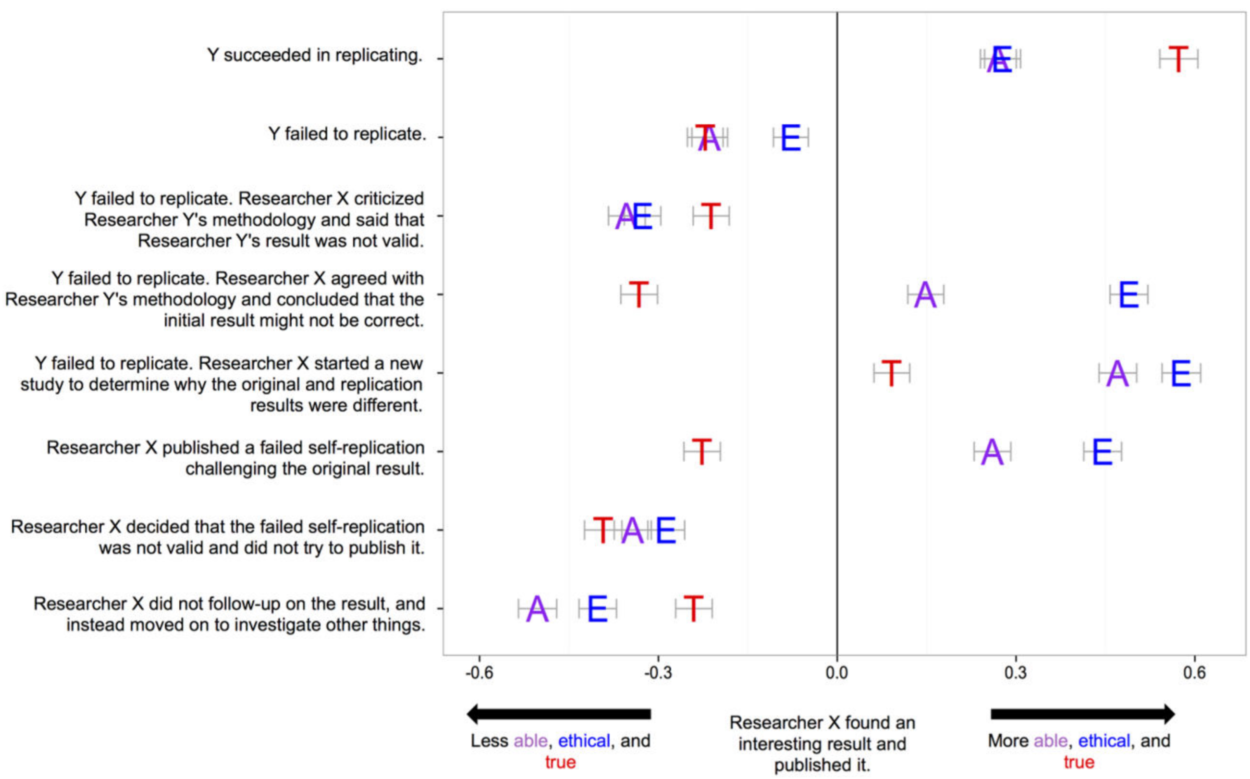


Fig 1. Effect of replication on perceived ability (purple) and ethics (blue) of Researcher X and truth (red) of the original result (*N* = 4,786).

Quiz

1) What is the major statistical problem researchers are facing in designing a replication report?
    a) Difficulty in searching for sample size
    b) Limited analysis method
    c) <u>Psychology's aversion of null results</u>
    d) Unable to replicate the data
2) What is the claim by Simmons &Simonsohn(2011) in the report of False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant?
    a) <u>The contradiction of statistical significance between small and larger sample size</u>
    b) The contradiction of statistical significance between consistent and inconsistent sample size
    c) The contradiction of statistical significance between the effect size
    d) The contradiction of statistical significance between H0 and H1
3) When there are many dependent variables and very low sampling size, ___
    a) There is a higher chance of obtaining more statistics
    b) <u>There is a higher chance of obtaining statistical significant results</u>
    c) There is a lower chance of obtaining statistical results
4) the p-value assumes the ____ hypothesis
    a) research
    b) probability
    c) <u>null</u>
5) What could occur when closed-science approach occurs?
    a) <u>Researcher Bias</u>
    b) Information Overload
    c) Typos
6) Which of the following methods is <u>not</u> a suitable method to increase credibility of a research?
    a) Open Science Framework (OSF)
    b) Open Notebook Journals
    c) <u>Personal Diary</u>
    d) Triangulation
7) What does a p-value of 0.05 mean?
    a) <u>A statistical value</u>
    b) Shows a statistical insignificance
    c) Shows a statistical significance
    d) None of the above
8) What is p-hacking?
    a) Manipulation of result
    b) Manipulation of hypothesis
    c) Manipulation of data
    d) <u>Manipulation of method</u>

9) What is the most important part when publishing a scientific paper? (B)
    a) Reputation of a publication
    b) <u>Contents of the paper</u>
    c) Citation counts of the paper
    d) New findings in the paper
10) Which of the following is **<u>not</u>** a reason researchers publish on prestige journals? (D)
    a) To promote readerships
    b) Appropriate peer reviews
    c) Precise evaluation of articles
    d) <u>Lacking time to promote work himself</u>
11) If we erroneously conclude that motorists are more likely to honk at low status cars than high status cars, we
    a) have made a Type I error
    b) have made a Type II error
    c) would have made that conclusion 5% of the time if the null hypothesis were true.
    d) <u>both a and c</u>
12) After conducting a statistical test for a research, you conclude that the mean score of male participants differs significantly from the mean score of female participants. You have:
    a) have made a Type I error
    b) <u>Accept null hypothesis</u>
    c) Reject null hypothesis

References

Alexander Michael Petersen et al. (2014) Reputation and impact in academic careers. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4217436/

Abelson R. Beliefs Are Like Possessions. (1986) Journal for the Theory of Social Behaviour. 1986;16(3):223–250.

Björn Brembs, Katherine Button, Marcus Munafò (2013) Deep impact: unintended consequences of journal rank. Front. Hum. Neurosci., 24 June 2013 | Retrieved from https://doi.org/10.3389/fnhum.2013.00291

Chambers, C., & Etchells, P. (2018, August 23). Open science is now the only way forward for psychology. Retrieved from https://www.theguardian.com/science/head-quarters/2018/aug/23/open-science-is-now-the-only-way-forward-for-psychology.

Charles R. Ebersole, Jordan R. Axt, Brian A. Nosek (2016). Scientists' Reputations Are Based on Getting It Right, Not Being Right. Published: May 12, 2016. Retrieved from https://doi.org/10.1371/journal.pbio.1002460

Chris Woolston (2015) Recognition: Build a reputation. Nature 521,113-115 (2015) doi:10.1038/nj7550-113a. Retrieved from https://www.nature.com/naturejobs/science/articles/10.1038/nj7550-113a

Edwards, A. (2016, May 11). Perspective: Science is still too closed. Retrieved from https://www.nature.com/articles/533S70a.

Falk, R., & Greenbaum, C. W. (1995). Significance Tests Die Hard. *Theory & Psychology*, *5*(1), 75-98.doi:10.1177/0959354395051004.

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories. *Perspectives on Psychological Science*, *7*(6), 555-561. doi:10.1177/1745691612459059.

Haller & Krauss (2002). Misinterpretations of Significance: A Problem Students Share with Their Teachers? Methods of Psychological Research Online 2002. 7(1).1-20

John, L. K., Loewenstein, G. F., & Prelec, D. (n.d.). Measuring the Prevalence of Questionable Research Practices with Incentives for Truth-Telling. *SSRN Electronic Journal*, ,doi:10.2139/ssrn.1996631. Retrieved from https://doi.org/10.2139/ssrn.1996631

Langin, K. (2019). For academics, what matters more: journal prestige or readership? Retrieved from https://www.sciencemag.org/careers/2019/07/academics-what-matters-more-journal-prestige-or-readership

Lloyd Sealy Library. (2019). Open Access Publishing. Retrieved from https://guides.lib.jjay.cuny.edu/c.php?g=399417&p=3150307

Loannisis.J.P.A. (2005) Why Most Published FIndings Are False. Retrieved from https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124

Meyer MN, Chabris C.(2014) Why Psychologists' Food Fight Matters. Slate Magazine. 2014 Jul 31.

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. A. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. L. Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni (2015). Promoting an open research culture: The TOP Guidelines for journals. *Science* 348(6242):1422-1425.

Nosek B, Alter G, Banks G, Borsboom D, Bowman S, Breckler S et al.(2015) Promoting an open research culture. Science. 2015;348(6242):1422–1425. pmid:26113702

Oakes, M. (1986). Statistical inference: A commentary for the social and behavioral sciences, Chichester: Wiley.

Rosenthal,R. (1979). The file drawer problem and tolerance for null results. Psychology Bulletin, 86, 638-641, doi:10.1037/0033-2909.86.3.638

.Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, *22*(11), 1359-1366. doi:10.1177/0956797611417632.

"7 Threats to Science's Reputation." (2018) National Academy of Sciences. 2018. The Science of Science Communication III: Inspiring Novel Collaborations and Building Capacity: Proceedings of a Colloquium. Washington, DC: The National Academies Press. doi: 10.17226/24958.

3 - Mass collaboration in psychology and initial findings about low replicability

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Chun Hei Matthew	Yeung	https://www.researchgate.net/profile/Matthew_Yeung6	osf.io/qc54k	u3543715@connect.hku.hk	matthewyeung45@gmail.com	3035437159
Tin Shing	Wong	https://www.researchgate.net/profile/Wong_Tin_Shing	http://osf.io/6fq7m	u3549142@hku.hk	davidw.1014@gmail.com	3035470060
Yiu Kan	Tsang		osf.io/ecmf4	u3539097@hku.hk		3035534646
Mills	Owen		osf.io/d8t4e	u3565436@hku.hk		3035654551
Ngo Chai	Yu	https://www.researchgate.net/profile/Ngo_Chai_Yu	osf.io/pwy9d	u3547001@hku.hk	phyllis102.yp@gmail.com	3035470010
Ho Yin	Shek	https://www.researchgate.net/profile/Hoyin_Shek	https://osf.io/dkzqj/	ferdi@hku.hk	ferdishek@gmail.com	3035497874

Please be very clear about the contributions of each member of the team to this task:

First name	Last name	Contribution to this task
Chun Hei Matthew	Yeung	Chapter introduction, Quiz & summary
Tin Shing	Wong	Chapter introduction, Quiz & summary
Yiu Kan	Tsang	In-depth report: *Origins and why*
Mills	Owen	In-depth report: *What was done & initial findings*
Ngo Chai	Yu	In-depth report: *Origins and why*
Ho Yin	Shek	In-depth report: *What was done & initial findings*

Managerial summary

In this chapter, we will discuss *mass collaborations* in psychology research. We will dive into what it is, where did it come from, how does it help resolve the issues mentioned in the previous chapter and what does it mean for future studies. In a larger context, *mass collaborations* are the combined effort of many individuals towards a particular endeavour. This has been the case throughout history, and in the present day researchers have banded together in an attempt to alleviate the replication crisis. By working together, they are able to greatly reduce the statistical and methodological errors found in many completed studies and identify deeply rooted problems that plague the psychology field like false effects and absence of standardised practice. Mass collaborations like the Reproducibility Project: Psychology help pioneer collaborative and transparent practices amongst researchers in the field. This project that dates back to 2015, has provided many heuristics to identify problematic studies and how we can reduce our errors and greatly enhance the scientific accuracy of our researchers. Ultimately through *mass collaborations,* a new movement of replication studies that subject present research to more scientific rigor is now developing.

In depth report

Introduction

Mankind since the age of old have cooperated to resolve tasks deemed impossible alone. In a similar manner, when facing the replication crisis, researchers have joined together to tackle the problem at hand. As mentioned in Chapter 1 and Chapter 2, the main problems at hand are possible false effects, presence of inconsistent data and the absence of statistical and methodological rigor. The false effect produces numerous fraudulent and overblown scientific results which is worthless and inefficient for further studies to work on. Furthering this, are inconsistent data possibly derived from misunderstood statistical methods and practices that cause data cannot be measured accurately across studies. This also puts into question the rigor of these methods used and recorded as practices are not strictly enforced and scientifically followed. These issues have plagued the research field and have caused many studies to press forward on results that are not replicable and may ultimately undermine the research built from it.

Researchers opted to work together through *mass collaboration* as a means to scrutinize these fundamental issues plaguing the field. It is a collaboration model that large numbers of contributors and researchers work independently but collaboratively in the research (Tapscott & Williams, 2006). Through the mass collaboration model, individuals can implant their innovation and creativity to the research project by running a self-organized collaboration framework and share their ideas to produce dynamic products (Ghazawneh, 2008). Typically, it usually takes place in the internet by web-based collaboration technologies. It aimed to subject existing research in the field to methodological rigor and peer review whilst maintaining a high level of transparency throughout replication. Also, it allowed facilitation of more accurate and precise practices.

What are mass collaborations and where did they come from

Mass collaboration involves people that either working as a group or separately, and exchanging ideas to bring advancement to the development of community (Collins, 2016). People with specialized occupation or interests join together and strive for a more productive and beneficial society. With the invention of Internet, it has created more methods of investigation and new medium for revealing the way that our world operates. In the world wide scientific web, there are various means to enable to flourish mass collaboration such as by digital libraries, web communities, massive open online courses and also crowdsourcing etc., (Collins, 2016).

Back to the development of cultural evolution, sentient animals started to learn how to pass their knowledge and doings to their offsprings. Homo sapiens learnt to make stone tools for hunting and fishing million years ago. From Ridley (2010), humankind used barter to exchange for goods is one of the greatest organisations, as when more people trade resources, the more division of work executed, thus, the produce can be better off. Such worksharing leads to a cycle for even more increasing trade and learning. The merchandises are becoming more and more profit-making and with a growing amount of people joining the effort of specialization of what they can foremost produce.

Next, the invention of cities are noted to aggregate people for specializing in professions and commerce. With more research developments, budgets and patents, more inventors and creative professions will be arisen and making the city more creative (Johnson, 2010). The significant diversification of ideas help generating new propositions and innovations. Take Silicon Valley as an example, it is a warm bed for startups and it provides a rich base of experts and knowledge. Firms with large capital and advanced technology are in high proximity that helps notions and proficiency to spread from firms to firms, such as the user-interface design be escalated from Xerox to Apple then to Windows (Johnson, 2010). The concentration of industries build a centre for specialized inventions and services as the materials are well supplied and organized. At the same time, the market is in good merit. Information flow effortlessly and lead to a close-knit community that to uphold a powerful environment for business. Silicon Valley, acting as a “networks of practice”, linking different communities and develop rapidly (Wenger, 1998).

Later on, the new phenomenon of the worldwide scientific community has played a major part in constructing the mass collaboration. For example, scientific publications in various fields demonstrate data analysis on methodology and grasp of prior literature with the help of peer reviews. Also, regular meetings are held by scientists and professionals to present their ideas and address questions about their findings. The feedback received provides better argument and support their experiments. Governments and other private enterprises will eventually fund for further development and assist more in the industry. When the internet is a great tool to spread mass collaboration through smartphones and the web community, it is a new norm that scientists of different fields are able to join together on projects through the net. They can share experimental designs, data, preliminary results and applications (Finholt & Olson, 1997). They can also co-publish their work even if they have never meet each other face-to-face. In recent years, digital libraries widely support the mass collaborations by providing various input state-of-affairs. The large capacity of resources multiple access to digital resources.

With a high-speed expansion of internet, social media platforms ascend and allow masses to put up messages and spreading news in different mediums. Shirky (2008) used Flickr for instance, it is an important agent for showing episodes before the traditional news outlets able to get the information. For example, photos of the destroyed places and missing ones first appeared in Flickr with the hashtag “tsunami” during the event of Indian Ocean tsunami (Shirky, 2008, p.36). It is a source for audience to report issues and analyze incidents happening all around the world straightaway. Also, the social networking sites such as Facebook, Instagram, Twitter etc. work as an organization to gather people for flash mobs or protests. This shares similar concept with crowdsourcing. It furnishes channels to collect funds for victims under political, social or environmental campaigns, or approaches to support new start-up firms. Kickstarter, one of the famous websites of crowdfunding, have been supporting entrepreneurs to accumulate money on their projects. People with similar ideologies that appreciate one's work will donate and the ideas of the original owners can be spread.

According to Collins (2016), Wikipedia is also one of the prime examples of mass collaboration. It is an online platform that people can pursue individual interests and efforts on providing information. One can focus clearly on how to make certain research more accurate and informative by gathering different resources from the Internet. Therefore, people working singly can contribute to the greater good as different participants are actively improving the information. Heavy above-head administration or management of corporation is not required. The content of the encyclopedia varies everyday and the topics people interested are in a fast-changing pace. The special feature of diverse languages connect people from all places and able to contribute and collaborate on the same subject matter.

With the looming of Internet, there is an uncontrollable outburst of mass collaboration circulating all around the world effectively. The new form of collaboration is to spread new and modern ideas to produce greater innovations. WIth the complex system of community, people need to work together and aware of the strategies to cope with multiple challenges. By keeping track of the changes of situation day by day, opportunities and risks appear and we should be highly adaptive to the unpredictable environment. Yet, the problem of hierarchy may be emerged. With numerous information flow in the market, it is hard to elaborate or organize without agency that holds much power. Therefore, elites would have monopolize the community. From Friedman (2004), he stated that power is being allocated by those who got crazy ideas. By speeding up innovations and inventions, more ideas came out in touch and more individuals can participate in the action of mass collaboration. The process can be way further from any prior success.

Mass collaboration: the Reproducibility Project

For these reasons, the first mass collaboration to investigate reproducibility in the Social Sciences was started in November 2011, culminating in a 2015 report, and was called the Reproducibility Project: Psychology (RRP) (Camerer et al., 2018; Open Science Collaboration, 2015). Not only was this the first collaboration of its kind in the history of the field, but it was of particularly grand scale, with 270 authors contributing to the replication of 100 studies published in journals in the year of 2008. Furthermore, the collaboration produced a protocol to ensure a standardized methodology of designing high quality replications would be carried out throughout the research (Open Science Collaboration, 2012). By adhering to this protocol, replications could be kept at a high level of fidelity to the original studies, and biases that could interfere with results are minimised at all stages of research. These experiments, and justifications for any deviation from protocol within them, were all reported on the Open Science Framework, to ensure maximum transparency occurred.

Of the original 100 studies, 97 had found a positive result (Open Science Collaboration, 2015). This is not surprising, as publication bias naturally tends to lead to only positive findings making their way into journals. However, the RRP concludes that upon replication, many of these experiments fail to provide equally compelling evidence.This conclusion is based on a number of quantitative and qualitative data. To start, one simple – albeit reductionist – tool to analyze how strongly the replications support the initial findings is to simply compare whether the same statistically significant effects were found in both the original study and replication. Because there is always a chance of a false positive in statistical tests, it is to be expected that there are only around 89 positive findings in these replications, rather than the 97 original found. However, the actual number of positive findings is much lower, with only 35 found, a reduction that is itself statistically significant ($P < .0001$). It is possible that this is due to the rather arbitrary marker of $P < .05$ – perhaps the replications are simply underpowered and so cannot detect the existence of an effect to such a precise level. However, an inspection of the distribution of P-values

(Fig 1.) does not support this argument, as it appears there is a wide distribution in P-values in replications that did not find an effect. Another statistic to look at is calculated effect size – if the findings are reproducible, it is to be expected that similar effect sizes are to be found. Similarly to the first findings, when comparing the reported effect sizes to the 95% confidence intervals, it is suggested that the failure rate of replication is 52.6%. A further analysis of effect sizes provides another alarming finding – of 99 studies for which both the original and replication effect sizes could be calculated, 82 showed a greater effect size in the original rather than the replication. Another way of investigating the success of the studies is in a meta-analytic sense, where an effect size is estimated based off of the findings of both studies. Of the 75 meta-analyses that could be generated, 51 (68%) succeeded in replicating, in the sense that the 95% confidence intervals did not contain an effect size of 0. Although this is more favorable regarding the replicability than the previous findings, it is only effective if the original data can be trusted as unbiased and replicable in and of itself, and since the earlier statistics do not support that claim the meta-analytic method should not be regarded as reliable. Finally, a simple subjective method was used, asking the replication team to report whether or not they regarded their replication as a success or not. In this case, 39% of studies were deemed successful. Although no single measure used here is robust enough to be used as the singular determinant of whether a replication succeeded or not, there is enough evidence here to suggest that many more studies than statistics would suggest report false positives, and even more report effect sizes that are larger than can be replicated.

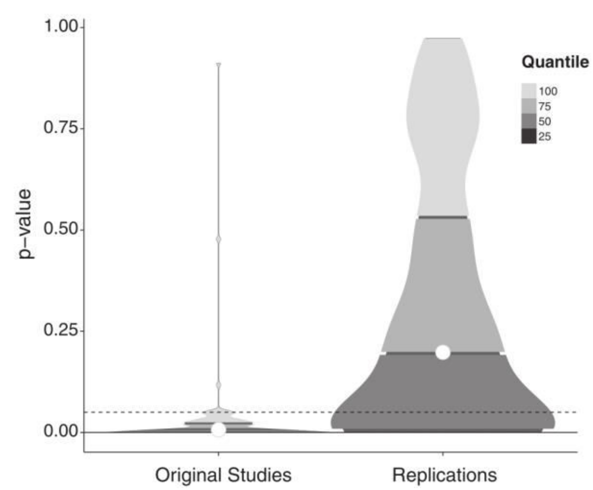


Figure 1. Density plots of P-values for the original 100 studies and their replications (Open Science Collaboration, 2015).

A further point of interest in this collaboration was an investigation into the correlates of replicability in the chosen studies, and although this is a less stressed part of the study, it is nevertheless important, as it provides a way of determining what studies in future are most or least likely to replicate. For instance, the P value of an effect is weakly negatively correlated with success of replication, and reported effect size is weakly positively correlated, so it seems findings that are less 'marginal' with more strong reported findings are more likely to replicate. Surprising and important results were also found to be less likely to replicate. In conclusion, the collaboration provided a number of useful heuristics to judge when a replication is more or less necessary.

Consequences for subsequent replications

As the Reproducibility Project requires conducting replications to obtain estimation of reproducibility of original studies. Open practices of the replications are ideally to increase the accountability of the replication team and the quality of designs and results (Open Science Collaboration, 2012). The Open Science Collaboration (2012) stated that the project follow a standardised protocol in attempt replications in order to minimise any irrelevant variation in both data collection and reporting methods, and also to maximise the quality of replications. Replications that are statistically significant or obtaining a directionally similar but not significant are considered as successful replications. Several factors might cause replications fail to replicate, the first factor is simple Type 2 error, where some findings will fail to replicate purely by chance. Failures of replication can also occur if the original effect is false; the actual effect size is lower than the original report; the original design of the study is flawed; or the replication methodology is different from the original study. Most importantly, failure to replicate an effect does not mean the original effect was false, many factors might contribute to the failure of the replications. For these reasons, the reproducibility project aims to investigate factors such as replication power, the evaluation of the replication study design, samples and effect sizes by the original authors in order to improve reproducibility (Open Science Collaborative, 2012).

So what is up next with the Reproducibility Project? Implications? The standardised protocol to replicate study of the project may provide a high reproducibility replications which might enhance the confidence in conventional research and peer review practices (Greenwald, 1975; John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011). Low quality replications might give ideas on reflection on the quality of standardised practice, motivate investigation on factors that affect reproducibility and ultimately leads to changes in practice of replications and standardised protocols as well as changes in publishing standards(Open Science Collaborative, 2015). Some people may think that the low reproducibility rate will damage the reputation of psychology study in the field of science. Such that it is important to investigate factors that might contribute and affect reproducibility rate and to address it. The reproducibility project provide the first open, systematic evidence of reproducibility from samples of studies in the field of psychology study (Open Science Collaborative, 2015). The project aims to maximise the generalizability of the replication results, however, not all studies are able to be replicated. For example, more resource intensive studies were less likely to be included than less resource intensive studies. Concerns on the reproducibility rate are widely spread in science study as reproducibility is a sign of credible scientific evidence, replication can increase when findings are reproduced, and promote innovation when they are not. This project provides accumulating evidence for many findings in psychological research, and suggests that there is still more work to do to verify whether we know what we think we know.

...

Quiz

Write 10 easy quiz 4 multiple choice questions to ensure reading and understanding of this chapter:

(Note highlighted are the correct answers)

What are the fundamental issues found as mentioned in 3.1?

1. What are the fundamental issues found by researchers as mentioned in section 3.1?
    - **a. False effect, Inconsistent data, Statistical and Method rigor**
    - b. Data falsification, corruption, publication manipulation
    - c. Placebo effect, Plagiarism, Non-disclosure of methods
    - d. Varying data, Publisher bias, partial method disclosure

2. What was the first mass collaboration on reproducibility called?

    - a. Replication Reproduction Research
    - **b. Reproducibility Project: Psychology**
    - c. Production of Reproducibility
    - d. Reproducibility Mass Collaboration project

3. How many authors did the RPP involve?

    - a. 1500
    - b. 50
    - c. 300
    - d. 270

4. What was the conclusion of the RPP?

    - a. The studies showed that replicated methodology was not accurate
    - b. The studies showed insignificant difference
    - c. The studies replicated fail to provide equally compelling evidence
    - d. The studies replicated were successful

5. Which is **not** a way to analyze how strongly the replication supports the initial findings?

    - a. Post-hoc T-test
    - a. Comparison of statistically significant effects
    - b. Calculated effect size
    - c. Replication team evaluation

6. According to Ridley (2010), what was mankind's greatest feat of organisation?

    - a. Trade

b. Systematic farming
c. War tactics
d. Constitution and the judicial system

7. With reference to the chapter, what is Wikipedia an example of?

    a. Online information hub
    b. Quick knowledge database
    c. Mass collaboration
    d. Replication study

8. How does Open Science Collaboration (2012) reduce irrelevant variations in replications?

    a. Certification by respective departments
    b. Additional researchers within the same project
    c. Extensive peer review
    d. Strict and Standardised protocol

9. What is a factor that can reduce the plausibility of a replication?

    a. High cost of resources
    b. Number of researchers
    c. Placebo effect
    d. Data falsification

10. How many studies (out of the original 100) in the RRP found positive results?

    a. 80
    b. 98
    c. 97
    d. 77

References

1. Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Altmejd, A. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour, 2(9), 637.
2. Collins, A. (2016). A Brief History of Mass Collaboration: How Innovations Over Time Have Enabled People to Work Together More Effectively. 10.1007/978-3-319-13536-6_2.
3. Finholt, T., & Olson, G. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboratories. *Psychological Science*, 8, 28-36.
4. Friedman, T. L. (2004) *The world is flat: A brief history of the twenty-first century.* New York: Farrar, Straus and Giroux.
5. Ghazawneh, A. (2008). Managing Mass Collaboration: Toward a Process Framework.
6. Johnson, S. (2010) *Where good ideas come from: The natural history of innovation.* New York: Penguin.
7. Many labs 4: Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.
8. Open Science Collaboration. (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. Perspectives on Psychological Science, 7, 657-660.
9. Ridley, M. (2010) *The rational optimist: How prosperity evolves.* New York: HarperCollins.
10. Shirky, C. (2008) *Here comes everybody: The power of organizing without organizations.* New York: Penguin Books.
11. Tapscott, D., & Williams, A. D. (2006). Wikinomics: How Mass Collaboration Changes Everything, 2006. *New York: Portfolio*.
12. Wenger, E. (1998) *Communities of practice: Learning, meaning, and identity.* New York: Cambridge University Press.

4 - Reactions/debate about low replicability or the “replication crisis” - is there a problem?

Team names and contribution

First Name	Last Name	Researchgate Profile	OSF Profile	Institutional Email	Personal Email	Student ID
Amy	Wu	https://www.researchgate.net/profile/Amy_Wu6	osf.io/2xame	Amywu128@connect.hku.hk	1395307963@qq.com	3035554189
Cheuk Lun	Chiu	https://www.researchgate.net/profile/Cheuk_Lun_Chiu2	http://osf.io/6fq7m	ccljoe@connect.hku.hk	chiucl95@gmail.com	3035470060
Sirui	Lu	https://www.researchgate.net/profile/Sirui_Lu3	osf.io/ecmf4	roselu@connect.hku.hk	2023362678@qq.com	3035534646
Susan	Papp	https://www.researchgate.net/profile/Susan_Papp	osf.io/d8t4e	sp710613@connect.hku.hk	susan.papp@ymail.com	3035654551
Svetla	Manolova	https://www.researchgate.net/profile/Svetla_Manolova	osf.io/pwy9d	svetla@connect.hku.hk	svetla.n.manolova@gmail.com	3035651676
Yiming	Wu	https://www.researchgate.net/profile/Yiming_Wu21	osf.io/d4b3e	wym98@connect.hku.hk	wuyiming98@126.com	3035449346
Yuge	Wang	https://www.researchgate.net/profile/Yuge_Wang8	https://osf.io/uqpjm/	u3556432@connect.hku.hk	rainy1021@outlook.com	3035564328

Managerial summary

When the surprising findings regarding low replicability mentioned in Chapter 3 first came out, they triggered an upsurge of discussion and practice in the field of psychology as well as other science realms. The debate on the replication crisis received mixed reactions as to its' nature and stirred a lot of debate among the scientific community. In this chapter, we will proceed from the previous findings to arranging and researching the thread of the heated debates on "replication crisis" ever since 2011 and 2012. Following the debates, this chapter will present the main arguments regarding the statistical, methodological and contextual perspectives. However, the process should not be confined to the pure debates and arguments. Instead, this chapter will move on to the discussion of solutions to such science dilemma, aiming at revealing its veil and deeply reflect on the way of doing science both in the past and future.

In depth report

Psychology in the eye of the hurricane: An open letter and serious debate

The failure of mass collaboration and initial findings about low replicability had become the small stone that made a big splash in the field of psychology. In the following years, the ripple effect did not stop but triggered the new debate regarding whether there was really a “replication crisis”.

Among the chaos of accusations and uncertainty, suspicions had also extended to the central phenomena of social psychology: the *priming* effect. The basic idea of priming studies is that the activation of certain concepts may affect people’s subsequent processing of other information or the behavior, even without the awareness of the influence (Vadillo, 2012). While in the field of social psychology, the process of priming is mainly defined in terms of how the events and actions, which trigger facilitative effects on subsequent associated responses, influence the activation of stored knowledge (Molden, 2014). One of the most cited experiments about the priming effect was conducted by John Bargh (1996), which found that the activation of concepts related to the elderly (such as words “slow”, “old” and “gray”) made the behavior of the participants resemble that of the elderly. For more than fifteen years, the priming effects had been firmly grounded in the core concepts of social psychology and researched extensively. However, in 2011 and 2012, questions had been raised about the robustness of such priming results.

Doyen, Klein, Pichon, and Cleeremans (2012) published their article in which they tried to replicate Bargh’s (1996) experiment mentioned above, but failed to replicate the original results. In their replication, Doyen and his colleagues made several adjustments. For instance, they doubled the number of volunteers, and measured the time with infrared sensors instead of stopwatch. They also recruited four experimenters who are not aware of the experimental purpose to carry out the experiment. In the following modifications, Doyen and his colleagues found that the only way for them to get the same results as the original study was by disclosing the experimental design and expectation to the participants. In regard to those unsuccessful results, Doyen and his colleagues expressed their suspicion that in Bargh’s original experiment, the experimenters could have primed the participants with their own expectations. However, their replication work and relevant suspicion drew Bargh’s attention and in March 2012, he replied with harsh criticisms against Doyen’s team, PLoS (the journal that published Doyen team’s paper), and some online science-journalism blogs including Ed Yong’s, claiming that the “priming effects replicate just fine”. Such reply led to strong reactions from people who were involved or who were reading their websites.

In Ed Yong’s next post, he replied to Bargh’s (2012) criticism point by point, emphasizing Doyen *et al.*’s (2012) improvement in measurement objectivity and sample size. However, he admitted that “a single failure to replicate doesn’t invalidate the original finding” and denied the suspicion to the entire concept of priming. In the end, he called for plenty of more replication, as questioning towards replicability was regarded as a significant problem in psychology or science as a whole.

The serious debate of priming effect was drawn to an end by an open letter from Nobel laureate Daniel Kahneman in September 2012. In his open letter, Kahneman expressed his deep concern about this situation and invited all the psychologists to work collectively to address this problem. In Kahneman's view, the storm of doubts originated from various sources, including the recent exposure of fraudulent researchers, multiple failures of replication, as well as the existence of the file drawer problem. He stated that those doubts undermine two major methodological pillars of the social psychology field: first, the preference for conceptual over literal replication and second, the use of meta-analysis. The former one implies social psychologists' preference on adopting conceptual replication in research, in which instead of replicating experiments with exact same data or conditions, they lay emphasis on testing the experiment's underlying hypothesis with modifications on methods and measures (Nussbaum, 2012; Diener & Biswas-Diener, 2019). The latter one lays emphasis on pooling data from various studies of a well studied topic, in order to get one combined answer and gain greater confidence in the research results.

Although not being strictly regarded as a social psychologist, Kahneman confessed himself as a general believer in the mechanisms of priming. His book "Think Fast and Slow", which was published in 2011, devoted an entire chapter to explain the mechanisms of priming and Kahneman emphasized that in study of associated memory, priming research could be adopted as a new approach (Vadillo, 2012; Kahneman, 2012). To support previous social priming results, he wrote "disbelief is not an option" in his book (Stafford, 2017). However, this statement gave rise to the broad discussion and query towards the reliability of social priming research. In such case, he expressed the concern that the doubts or crisis in the fields will in fact affect the large population of colleagues who, just like himself, had accepted the surprising results as facts and devoted themselves to the related researches. And the replication crisis as well as debates had cast a shadow on their academic career. His concerns became real that in 2017 Schimmack, Moritz and Kesavan reviewed chapter 4 of *Thinking Fast and Slow*, picking out the references with evidence for social priming and re-calculated the statistical reliability. They reached the conclusion that it was nearly impossible to get the pattern of 100% significant results, which was surprising but accepted by Kahneman. He admitted that he "placed too much faith in underpowered studies…" and "have changed my views about the size of behavioral priming effects – they cannot be as large and as robust as my chapter suggested." (Stafford, 2017). Back to his open letter, Kahneman (2012) stated that the reason for him to write the letter was that "seeing a train wreck looming". The first expected victims would be the young people on the job market whose controversial and suspicious academic background left them in disadvantaged competition position. The second concern was that the lesson of subliminal perception and dissonance reduction, which "went into prolonged eclipse" after similar attacks on replicability, would be repeated.

In order to solve those issues, Kahneman's (2012) main point was that work should be done through collective effort, that an association with board members of prominent social psychologists should be found and the first mission was to re-examine the replicability of priming results. Then there should be a follow-up protocol helping to guarantee responsibility among colleagues outside the field. Kahneman's proposed protocol was the collaborated replication by groups of laboratories with good reputation. For example, in the group of five laboratories, each picks one specific effect in social psychology and the laboratories do mutual replicas in a cycle. The replication lab should send members to the research center that originally published the study to make sure the replication is done under the same conditions as the original experiment. In return, the original research center should also send members to vet the procedure of the replicating lab. Moreover, the replication requires large number of subjects in case encountering the statistical power problem. With technology support, the entire replication process should be recorded (i.e. videotaping) to ensure every detail is preserved and can be copied by others. Last but not least, Kahneman (2012) suggested that the researchers and authors should commit themselves to publish the work beforehand, and make the replication data public that everyone can analyze them.

After Kahneman issued his open letter, Ed Yong raised some questions in response to his call. Nobert Schwarz (2012) shared his answers publicly to reply to questions down the road. He believed that the skeptical attitude towards priming studies should be attributed to the more surprising findings in this area than other domains, instead of saying that “work in this area is more or less replicable than work in other areas.” He illustrated this explanation by analogizing it to the climate change debate: the consensus of the vast majority of professional psychologists in this area was challenged by some persistent skeptics raised by a few people who were not as familiar with priming studies as the former psychologists did. Their skepticism aroused vast concerns about the replication crisis, which was not the case. The broad doubt was incompatible, and much about these phenomena remaining to be further examined.

Ed Yong (2012) responded to Kahneman’s open letter and Kahneman’s answers by reporting, “Nobel laureate challenges psychologists to clean up their act”. He explained that the skepticism was due to “failed attempts to replicate classic priming studies, increasing concerns about replicability in psychology more broadly, and the exposure of fraudulent social psychologists who used priming techniques in their work”. He mentioned other psychologists’ worries that the debate about priming was turned into self-defensiveness, missing a good chance to conduct a scientific discussion about data. Afterward, Daniel Kahneman complained about the damaging title of this piece by Ed Yong. He emphasized that what he was trying to convey was a friendly suggestion to deal with the image problem faced by his colleagues.

By reading through these debates, we can see that people show different attitudes towards the disability to replicate some research. Some people regard it as a severe replication crisis, doubting the credibility of psychology. Some scientists view it as an effect of different findings under different time and technology scales. Some people, like Schwarz, believe there is no need to feel anxious about broader scientific trouble. Many, however, agree with Kahneman’s suggestions that more replication should be carried out, and the undermining effects of publication bias should be considered. It is meaningful no matter to provide evidence for rehabilitating the field and protecting the scholars’ reputations, or to deal with the replication crisis, which was believed by some people to have been happening.

Views from the Opposite Sides: Is there really a Replication Crisis?

Replication Crisis Doesn't Exist ——Depending on the *Ways* to look at it

Having settled on what the debate(s) essentially is/are, who is involved and how the opposing sides reason, common sense dictates that we take a step back and try to look at this dispute from different perspectives, clearly, to be able to judge for ourselves whether the provided arguments are plausible.

As the aforementioned narrative explained the reproducibility crisis "...postulates that a large growing proportion of studies published across disciplines are unreliable due to the declining quality and integrity of research and publication practices, largely because of growing pressures to publish and other ills affecting the contemporary scientific profession" (Fanelli, 2018, p. 2628). However, as the term "reproducibility" itself is a broad concept, comprising aspects, such as statistical problems, experimental design flaws, methodological issues, even suspicious ethical practices (Cyranoski , & Ledford, 2018). It is difficult to define which should be the specific matter of attention in the argument. Whether there should be a specific criteria, which would indicate if a paper is replicable or not or even how many of the priorly mentioned aspects of reproducibility have to be violated to 'condemn' a paper as non-reproducible without major revisions? A simple example would be that of fields such as computer science and mathematics that do not experience statistical problems or experimental design issues but still experience replication failures (Redish et al., 2018).

Bearing this in mind, Redish et al. (2018) have suggested a new perspective that centers reproducibility as a key part of empirical exploration in any scientific field. They argue that researchers should not overlook the replication crisis debate as such but instead should view reproducibility as "... a core feature of the process of scientific inquiry". In other words, not being able to replicate certain findings is not essentially a "failure" but is an opportunity to further explore the different dimensions and limitations of an existing phenomenon. Failures to replicate (or i.e. failure to generalise a study's findings) would be the drive for scientists to explore the underlying factors, responsible for the different outcomes since after all "...the key to science is that every answer opens up new questions" (as cited in Redish et al., 2018). Realistically, this would put emphasis on key points people tend to forget such as: i) science needs time to "reconcile conflicting results"; ii) reproducibility failures are a part of research, which do not imply that 'bad science' has been conducted; iii) more metanalyses should be conducted and conclusions should not be based on single research papers (Redish et al., 2018). Integrating the conflicting observations will then most probably lead to the formation of a consistent theory. Therefore, the stages of this empirical process themselves would naturally promote better open-science policies and transparency.

More polarizing views on the topic have been expressed, however. Researchers such as Daniele Fanelli have argued that in reality the reproducibility crisis is exaggerated or at least "partially misguided" (Fanelli, 2018). According to him, recent evidence have disproven the widely spread belief that questionable research practices are distorting the majority of the literature, making generalizations unreliable and findings controversial. Instead scientific misconducts effects "while nonnegligible, are relatively small and therefore unlikely to have a major impact on the literature". (Fanelli, 2018; Fiedler &

Schwarz, 2016). Major concerns such as p-hacking(subconscious or conscious manipulation of gathered data for an experiment so as to produce a desired p-value) and a high rate of false-positives, despite being more common in scientific disciplines “appear to be contained” and do not seem to have any significant impact on conclusions from meta-analyses, therefore implying that the majority of studies are measuring true effects (Fanelli, 2018). A logical argument against this nevertheless could be pointing towards the amount of publication biases present in research. Meta-analyses have been conducted measuring the strength of evidence supporting this consideration (Fanelli, Costas, & Ioannidis, 2017). Nonetheless the analysis itself could be a “victim” of publication bias and moreover publication bias is thought to vary substantially across different fields.

Interestingly enough, as concerns about science’s reliability have been rising, there has been no evidence suggesting that questionable research practices are proportionately increasing as the stigma around the topic states (as cited in Fanelli, 2018). In contrast, evidence is showing that published studies are getting more complex and detailed and are subject to more precise statistical analyses. This could potentially explain the rise in reported positive statistical results (as cited in Fanelli, 2018), as their authors might be omitting the negative results from the abstracts but instead including them in the following thorough investigation.

Having all these empirical arguments (as science fundamentally emphasises on empirical evidence), we can logically conclude that the reproducibility crisis narrative is not supported enough factually to be manifested as a “crisis”, “discrediting” whole fields of research. While science is indeed facing problems, such as questionable research designs, p-hacking, biased and falsified results or even underpowered statistical analyses, it is certainly not a new challenge as there always have been doubts as to the ecological validity of findings (Mullane & Williams, 2017) (but as aforementioned this is what science is about essentially - implementing contrasting theories in order to objectively judge the qualities of a phenomenon). Instead of thinking of this as a major downside, it is important to recognise the fact that contemporary science is actually facing new opportunities, revolutionising the way we do science. As a result of the “crisis”, open-science and transparency have shown to be fundamental as to the resolving of the debates, but moreover large-scale assessments and meta-analyses have proven to be an indispensable criterion of the accuracy of research findings that it is becoming the norm. As Fanelli (2018) states that our contemporary situation is actually “... a narrative of epochal changes and empowerment of scientists”. Therefore, in conclusion, one could argue that the key to progress lies not in the debate as such but paradoxically in the implementation of opposing views, since it appears that the truth lies midway.

Apart from the perspective of looking at the “low reproducibility” that leads to the discourse of “replication crisis” in the sciences, the existence of “low reproducibility” and “replication crisis” may be due to the issues of error, power and methodological endorsement. Gilbert, King, Pettigrew and Wilson (2016) mentioned that the reproducibility of psychological science is however quite high, which was different with the conclusion made by The Open Science Collaboration (OSC). Gilbert *et al.* (2016) compared OSC’s replication with The “Many Labs” Project (MLP), which found a higher reproducibility rate of the original articles. First, in terms of error in the study, OSC did not take all possible errors into account when setting the benchmark for error in the replicated study, which was compared with the results that adopted a tighter benchmark. Therefore, it might lower the replication rate of the original study. OSC just simply assumes that sampling error is the only error in the data, implying that they assume the only difference between their replication and the original study is the sample they drew from the original population. However, Gilbert *et al.* (2016) mentioned that there were many sources of errors that OSC did not take into account, such as different population and procedures from the original study, which made the infidelities allowed in OSC's replication. With the wrong assumption of error existed in the replication, OSC would therefore adopt a loose benchmark for error in their replication when compared with the original one having a tighter benchmark. However, when considering all these possible errors and infidelities, MLP's replication implied that more than 34% of OSC's replication should fail by chance despite the true effect described in their replication. Therefore, OSC's study about "low reproducibility" of the original finding was still inconclusive with less statistical support.

Second, OSC's replication also suffers from the criticism of having a low power due to its little attempts for replication. Compared with MLP which replicated each of the original study for 35-36 times and then pooled the data, OSC just had one attempt for each replication, presumably leading to lower power. Therefore, when adopting MLP's method in replication, they attained an 85% of replication rate. Finally, Gilbert *et al.* (2016) argued that the "low reproducibility" was due to lack of methodological endorsement from the original author. It was found that endorsed protocols were almost four times to produce successful replication as the unendorsed one, implying that fidelity of replication to the original finding is important for a successful replication. This therefore explains that "replication crisis" is just statistical issues of error and power, as well as the issue of methodological endorsement. It thus imply that crisis of low reproducibility can be solved if these issues are eliminated from the replication.

Replication Crisis Does Exist ——Arguments from the *Methodological*, *Statistical* and *Contextual* Perspectives

Methodological

The "replication crisis" may be just a debate of opinion and perspective, avoidable methodological and statistical issues, and unavoidable contextual differences. However, low reproducibility of scientific studies undeniably posts a crisis in the research of social and life sciences, which is something that we as (future) researchers cannot ignore. Rather than arguing about whether a "replication crisis" exists, it is far more important for us to analyze the reasons and problems behind it, also the focus of this part.

The replication crisis exists in the methodological aspect. Many factors can influence the result of a replication. According to Schwarz and Clore (2016), the similarity of the procedures used in a replication and the original study is a significant predictor of whether results will be replicated since in the OSC's (2015) reproducibility project, 11 replications used procedures that the original authors considered inappropriate prior of data and 10 of them failed. However, although other 89 replication's experiments replicated in the way that original authors considered appropriate, the success rate is still unsatisfactory which only less than half the experiment is replicable. What are the problems? Gigerenzer (2018) used strategic-game hypothesis to explain the low replication rate. Simply put, it can be attributed to the low transparency of the original study, which allows the researchers to manipulate and publish the desirable data only. This perspective is explained in detail by Schwarz and Clore (2016).

First of all, the strength of manipulations will influence the result of the replication. During the experiment, some of the researchers may try to induce specific moods and observe their subsequent effect. However, the comparability and strength of the experimental manipulations are more often assumed than assessed (Schwarz & Clore, 2016), which means that in some cases, the intensity of the mood will vary with each person. The converge of those differently intense moods will cause a broad range of variables, which may cause the results of the replication to be different from the original one. These differences could cause replication to fail.

Secondly, the heterogeneity of the participants should be taken into account, especially with the replication, which includes large N (Schwarz & Clore, 2016). The larger sample, may contain more people from different places, with a different culture, growing up in different environments, which will influence the diversity of the result. This situation will increase the variance and reduce the effect-size estimates for the experimental treatment. The replication experts may not determine whether the replication meets the condition of an informative comparison. As a result, replication may fail.

Thirdly, the partial report of the dependent variables and the authors' intuitions in the original article might cause the failure of the replication. According to Fiedler and Schwarz (2016), failing to report all of a study's dependent measures, a high admission rate, may cause intentional concealing of unwanted results regarding the test hypothesis. This situation may predetermine an unsuccessful replication, which cannot obtain the same effect as the original paper. At the same time, the research practices addressed in the survey constitute a convenience sample of research practices that may have been selected based on the authors' intuitions. This could lead to participants potentially not being representative enough of the investigated population, leading to the failure of the replication.

Last but not least, the original experiment may contain some problems in the survey design period and report period, which may cause data disturbance (Fiedler & Schwarz, 2016). In the survey design period of the original experiment, there might be some questionable research practices (QRPs), which implies the potential presence of ambiguous questions or the response format obscuring the intended communication of self-report, influencing participants' responses. Unfortunately, the QRPs are common in psychology, with occurrence rate up to 100%, which causes more difficulties in successfully replicating these experiments. The results of the experiments also could differ if the replicators add more dependent variables since it increases the level of noise in its experiments (Schwarz & Clore, 2016). Also, in some experiments, the survey might mistake the proportion of individuals that ever engaged in behavior as a measure of the behavior's prevalence can lead to misunderstanding. Because of the existing problems in the original article, when the replication amend the error, the result will be different, which leads to the failure of replication.

Statistical

Apart from the misguided methodology used by researchers, Gigerenzer (2018) also argued that researchers' internalization of the statistical ritual (null ritual) in contemporary scientific study, which may lead to replication delusion, the illusion of certainty and Bayesian wishful thinking, may further fuel the replication crisis. Gigerenzer (2018) condemned that most of the researchers treated some statistical rules or steps (some even having loopholes) as ritual and thus neglected other good scientific practices. Therefore, implying that the researchers do not care about the statistical power due to its irrelevance to null ritual.

The null ritual can be attributed to the non-alignment of fusion of Fisher's and Neyman and Pearson's hypothesis testing, leading to several problems in hypothesis testing. First, specifying the null hypothesis but not an alternative hypothesis, which follows Fisher's practice, may eliminate the researchers' judgement of hypothesis. Fishers (1955; 1956, as cited in Gigerenzer, 2018) explained that his way of hypothesis testing aimed to test whether a hypothesis should be nullified instead of whether it presumes nil difference. We should bear in mind that two are different, which may greatly affect our statistical inference. Second, the use of fixed, conventional and inaccurate rounding-up significance level (e.g. p <.05), which is not supported by both, will make the researchers ignore the beta of the data (the probability of the test committing Type 2 error). This implies that the researchers following the statistical ritual may neglect the importance to make a judgment about balance between Type 1 and Type 2 error. Therefore, Gigerenzer (2018) believed that strictly following the null ritual leads to replication delusion/ fallacy among researchers who may not truly understand the p value but regard it as a ritual. Most of the them may wrongly assume that the probability of result can be replicated is 95% when having p=.05. However, we need to bear in mind that replication and original study are two independent studies. Just consider a simple example: a die twice, which could be fair or loaded (unfair), was thrown twice, and you got the result "one" for both times. Therefore, the p value of these two throws were .03 (1/36) assuming that it was a fair die. Then does it imply that there will be 97% of having two "one" for the next two throws? Therefore, when researchers have this logical fallacy, this may make the replication appear to be superfluous. Gigerenzer (2018) also argued that the replication had a low effect size because the researchers did not know about the power, which is due to the internalization of null ritual. For example, Bakker *et al*. (2016, as cited in Gigerenzer, 2018) reported that 89% of 214 author overestimated the power of their researcher designs. This implied most of the researchers actually had an inadequate understanding of statistical power, which may greatly affected the effect size of a research.

Second, the illusion of certainty and the Bayesian wishful thinking are closely related. Before explaining the details, we need to define an important statistical concept p value, which refers to *"the probability of data, assuming the null hypothesis is true"* (Gigerenzer, 2018, p.204). For example, p=.05 refers to the probability of obtaining the observed results or more extreme is 5% when the null hypothesis, including all assumptions made, is true. It is important for researchers to bear in mind that p value can only inform us the probability of the *given data* assuming the null hypothesis is true but not certainty - probability of a null or alternative hypothesis is true (Bayesian wishful thinking) (Gigerenzer, 2018). Also, most of the researchers still believe that significance level can inform us of the certainty of hypothesis (illusion of certainty). If researchers overestimate what can be inferred from p value and proven or disproven the original finding in this sense, replication is totally superfluous as their wrong assumption already overestimate the probability of successful replication. Therefore, an unsuccessful replication may seem surprising to researchers.

Contextual

Besides the arguments on the existence of statistical and methodological problems, one of the major prevailing debates is the contextual sensitivity in specific reproducibility, which is about the extent to which failures of reproducing certain results might also reflect contextual differences between the original study and the replication attempt. In this way, the contextual differences are also termed as "hidden moderators".

While some scientists believed that the hidden moderators were unlikely to influence the direct replication results, for the reason that the replication work adopted exactly the same methods used in the original study and in this case, context moderators were squeezed out (Srivastava, 2015; Roberts, 2015), many other scientists have committed to the research on the influence of context on individuals and discovered extensive evidence that contextual factors have altering effect on human behavior. One of the most impressive studies of replication's contextual sensitivity was done by Van Bavel and his colleagues in 2016, in which they recorded the Reproducibility Project's 100 original studies (OSF, 2015) and individually evaluated each research topic's contextual sensitivity. The study's results suggested that the research topic's contextual sensitivity was associated with replication success. More specifically, the correlation between the contextual sensitivity and the replication success was negative: $r(98)=-0.23$, $P=0.024$, indicating that the higher a topic contextual sensitivity was, the less likely the replication attempt to be successful.

The study adopted binary logistic regression models and linear regression models to test the contextual sensitivity variable as well as other variables which were thought as predictive towards replication success. There were two analysis models. Besides the contextual sensitivity, Model 1 includes 1) the effect size of the original study; 2) whether the original result was surprising; 3) the power of the replication attempt; and 4) whether the replication result was surprising, while Model 2 include the above four variables plus the 1) sample size of the original study; and 2) the similarity of the replication. The results showed that even after statistically adjusting for characteristics of those predictive variables, contextual sensitivity remained a significant predictor for replication success. Another interpretation of those results was that contextual sensitivity was a key factor in predicting replication success over other important methodological characteristics. A supplementary findings towards the association between the contextual sensitivity and replication success is the same across multiple psychology subdisciplines.

The study results also suggested that the original authors' endorsement could be effective predictors of future replication success when taking the contextual sensitivity into consideration. And when a research topic is considered with high contextual sensitivity, conceptual replication rather than direct replication could be adopted to avoid potential problems and improve the successful replication rate. In summary of all the results above, hidden moderators such as the contextual sensitivity should not be ignored in the replication studies but a measure of scrutiny should also be adopted.

Beyond the Debate —— Improvements, Reflections and Getting Closer to the Truth

One thing we have to be aware of, or gain most from this debate is that science is still advancing and there are still a lot that we need to do to improve. Publication works in the psychological field is still undergoing improvement and through this debate, it is crucial to understand the potential or existing flaws in psychological publications and how experts in the field should take these criticisms into account to tackle and build up where it is still weak and vulnerable to fraudity.

Some suggestions to improve the replicability

Increasing replicability is increasing the success rate and legitimacy of psychological research. One of the main reasons for nonreplication is the compliance for errors, especially on important facets of the research design (Asendorpf *et al*, 2013). Thus, the following methods of reducing sources of error written by Asendorpf *et al.* (2013) are proposed:

Increasing Sample Size

Reviews in the area of psychological studies have showed little improvement in sample sizes. As sample sizes increase, the statistical power increases and CI width decreases, this allows the results obtained to be more likely replicable as compared to a smaller sample size, where the statistical power is too low for replication standards. Thus, it is important to increase sample sizes as most of the psychological reports published showed underpowered results.

Increasing reliability of measures

One way to increase replicability is to decrease measure variance that can be attributed to error. As apart from increasing sample sizes, Cohen’s d and Pearson’s r, the two most common estimators of effect size can also be used to decrease error. However the standard deviations in these measures can be also affected by measurement errors, thus it is important to decrease measure variances that attributes to error. In order to do so, measure reliability should be increased, which is by definition, *the proportion of measure variation attributable to true variation* (Asendorpf *et al.*, 2013, p.111).

Increasing study design sensitivity

Building on the previous method of decreasing error variance, having a better control over methodological sources of error can effectively decrease error variance without restricting true variation. This can be achieved by distinguishing between random errors and systematic errors, and then effectively eliminate the sources of systematic errors where possible (Asendorpf *et al,* 2013).

Increasing adequacy of statistical analyses

Using the correct method of statistical analyses better suited to the study design is an important factor in reducing errors when analyzing data. Thus, it is important to test the appropriateness of method-required assumptions, taking into account all stimulus and treating them as random factors rather than fixed, as well as removing influences of covariates.

Avoid multiple underpowered studies

A common practice by scientists on multiple studies showing effects with underpowered results are conducted where they believe it tends to prove the reliability of the study, has been found to be a false assumption. In many cases, multiple underpowered studies also tend to show design and sampling biases, thus, excessive use of this practice should be discouraged and avoided in the scientific fields. Therefore, a pre-registration is important to enhance the transparency and reliability of the research design and replication.

Consider error introduced by multiple testing

It is more common to find a significant relationship between two variables in underpowered studies. Thus, multiple testing conducted on such experiments produce many different sets of significant relations and when the publication often exaggerate results, this often creates more fraud findings and confirmation bias in the field of psychology. One method to reduce this phenomenon proposed by Asendorpf et al.(2013) would be the adoption of a more modern variant (random permutation test) in replacement of the famous Bonferroni procedure that actually diminish statistical power. Along with it, he proposes, should be the non statistical solutions such as separating a priori hypotheses from the exploratory post hoc hypotheses.

The Calling for Role-Based Collective Work

In the field of connecting science to the world, not only psychologists themselves, but also journalists, reviewers, teachers of the related field, and institutions all play an important role in the replication crisis. Each role has their own responsibilities and methods to decrease nonreplication and filter out fraud or non replicable psychological findings. The aforementioned roles have their respective ways in doing so:

Authors

While the authors of the publications take credit for their work, at the same time, they should also be responsible to assess the replicability of their own research published. Authors of scientific publications should prioritize contribution to psychological research and increased transparency of research when publishing a report. By increasing research transparency, it allows scientific progress to accelerate, where experts of the same area can participate in published yet transparent publication discussions and engage in debates as well as replications.

Reviewers, editors, and journalists

The media and many other actors that contribute to the legitimacy and promotion of scientific publications often is a key drive in whether science is hindered or improving. Reviewers and journalists should not only focus on reports that is seemingly ‘perfect’, even when the results are underpowered or multiple underpowered testing has been conducted. Rather, the reports that effectively and correctly utilized scientific methods but showing complete opposite results of the hypothesis should also be taken into account and published. In short, both positive and negative replication studies should be incentivised. In this way, good practices of scientific experiments are not discouraged even when the results shown are inconsistent with the initial hypothesis. Real scientific reports should be encouraged, scientific progress from the outside cover of multiple significant findings will only continue to hinder science and journalists, reviewers and editors all have a responsibility in promoting real practices of science.

Teachers of research methods and statistics

It is important for teachers to establish a scientific culture of “getting it right” instead of “getting it published” (Asendorpf *et al.*, 2013, p.111). Under such a culture, the reproducibility and replicability of scientific findings would gradually increase as the standard of good practices in scientific experiments is valued. Teachers should also teach concepts necessary to understand replicable science and not just published and focusing on the outcome science. It is vital for research method courses to emphasize that questionable research practices (p-hacking, HARKing etc) are not to be used. Teachers should also discuss more about the importance of research ethics in research methods. Transparency should be encouraged and understood to students at a young age and even replication studies itself is a priority for students to understand its importance in the scientific field and how it should be practiced.

Institutions

The role of institutions are of crucial importance as most of the creation, funding and dissemination of psychological research occurs within and with the support of institutions. As most of the rewards from institutions are primarily based on quantity of publications over the quality, institutions should shift their incentives to focus more on the quality of psychological research, the values institutions create should establish quality work and not quantity work. The distribution of experiment funding are under the control of institutions, and this could act as an effective tool in controlling the publications or quality of psychological research, where a portion of the fund can be devoted into replication assessments or replicability of the report. Lastly, if all focus and attention is given to successful replications and significant findings, then the actual null findings or potential problems in previous publications would be neglected and new discoveries in deliberately avoided areas would remain hidden. So as to say, it is important that institutions start to also acknowledge null findings or non replicable replication results. In this way, discovery and building on previous knowledge is not discouraged in the scientific field of research.

Last but not Least——Treatments towards Criticism

In 2016, Susan Fiske, former president of the Association for Psychological Science and researcher at Princeton published an open letter to address the criticism that has been spread around social media in relation to the ongoing replication crisis in psychology (Fiske, 2016). She called those who openly criticise experiments and researchers ‘methodological terrorists’, and further urged for such criticism to be made in a closed environment. She proposed this should happen privately and in more controlled places such as journals, or not at all. Her reasoning was that such criticism could hurt reputable researchers' careers and that it "undermines science". This caused an onslaught of criticism from various directions within the scientific community who believed Fiske wanted to cover up faulty research, instead of bringing it into the light in order to set it right.

Various people responded to Fiske, such as Andrew Gelman, a statistician from Princeton. He claimed Fiske's stance in the matter were rooted in her attempts at protecting her career, as he found papers published by her contained faults such as statistical errors (Gelman, 2016). He believed Fiske should take responsibility for the faulty contributions she's made instead of blaming unknown 'terrorists' for bringing light to it. Gelman also believes open debate and criticism is good for psychology and science as a whole, and is something that should be encouraged (Letzter, 2016).

Another person who replied to Fiske was Tal Yarkoni, a professor at the University of Austin, Texas. He claimed there was no tone problem in psychology that Fiske claimed there was, and suggested the current way of open criticism is the best way to go about such things happening in psychology research (Yarkoni, 2016). He proposed a model of the current events within psychology, a "fire-on-engine model". If your car is on fire and someone yells that your engine is on fire, you should not be concerned that this person was screaming at you, but at the problem at hand. He pointed out that Fiske's priorities are wrong, she is focusing on the people doing the science instead of on science itself, which naturally should be priority number one. Yarkoni also claimed that the 'debunkers' who bring faults within psychological research to light are the ones who suffer the most, and that even obvious faults can take an extremely long time to be corrected (Letzter, 2016).

There were two examples which showed that taking a traditional route of approaching faulty research didn't work in favour of those attempting to bring light to it. Steven Ludeke was one person who noticed errors that had been published in experiments. Even though he approached the researchers privately, they refused to cooperate for some time and it took over a year to solve the whole issue (Letzter, 2016). Another person was Marcus Crede who criticised certain findings to be irregular and weak, and in response he was criticised by the original researcher and called racist (Letzter, 2016). Both of these men, who have personally tried to amend such mistakes in a more private way, agree that this route of criticism does not work.

Open discussion and critique may be more helpful in solving issues such as mistakes to be fixed quicker or removed, and faulty papers to be withdrawn as it puts pressure on those who have conducted faulty research. There is a risk of there being cover-ups and such when this is done in private. These two articles discussed the fact that the research itself should always be prioritised, and not the researchers and their reputation. If bad science has been done, then it has to be addressed as quickly as possible.

Quiz

Write 10 easy quiz 4 multiple choice questions to ensure reading and understanding of this chapter:

1. Of which effect in which field did the debate about replication crisis begin?
    a. Context effect; cognitive psychology
    b. Priming effect; social psychology
    c. Placebo effect; abnormal psychology
    d. Spotlight effect; social psychology

2. Which of the following options is NOT mentioned by Kahneman in his open letter?
    a. The first victims should be the young people on the job market
    b. An association with prominent social psychologists should be formed
    c. Publishing the work beforehand and make the data available for all to analyze
    d. Challenging psychologists to clean up their act

3. Which of the following options is NOT a key point people tend to forget when advocating that the ‘reproducibility crisis’ extsts?
    a. Science needs time to ‘reconcile conflicting results’
    b. Reproducibility failures are a part of research
    c. Conclusions should not be based on single research papers
    d. Failing to replicate implies methodological issues

4. Which of the following contributes to scientific errors?
    a. Large sample size leads to more underpowered results;
    b. Distinguishing between random and systematic errors;
    c. Conducting multiple testing with significant results;
    d. Using methods of statistical analyses suitable for the study design;

5. Researchers conducted a research which aimed to test the efficacy of drug X, in which the null hypothesis H0 (no effect) was tested against an alternative hypothesis Ha (some effect). They found that the result was in favor of an alternative hypothesis with *p* value <0.05. What could be concluded?
    a. H0 is false.
    b. Ha is true.
    c. Both (a) and (b).
    d. None of the above.

6. Researchers conducted a research which aimed to investigate whether Treatment X can help improving the performance on Task Y. They compared the means of the control and experimental groups with 50 participants in each group). They used an independent means t-test and obtained a result that is significant (t=2.7, df=48, p=.01). Which of the following sentences is correct?

a. Researchers could absolutely disprove the null hypothesis (i.e. there is no difference between the population means).
b. Treatment X can help improving the performance on Task Y.
c. If the researchers repeat the experiment with a great number of times, hypothetically, they would obtain a significant result on 99% of occasions.
d. None of the above.

7. Redish *et al.* (2018) have suggested a new perspective as to the role of reproducibility in science, which of the following best describes their approach?
   a. Reproducibility should be seen as an essential part of the scientific process, leading research to progress.
   b. Reproducibility should not be overlooked by scientists as it is not indicative enough of whether an effect exists.
   c. Reproducibility should be the major indicator of whether bad science has been conducted.
   d. Reproducibility's role is important but findings from original studies are a better estimate of whether an effect truly exists.
8. Who were the 'methodological terrorists' that Susan Fiske criticised in her column?
   a. The researchers contributing to the replication crisis with faulty experiments.
   b. People criticising researchers involved in the replication crisis on social media.
   c. Those wanting 'open criticism' of faulty experiments.
   d. Researchers Fiske had previously worked with.

9. Where did Susan Fiske suggest criticism towards researcher's experiments be directed?
   a. In journals, through emails, or in the New York Times.
   b. In journals, through personal letters, or face-to-face.
   c. In private, through meetings, or through Skype.
   d. In private, in journals, or not at all.

10. According to Fiedler and Schwarz (2016), why author reports all of the dependent variables is important?
   a. To reduce the possibility of misunderstanding the data.
   b. To avoid author concealing the unwanted results.
   c. For better improving the quality of replication.
   d. To provide the original procedures.

References

Asendorpf, Jens B., Conner, Mark, De Fruyt, Filip, De Houwer, Jan, Denissen, Jaap J. A., Fiedler, Klaus, . . . Wicherts, Jelte M. (2013). Recommendations for Increasing Replicability in Psychology. *European Journal of Personality, 27*(2), 108-119. https://doi.org/10.1002/per.1919

Bargh, J., Chen, M., & Burrows, L. (1996). Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action. Journal of Personality and Social Psychology, 71(2), 230-244. https://psycnet.apa.org/doiLanding?doi=10.1037%2F0022-3514.71.2.230

Bargh, J. (2012, March 5). Nothing in Their Heads. Psychology Today Blogpost. Retrieved from https://osf.io/bgb3c/ .

Bragh, J. (2012, May 11). Priming Effects Replicate Just Fine, Thanks. *Psychology Today*. Retrieved from https://www.psychologytoday.com/us/blog/the-natural-unconscious/201205/priming-effects-replicate-just-fine-thanks

Cyranoski, D., & Ledford, H. (2018). Genome-edited baby claim provokes international outcry. *Nature*, *563*(7733), 607-608. doi: 10.1038/d41586-018-07545-0

Decision Science News. (2012, October 5). Kahneman on the storm of doubts surrounding social priming research. Retrieved from http://www.decisionsciencenews.com/2012/10/05/kahneman-on-the-storm-of-doubts-surrounding-social-priming-research

Diener, E. & Biswas-Diener, R. (2019). The replication crisis in psychology. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/q4cvydeh

Doyen, S., Klein, O., Pichon, C., Cleeremans, A., & Lauwereyns, J. (2012). Behavioral Priming: It's All in the Mind, but Whose Mind? (Behavioral Priming Reexamined). *PLoS ONE, 7*(1), E29081. https://doi.org/10.1371/journal.pone.0029081

Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to?. *Proceedings of the National Academy of Sciences, 115*(11), 2628-2631. https://doi.org/10.1073/pnas.1708272114

Fanelli, D., Costas, R., & Ioannidis, J. P. (2017). Meta-assessment of bias in science. *Proceedings of the National Academy of Sciences*, *114*(14), 3714-3719. https://doi.org/10.1073/pnas.1618569114

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, *7*(1), 45-52. https://doi.org/10.1177/1948550615612150

Fiske, S. T. (2016). Mob rule or wisdom of crowds? *APS Observer*. Retrieved from http://datacolada.org/wp-content/uploads/2016/09/Fiske-presidential-guest-column_APS-Observer_copy-edited.pdf

Gelman, A. (2016, September 21). What has happened down here is the winds have changed. Retrieved from https://statmodeling.stat.columbia.edu/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/

Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science, 1*(2), 198-218. https://doi.org/10.1177/2515245918771329

Gilbert, D., King, G., Pettigrew, S., & Wilson, T. (2016). Comment on "Estimating the reproducibility of psychological science". *Science (New York, N.Y.), 351*(6277), 1037. https://doi.org/10.1126/science.aad7243

Kahneman, D. (2012, September 26). A proposal to deal with questions about priming effects. Decision Science News. Retrieved from https://osf.io/dw7m4/ .

Koole, S. L., & Lakens, D. (2012). Rewarding Replications: A Sure and Simple Way to Improve Psychological Science. *Perspectives on Psychological Science*, *7*(6), 608–614. https://doi.org/10.1177/1745691612462586

Letzter, R. (2016, September 23). Scientists are furious after a famous psychologist accused her peers of 'methodological terrorism.'. *Business Insider*. Retrieved from https://www.businessinsider.com/susan-fiske-methodological-terrorism-2016-9

Molden, D.C. (2014). Understanding Priming Effects in Social Psychology: What Is "Social Priming" and How Does It Occur? *Social Cognition, 32*, 1-11. https://doi.org/10.1521/soco.2014.32.supp.1

Mullane, K., & Williams, M. (2017). Enhancing reproducibility: Failures from reproducibility initiatives underline core challenges. *Biochemical pharmacology*, *138*, 7-18. https://doi.org/10.1016/j.bcp.2017.04.008

Nussbaum, D. (2012, May 3). Conceptual Replication. Retrieved from http://davenussbaum.com/blog/conceptual-replication-part-i

Open Science Collaboration (2015) Psychology. Estimating the reproducibility of psychological science. *Science 349*(6251):aac4716. https://doi.org/10.1126/science.aac4716

Redish, A. D., Kummerfeld, E., Morris, R. L., & Love, A. C. (2018). Opinion: Reproducibility failures are essential to scientific inquiry. *Proceedings of the National Academy of Sciences, 115*(20), 5042-5046. https://doi.org/10.1073/pnas.1806370115

Roberts, B. (2015). The new rules of research. Retrieved from https://pigee.wordpress.com/2015/09/17/the-new-rules-of-research/

Schimmack, U., Moritz, H., & Kesavan, K. (2017, February 2). Reconstruction of a Train Wreck: How Priming Research Went off the Rails. Retrieved from https://replicationindex.com/2017/02/02/reconstruction-of-a-train-wreck-how-priming-research-went-of-the-rails/comment-page-1/#comment-1454

Schwarz, N. (2012, October 2). Response to Ed Yong's questions. The Open Science Framework. Retrieved from https://osf.io/v3435/

Schwarz, N., & Clore, G. (2016). Evaluating Psychological Research Requires More Than Attention to the N: A Comment on Simonsohn's (2015) "Small Telescopes". *Psychological Science, 27*(10), 1407-1409. https://doi.org/10.1177/0956797616653102

Simons, DJ. (2014). The value of direct replication. *Perspectives on Psychological Science 9*(1),76–80. https://doi.org/10.1177/1745691613514755

Singal, J. (2016, October 12). *Inside Psychology's 'Methodological terrorism debate'*. Retrieved from https://www.thecut.com/2016/10/inside-psychologys-methodological-terrorism-debate.html.

Srivastava, S. (2015). Moderator interpretations of the Reproducibility Project. Retrieved from https://thehardestscience.com/2015/09/02/moderator-interpretations-of-the-reproducibility-project/

Stafford, T. (2017, February 16). The Social Priming Studies in "Thinking Fast and Slow" are not very replicable. Retrieved from https://mindhacks.com/2017/02/16/how-replicable-are-the-social-priming-studies-in-thinking-fast-and-slow/

Vadillo, M. (2012, October 17). *Returning Trust in Social Psychology*. Retrieved from https://psicotecablog.wordpress.com/2012/10/17/devolviendo-la-confianza-en-la-psicologia-social/

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences, 113*(23), 6454-6459. https://doi.org/10.1073/pnas.1521897113

Yarkoni, T. (2016, October 1). There is no "tone" problem in psychology. Retrieved from https://www.talyarkoni.org/blog/2016/10/01/there-is-no-tone-problem-in-psychology/

Yong, E. (2012, March 10). A failed replication draws a scathing personal attack from a psychology professor. Discover. Retrieved from https://osf.io/esgdh/ .

Yong, E. (2012, October 3). Nobel laureate challenges psychologists to clean up their act. Nature News. Retrieved from https://osf.io/b5fe6/

5 - Beyond Psychology: Replications in other fields

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Madeleine	Mendes	https://www.researchgate.net/profile/Madeleine_Mendes	https://osf.io/zdgcq/	u3565270@connect.hku.hk	maddy.mendes@hotmail.co.uk	3035652709
Tallulah	Thomas	https://www.researchgate.net/profile/Lulu_Thomas2	osf.io/encv5	u3564907@connect.hku.hk	tallulah.thomas.uk@gmail.com	3035649075
Anne	Zonneveld	https://www.researchgate.net/profile/Anne_Zonneveld	https://osf.io/profile/	u3565158@connect.hku.hk	annewzonneveld@gmail.com	3035651585
Caoimhe	Mackin-Lau	https://www.researchgate.net/profile/Caoimhe_Mackin-Lau	https://osf.io/b4u6a/	u3555319@connect.hku.hk	caoimheml12@gmail.com	3035553197
Karoline	Kristensen	https://www.researchgate.net/profile/Karoline_Kristensen2	osf.io/xzmyv	u3565422@connect.hku.hk	karohk@gmail.com	3035654226

Managerial summary

So far, we have discussed the reproducibility crisis in psychology, which emerged after it was found that classic studies were unable to be replicated and similar findings could not be reproduced. Though psychology seems to be the field most affected, the reproducibility crisis has appeared in other areas, including but not limited to biology, medicine, and philosophy. The reproducibility of experiments is considered a fundamental assumption of the scientific method, so there can be disastrous consequences in fields where significant theories may be based on experimental research that cannot be replicable.

This chapter will introduce the reproducibility crisis in areas of academia other than psychology and how the crisis has been demonstrated in research, as well as potential reasons for the failures in replication. This chapter will also discuss the relatively high replicability rates found within particular scientific fields, such as experimental philosophy, and how we might use these studies as models for future experiments, to increase study replicability across other scientific domains. Raising awareness of the issue so significant and disruptive to the integrity of scientific research as a whole, we hope to blow the roof off the crisis encouraging the movement of open-science to develop and no longer remain such an under-reported myth-like phenomenon.

In depth report

Introduction

A recent reproducibility crisis has emerged in empirical science. A possible cause may be the "publish or perish" mantra that notoriously prevails in today's academia. Certain findings, namely novel, significant results, are more likely to be published (Fanelli, 2012). Therefore, researchers are more tempted to participate in questionable research practices (QRPs) to make their findings quickly publishable. Studies suggest that 34% of empirical researchers conduct QRPs such as selective publishing of variables, conditions and analyses (only the ones that yield significant results), optional stopping (i.e. continuing data selection until significant results); also known as 'p-hacking' (Fanelli, 2009). This violation of scientific integrity heavily impacts the confidence and validity of empirical research, since significant findings might actually be false because of the high chance of false positives (Prinz et al., 2005). Thus, it is clear that if inconsistencies exist within one domain of science, psychology, the likelihood that they are more prevalent in many other regions is considerable.

Cancer Research

Extending the crisis beyond psychology, we witness inconsistencies of closer to home matters, particularly in the fields of: medicine; such as unreliability of health & illness research, and economics. First exploring the credibility crisis in medicine we look into a field that interests and affects most, cancer biology. This epidemic affects all individuals at some point, cross-culturally. Thus, cancer prevention research is highly sought-after research. Thus we, as scientists, are motivated to carry-out the research accurately and reliably, so that appropriate treatment implications may be derived; minimising risk that any errors may lead to unsolicited mortality of cancer patients. Reliable open research findings instil confidence these treatments will both be effective and appropriate.

However, when closely analysing the field of cancer research we also discover inconsistency. Perhaps motivated by the aforementioned QRP's (more-harshly described as "Sloppy data analysis, contaminated lab materials, and poor experimental design" by Engber, 2016) and their ability to shortcut research to publishing. In research paper; "the economics of reproducibility in preclinical research." by Freedman, Cockburn and Simcoe (2015) they claim to assume "half of all biomedical literature findings rest on shaky ground", thus suggesting very low replicability of the field. From this, they further report a suggested $28 million spending waste per annum devoted to research costs relative to US spending habits 2015. A possible explanation suggested for prevalence of crisis within life-affecting cancer research is that resource demands necessary to reproduce biomedical research studies at their equivalent scale are supposedly far greater than the demands of most psychological studies (Engber, 2016). This also therefore contributes to masking the crisis from being extended to the field of biomedicine, as the data is simply not available to critique.

Further evidence in support of the crises roots in cancer research: a large-scale reproducibility project by Davis (2014) found only 5 of 14 biomedical research studies were able to replicate convincingly important parts from their original study; only another 4 out of the 15 replicating some parts from original, totalling at 6 of 15 (40%) biomedical replications not-replicating their original studies.

Neuroscience

We followed this matter within further biological sciences. Neuroscience often makes the connection between biology and psychology; making the fields of psychology and neuroscience fairly closely-related. Just like psychology, neuroscience is also affected by the crisis. For example, Boekel et al. (2015) did a replication with pre-registration of five structural brain-behaviour correlations including a total of 17 effects of which all but one indicated evidence in favor of the null hypothesis (based on Bayesian hypothesis tests). The major aspect of the problem is low statistical power, which can be caused by low sample sizes or small effects. A low statistical power negatively affects the chance that a significant finding reflects a voracious effect. It is hard to determine the actual average statistical power in neuroscience, since true effect sizes are unknown. The best method to do this is to analyse the summary effects of meta-analyses. Button et al. (2013) did this and established a median statistical power of 21% in neuroscience. However these summary effects, including the power estimate calculated from these summary effects, may be inflated, especially when the original studies included low sample size (which is often the case in neuroscience). This is because of the fact that for example in the case of a small study a medium true effect will only pass the threshold for discovery, if the magnitude is overestimated (by chance). This is called ‘winner’s curse’: the lucky scientist who discovers the effect is cursed by finding an inflated estimate of this same effect. Because of this winner’s curse and the publication bias, the actual power in neuroscience is likely to be even lower.
Button et al. (2013) also did analyses for specific subfields of neuroscience such as brain imaging and animal behaviour studies. However, considering the results of their analysis, we have to keep in mind that meta-analyses in these subfields are fairly under-represented:

MRI (magnetic resonance imaging) studies, a certain type of brain imaging studies, often have very small sample sizes, since fMRI scans are costly, and thus have very little power to detect differences between compared conditions. Also, individual effect sizes are very small because of the highly complex nature of the brain. An excess significance bias has been shown for structural MRI studies and there seems to be a comparable problem for fMRI (functional MRI) studies (Ioannidis, 2011; David et al., 2013). After conducting an analysis of the summary effects of multiple meta-analyses, Button et al. (2013) found the median power in structural MRI studies to be only 8%.

A similar problem can be found for behavioural animal studies, another subfield of neuroscience. Button et al. (2013) analysed two meta-analyses about water maze and radial maze performances and found that the used sample sizes were very small, since training rodents is a demanding task and housing can be costly, resulting in a median statistical power of 18% and 31%. These studies were thus severely underpowered and could only have picked up on very large effects.

A real complication on meta-analysis’ explanatory power of findings in neuroscience exists the publication bias. Significant findings tend to be favoured in publication. Thus, meta-analyses of only-published data will be a distortion of the true effects and findings within the whole field of neuroscience. Ideally meta-analyses would be carried out on wider study samples; including studies that did not make publication due to favourability of significance.

Microarray gene expression

The reproducibility crisis also affects other biological fields, such as microarray gene expression. Since these studies are highly complex, guidelines encourage transparent methods and public availability of data. E.g. the Uniform Guidelines of the International Committee of Medical Journal Editors say that authors have to "identify the methods, apparatus and procedures in sufficient detail to allow other workers to reproduce the results" (International Committee of Medical Journal Editors, 2006). Many journals, such as Nature Genetics, also require public data availability as a prerequisite for publication. However, it turns out that not all the data is publicly available and that described data analyses are often not detailed enough for repeatability. Ioannidis et al. (2009) evaluated the repeatability of 18 articles on published microarray gene expression analyses and could only replicate two analyses in principle, six partially and ten could not be reproduced. This was due to different reasons (see Fig 1).

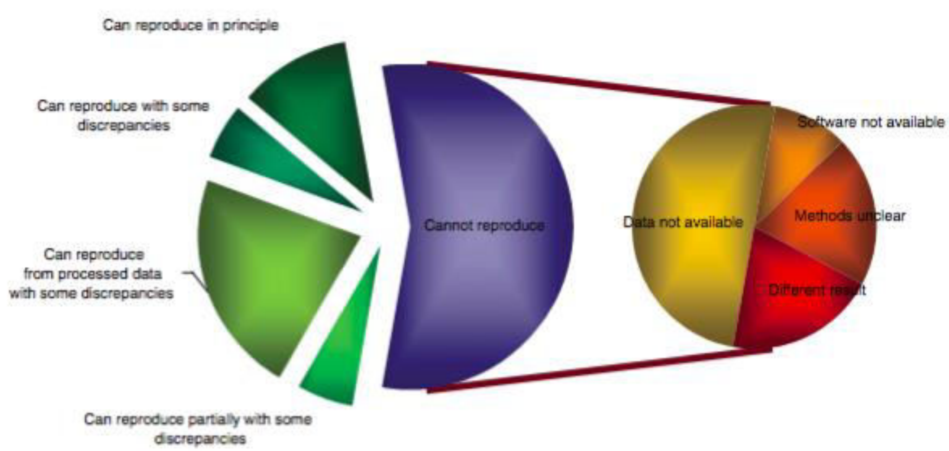


Figure 1. Summary of efforts to replicate published microarray gene expression analyses (Ioannidis et al., 2009).

For two articles the data turned out to be unavailable. Additionally, for some articles only summary analyzed data was available and for other articles it was not possible to determine which dataset corresponded to which data analysis. Therefore, only 10/18 articles were actually appropriate for an attempt at reproduction of the data analyses. Thus, reproduction of microarray gene expression analyses requires greater detail than provided by the standard platforms (Draghici et al., 2006) or publicly available data. This limits the options for integration of results from different studies, and the option to use the data for additional research.

Biomedicine

The modern scientific method reflects the idea that experiments are able to be replicated, and forms the basis of the design of the majority of scientific experiments. The reproducibility crisis has been widely reported in experimental psychology and other fields of science including cancer biology, however the similar crisis in the field of biomedicine is not nearly as well-publicized. It's been reported that around half of experimental results cannot be replicated in psychology, but the reality seems to be much worse for biomedicine. A 2011 study from Bayer HealthCare failed to replicate 75% of what were believed to be significant findings regarding pre-clinical cancer biology, and only 11% of results were able to be replicated in another study. This is important because if irreproducible findings lead to the creation of new fields and open up new areas of research, results subsequently conducted in these studies can be considered baseless, and may not be accurate or methodologically valid.

Prinz, Schlange & Asadullah (2011) conducted an analysis of early projects in target identification and validation in strategic research fields of oncology, women's health and cardiovascular diseases. This was prompted by a report by Arrowsmith (2011) who found there was a 10% drop over a two year period in the success rates of new development projects in Phase II trials. These trials aim to provide evidence and support to the effectiveness of therapeutic treatments. They found that only 20-25% of published data was completely in line with in-house findings, and there were inconsistencies between published and in-house data in nearly two thirds of the projects. This resulted in either increased duration of the process, or the termination of the projects as there was not enough evidence to justify continuing the research.

So why is it that so much of the data and findings in biomedicine cannot be replicated? There are several possible reasons. Firstly, conducting research in biomedical science can be extremely expensive and time consuming, as well as incredibly challenging, as seen earlier with the issues surrounding cancer replicability rates. The time and money that may need to be put into a single biomedical study could be overwhelming for the researchers, driving them to use smaller samples and multiple post-hoc measures and analysis to gain the desired results.

Secondly, certain incentives can make researchers biased in their results and conduct. In a field such as biomedicine, which is very competitive and is constantly being expanded, there is a rush (whether conscious or not) for researchers to be the first to publish a new paper, or report certain findings. Commercial interest, particularly from pharmaceutical companies, adds to the pressure for beneficial results to be found and reported. Ebrahim, Bance, Athale, Malachowski, & Ioannidis (2016) reported that studies with authors associated with the manufacturers of antidepressants were 22 times less likely to report negative statements and cautions about the drug compared to other studies. This finding seems to indicate that the importance of monetary and industrial benefits exceeds that of accurate and scientifically valid research.

Linked to this is the publication bias, where negative results are rarely published. Many journals are not interested in publishing studies with non-significant findings, and many researchers are eager to publish the most exciting and groundbreaking results. Pharmaceutical companies additionally will not want to submit evidence that their drugs don't have the desired effect, and researchers may not wish to reveal that their studies didn't yield the desired results.

However, publishing any negative results is important for a number of reasons; mainly, we learn from them in terms of developing appropriate methodologies to study certain subjects, and how to improve research into certain areas. Research falsely presented as true can lead to a waste of time and resources in replication attempts, materials which could have been put towards more promising areas of research. Given that scientists today are mutually dependent on each others research and are dependent on their work being cited by others (Persson & Kusnitzoff, 2017), it's important that all findings are published in order to prevent collections of faulty research piling up

Issues of scientific fraud also contribute to the reproducibility crisis. A study published in 2001 in the *Journal of the American Academy of Child and Adolescent Psychiatry* helped to approve the use of the antidepressant paroxetine, or Paxil, in treating adolescents with depression. However, since then, the authors of the study have faced criticism for supposed ghostwriting, false reporting, conflicts of interest and issues with data analysis. The authors were accused of downplaying the more negative side effects of the drug, and not publishing findings that showed the drug increased the risk of suicidal behaviour. More subtle forms of fraud include the use of p-hacking (manipulation of the data or methods of analysis in order to reach the desired significance level). Simple modifications, such as decreasing the sample size, can lead to an increase in the measured effects, which thus leads to inflated results. Scientific fraud is dangerous in that potentially harmful findings can be hidden or alternatively presented as safe and scientifically proven, both ethically and morally questionable.

The reproducibility crisis in biomedicine is dangerous in the same way as it is for other fields of study. Issues of academic honesty not only call the validity of research into question, but are also capable of damaging the reputation of parts of the scientific community. Because the majority of research in biomedicine is aimed towards developing treatments for various illnesses and improving the overall quality of life, there could be devastating effects on scientific developments and the flow of the healthcare system if research stalls or is not being conducted in an honest manner. Overall, the key issue with the reproducibility crisis in biomedicine is whether the science being conducted within this field is even science at all (Gobry, 2017).

Economics

The crisis may be extended to the field of economics (Camerer et al. 2019). Attempting to advance and establish results of existing research in Economics, replications have been carried out to try to validate published results. A cross-journal analysis by Chang and Li (2015) replicated 67 articles from 13 different established economic journals, 2008 to 2013. Significantly, merely 22 out of 67 papers (33%) were able to replicate based on the data available. However, due to missing data and code replication files, a considerable amount of papers were unable to be replicated as a result. If the papers which did not provide data, for confidentiality reasons, these were excluded. Thus, 29 out of the 59 papers were able to replicate; below 50%. An important factor heavily criticised, across fields, throughout the crisis is the exclusion of data availability in published papers. Also the case in Economics, wherein 58% of the papers evaluated did not include data, and thus were unable to replicate. In economics, the fact that the authors did not submit data files "constitutes approximately half of our failed replication attempts" (Chang & Li, 2015, p. 11). Thus, seeing as less than half of the papers examined by Chang and Li (2015) were able to successfully replicate, the authors concluded that research in the field of economics is generally not replicable. These implications call for a change in the criteria for the inclusion of data in economic journals.

Experimental philosophy

Experimental philosophy is a relatively new field of study, characterised by its aim to unite the important questions asked in the field of philosophy with the empirical evidence used in psychology.

Following the reproducibility project in the field of psychology, a study was conducted to test the replicability of experimental philosophy studies: 20 research teams ran high quality replications of 40 studies within the field published between 2003 and 2015. They found that 70% of the experimental philosophy studies were highly replicable (Cova et al., 2018).

The researchers followed three different methods for reporting a replication as a success or a failure:

1. **Checking the statistical significance of the replication results.**
2. **Subjective assessment by the replicating team** – so that study limitations, like the use of different equipment are accounted for.
3. **Comparison of the original and replication effect size** – the replication crisis has highlighted the importance of effect size, for example, a study can report the presence of an effect, when in reality, the effect size may be only marginally significant.

However, a limitation of replicating studies has created a bias, as researchers have tended to replicate more surprising and dramatic results, excluding the less surprising findings. Therefore, although current figures suggest low replicability rates of studies in many scientific spheres, the proportion of replications made are more likely to be unsuccessful, thus replicability rates may be over-exaggerated by this bias. Alternatively, current replicability predictions may be under-exaggerated, as overlooked, less surprising studies may also have poor replicability.

To combat this, the researchers investigating the replicability of experimental philosophy, chose three papers from each year (2003 to 2015): one as the most cited paper for this year, and two at random (except for 2003, whereby only two papers were available). This gave a total of 38 studies, to which 4 studies were added in case some of the originally selected studies could not be replicated (1 could not be).

Using journals listed on the *Experimental Philosophy* website (a credible and widely used resource within the field of experimental philosophy), a list of 35 journals was composed, including only published papers that clearly fit in the field of experimental philosophy and that analysed empirical data.

The nature of the participants used forms the base of the study, thus the researchers were explicitly instructed to report on the size and nature of the original sample - an underpowered study, with too few

participants can give a falsely large effect size. Moreover, researchers noted whether the original study included a selection procedure, when finding participants, as a sample taken from a small population may not be representative of a larger population, thus giving the study poor external validity. As well as this, a randomization procedure is necessary in a study when assigning participants to conditions, as consciously or subconsciously a researcher may assign candidates to a condition they believe will yield successful results.
The researchers also looked to see whether enough information - of the participant's background and prior knowledge - had been obtained in the original study, to factor in any limitations of the experiment.

There are several possible explanations as to why experimental philosophy has obtained such high replicability rates. Firstly, the higher effect sizes reported in experimental philosophy studies, may contribute to the high replicability rate, as effect sizes have been found to be good predictors of an effect's replicability. This is because a strong effect will be less sensitive to study limitations and thus are more easily replicated. Effects were especially large in the early years of experimental philosophy but have tended to reduce with time – this is something that should be considered for progression of producing more replicable studies.

Secondly, it has been found that experimental philosophers are more likely to publish their papers that achieved a null result. This could be explained by promoting that; results of any kind are beneficial to our understanding, within the study of philosophy in particular.

Lastly, experimental philosophy replications are usually low-budget, as the effects being studied are often easy to recreate in a lab setting. Therefore, researchers may be more inclined to publish non-significant findings, as a lost cost experiment will be easy for others to use and the researcher's reputation won't be affected from a high-cost experiment with non-significant results. This reduces the file-drawer effect, where papers that achieved a null result are not published. Thus, ignoring the effect of a type II error (falsely accepting the null hypothesis).

Conclusion

The reproducibility crisis in biomedicine and other scientific field is very much occuring and has detrimental consequences for the field of science overall. Issues of academic dishonesty not only call the validity of research into question, but are also capable of damaging the reputation of parts of the scientific community. Particularly in areas such as biomedicine and cancer research, where research is directly involved in people's health and quality of life, issues with reproducibility can contribute to wider issues regarding the ethics and morality of conduct. The crisis was also noted in economics, which can have a significant impact on individual socioeconomic systems and the global economy. However, the major contributing factor to the crisis in economics was said to be the unavailability of data in research publication (Chang & Li, 2015). This being the main benefactor to the prevalence of the crisis in economics is far more redeemable than the position of the biological sciences, and further demonstrates that we must incentivize the inclusion of all appropriate data and disincentivize disinclusion (applaud and make data inclusion a necessity for publishing; but actually reinforcing its regulation!)

Due to the high replicability rate of experimental philosophy studies, perhaps scientific fields which did not perform as well, such as, but not limited to psychology, should either adjust their studies to be more in line with those within experimental philosophy- for example, conducting lower budget studies that can be more readily replicated. This lower-budget study approach can be unrealistic for scientific domains (such as biomedicine) due to complexity and a need for absolute precision, but the underlying principles could be adapted to suit different areas of study. Conversely, researchers could adjust their mindsets to acknowledge the limitations discussed and be more willing to publish unsuccessful results.

Quiz:

1.What percentage of experimental philosophy studies were found to be highly replicable?

A. 36%
B. 12%
C. 89%
D. 70%

2. What has been found to be a good predictor of study reliability?

A. Sample size
B. Citations
C. Effect size
D. Date of publishing

ANSWERS: D & C

3. What possible cause for biomedical research's ability to mask credibility crisis did Engber, 2016 suggest:

(Select all that apply)

A. The data doesn't exist
B. B. The resource demands for replication are too great.
C. The significance level for biological research is stricter (p= 0.01)
D. Cancer research findings are more objective than psychological or other fields.

ANSWER: A & B

4. What percentage of cancer studies by Davis (2014) did not replicate in his reproducibility project findings?

A. 11%
B. 33%
C. 40%
D. 60%

Answer: C.

5. Authors of the 2001 study approving the use of paroxetine in treating adolescents with depression faced heavy criticism, but why?

A. They didn't include a larger and more varied sample size
B. They didn't discuss how the drug could be used to treat other disorders
C. They didn't present all of their collected data

D. They didn’t publish findings indicating the drug increased suicidal behaviours

ANSWER: D

6. Which of these is NOT a reason given as to why negative results in research are not published?

A. Scientific journals and magazines are not interested in reporting negative results
B. Researchers don’t want to show they found non-significant results
C. Researchers don’t want their methodology to be questioned
D. Researchers are more interested in publishing exciting, groundbreaking findings

Answer: C

7. According to Fanelli et al. (2009), what percentage empirical researchers have participated in QRP?
A. 34%
B. 61%
C. 18%
D. 8%

8. What are the two subfields of neuroscience that Button et al. (2013) examined concerning the reproducibility crisis?

A. Neuropharmacological studies and behaviour studies
B. Molecular neuroscience and neuropharmacological studies
C. Brain imaging studies and animal behaviour studies
D. Molecular neuroscience and brain imaging studies.

9. According to the replication research conducted by Chang and Li (2015), how well did the papers in economics replicate?

A. Merely 29 out of 68 papers
B. Around 20-25 % of published papers
C. Merely 22 out of 67 papers
D. Around 59 % of published papers

ANSWER: C

10. In the cross-journal analysis conducted by Chang and Li (2015), what was the primary reason for failed replication attempts?

A. The authors withheld data and did not want to contribute to the study
B. Data was excluded due to confidentiality reasons
C. 33% of the papers did not include data

D. The fact that the authors did not submit data files constitutes to approximately half of the failed replication attempts

ANSWER: D

References

Boekel, W., Wagenmakers, E. J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, *66*, 115-133.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T., Huber, J., Johannesson, M., … Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*, (6280), 1433-1436

Chang, A. C., & Phillip Li (2015). Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not"," *Finance and Economics Discussion Series,* 2015-083.

Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., ... & Cushman, F. (2018). *Estimating the reproducibility of experimental philosophy*. Review of Philosophy and Psychology, 1-36. DOI: https://doi.org/10.1007/s13164-018-0400-9

David, S. P., Ware, J. J., Chu, I. M., Loftus, P. D., Fusar-Poli, P., Radua, J., ... & Ioannidis, J. P. (2013). Potential reporting bias in fMRI studies of the brain. *PloS one*, *8*(7), e70104.

Davis, R. J. (2014, 10th Dec). Reproducibility Project: Cancer Biology. [Weblog]. Retrieved 8 November 2019, from https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology

Draghici, S., Khatri, P., Eklund, A. C., & Szallasi, Z. (2006). Reliability and reproducibility issues in DNA microarray measurements. *TRENDS in Genetics*, *22*(2), 101-109.

Ebrahim, S., Bance, S., Athale, A., Malachowski, C., & Ioannidis, J. P. A. (2016, February). Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. DOI: 10.1016/j.jclinepi.2015.08.021

Engber, D. (2016, 19th April). Cancer Research Is Broken. Future Tense. Retrieved 6 November 2019, from https://slate.com/technology/2016/04/biomedicine-facing-a-worse-replication-crisis-than-the-one-plaguing-psychology.html

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS one*, *4*(5), e5738.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891-904.

Freedman, L., Cockburn, I. & Simcoe, T. (2015, 9th June). The Economics of Reproducibility in Preclinical Research. Plos Biology. Retrieved 6 November 2019, from https://doi.org/10.1371/journal.pbio.1002165

Gobry, P. (2017, November 17). Research Replication Crisis Is A Growing Problem. *National Review.* Retrieved from https://www.nationalreview.com/2017/11/research-replication-crisis-growing-problem/

Havenaar, M. (2018, November 29). 6 factors causing replication crisis in medical research. [Blog post]. Castor EDC. Retrieved from https://www.castoredc.com/blog/replication-crisis-medical-research/

International Committee of Medical Journal Editors. (2006). Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication, III. Publishing and Editorial Issues Related to Publication In Biomedical Journals, D. Overlapping Publications, 3. Acceptable Secondary Publication. *http://www. icmje. org/*.

Ioannidis, J. P. (2011). Excess significance bias in the literature on brain volume abnormalities. *Archives of general psychiatry*, *68*(8), 773-780.

Ioannidis, J. P., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A. C., ... & Mangion, J. (2009). Repeatability of published microarray gene expression analyses. *Nature genetics*, *41*(2), 149.

Nature News. *Phase II trials.* Retrieved from https://www.nature.com/subjects/phase-ii-trials Accessed November 7, 2019.

Persson, C. P, Kusnitzoff, J.U. (2017, February 8). Basic research crisis? Many results cannot be replicated. (C. Jex, Trans.) *ScienceNordic.* Retrieved from https://sciencenordic.com/academia-basic-research-basic-research-crisis/basic-research-crisis-many-results-cannot-be-replicated/1442310

Palus, S. (2015, September 18). Re-analysis of controversial Paxil study shows drug “ineffective and unsafe” for teens. *Retraction Watch.* Retrieved from https://retractionwatch.com/2015/09/16/re-analysis-of-controversial-paxil-study-shows-drug-ineffective-and-unsafe-for-teens/

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets?. *Nature reviews Drug discovery*, *10*(9), 712.

6 - Grassroot movements

Managerial summary

In the chapter, we discussed the background information that led to the grassroot movement, which begins with the replication crisis in psychology, when scientists realised that it is difficult for them to achieve the same result as their peers when they are repeating their experiments. This subsequently led to the discussion of a more transparent and open practice standard among the scientific community, known as the "Open Science" principles. The grassroot communities of the research adopted and advocated for open science. By encouraging the researchers to fully disclose all the data and tools of their studies, the ability to replicate research results theoretically would increase. Since then, the principles are gaining acceptance in the scientific community, and interactions between communities help consolidate the common practices. Support and positive feedback are observed in these practices and the communities are coordinated to adhere to the open science principles. In the future, we expect to see them to be more widely applied to the scientific community, with the ease of technological advancement. Challenges must be overcome as well, with the cost of disclosure being a more significant barrier.

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Lok Kwan	Chu	https://www.researchgate.net/profile/Queenie_Chu	http://osf.io/uhp3q	clkqnie@hku.hk	clkqnie@gmail.com	3035608526
Wan Ho	Au	https://www.researchgate.net/profile/Leon_Au	http://osf.io/4v7dz	u3528770@hku.hk	auwanho@gmail.com	3035287700
Yue	Chen	https://www.researchgate.net/profile/Isaac_Yue_Chen	http://osf.io/v4ykz	chenyue1@hku.hk	cy842568212@gmail.com	3035533824
Yan Ping	Ho	https://www.researchgate.net/profile/Yan_Ping_Ho	http://osf.io/c9f2g	yph8614@hku.hk	alvayph@gmail.com	3035608265
Chung Kit	Shum	https://www.researchgate.net/profile/Chung_Kit_Shum	http://osf.io/ggnst	sck.elsin@hku.hk	Sck.elsin@gmail.com	3035379046
Chun Wang	Wong	https://www.researchgate.net/profile/Chun_Wang_Wong2	http://osf.io/cy4vu	cwwong28@hku.hk	mcklnericsoncwong@gmail.com	3035284978
Myrte Louise	Druyvesteyn	https://www.researchgate.net/profile/Myrte_Druyvesteyn	https://osf.io/hfqse/	mldruyv@hku.hk	m.l.druyvesteyn@gmail.com	n.a

In depth report

The rise of grassroot movements and Open Science

When being asked about the core values of science, researchers can easily recall terms including but not limited to transparency, openness and reproducibility, for they are the vital features of science and norms among researchers (Nosek et al., 2015; Morey et al, 2016; Johnson et al, 2017). The previous chapters have illustrated how the “replication crisis” has captured attention from scholars among various disciplines. In light of the unfavorable situation in academia, myriads of grassroot campaigns have evolved with researchers’ increasing interest in Open Science, replications and the improvements in experimental practices (Tackett, 2019). Collaborative efforts in systematic replication studies have been published following a set of principles that advocate research openness which allow validation of the findings and subsequent novel analyses and extensions (Caplan & Redman, 2018; Morey et al, 2016; Maxwell, Lau & Howard, 2015). The roles of reviewers and publishers with respect to the beliefs of Open Science has also been inquired into (Morey et al, 2016; Maxwell, Lau & Howard, 2015). The grassroot movements have been developing continuously, striving to reaffirm scientific research standards and uphold core values of science.

The common aim of the grassroot movement is to promote and uphold the values of Open Science, in which the communities try to improve the current research practices. One of the early efforts done was by a group of scientists in the United Kingdom (ORION, n.d.). The contributos, Peter Murray-Rust, Cameron Neylon, Rufus Pollock and John Wilbanks, wrote the Panton Principles in July 2009. The first draft was written at the Panton Arms in Cambridge, which was a pub popularly visited by scientists from the University of Cambridge. It was then refined by the Open Knowledge Foundation, which is a non-profit organization and a leader in the field committed to spread the idea of open knowledge (Murray-Rust, Neylon, Pollock, Wilbanks, 2010). The Panton Principles was officially launched in February 2010 (Gray, 2010).

They believed that “science is based on building on, reusing and openly criticizing the published body of scientific knowledge” (Murray-Rust, Neylon, Pollock, Wilbanks, 2010). This idea follows that open science data is important to the function of science and the spread of scientific knowledge to the society. As they were awareness of the need to make precise suggestions of open science, they adopted The Open Definition which states that “open data and content can be freely used, modified, and shared by anymore for any purpose” (Open Definition, n.d.). This can then promise quality and improve compatibility between different sources of materials. To promote this concept, they set up 4 principles to help male data more accessible, simply put, to apply licence for data. First, it was suggested to include an “explicit and robust” statement regarding the re-use and repurposing of any data element when publishing the data. Second, the statement is required to adopt a recognized license which should be considered suitable for data. Third, it is necessary to make sure the data is aligned with The Open Definition, so that it can be used by others with minimized disputes. With the protection of licence, the data is difficult to be repurposed for commercial use. Fourth, they recommended that data should be put in the public domain, especially for those publicly funded research (Murray-Rust, Neylon, Pollock, Wilbanks, 2010).

With the 4 principles laid out, they hoped to promote a norm in society that alters a bit of traditional research practice. They were suggesting the usefulness of applying license to data to make research reach open science. Their ideas promote the transparency of materials used in the research process. The increased accessibility then increases sharing and spreading of scientific outcome, hopefully achieving scientific knowledge advancement.

Open Science in the community

Community have been formed at university level, district level or regional level. The communities served various activities and have different functions in achieving open science. They had a wide range of indirect and direct contributions to open science, from advocating open science to facilitating research or replications using the open science frameworks or guidelines.

Small institutions and communities in the district often contributed through advocacy and promotion. Their aimed at letting more academics, including but not limited to undergraduates, post-graduates, teachers, professors, to promote open science that the general public can notice and understand. These communities, such as Open Science Community Leiden, take the position of educators to spread the idea of transparent replicable research to others. To fulfill such educational needs, most of these communities would organise formal and informal events. Symposia and conferences can be held to share the latest open science concerns within the community or to the public, while training workshop could prepare academics to perform research in a transparent way and to meet open science requirements. Workshops could be as simple as introducing software such as RStudio to handle data and codes or teaching about procedures of performing pre-registration before a study. Casual cafes meetings are often held by the community to allow members to exchange information or study on the topic of open-science-related issues. Communities developed by universities and districts promotes the circulation of open science research, reproducible research papers and their data sets. Academics can exchange information and latest research quickly within the community.

Some communities aimed to improve the way open science was currently performed and contributed by investigating future development of open science. Open science was not performed in all research disciplines for the time being, and the community is demanding for a higher degree of implementation across different disciplines. Academics' view on the future direction have been gathered by the communities to make a significant impact. For example, Global Young Academy, with 258 alumni and 200 members from more than 80 countries, voiced out the ideas gathered from young researchers to seek suitable scientific outputs being disseminated, target recipients and the expectations on these outputs. They tried to investigate the difficulties and chances to implement open science in a wide range of research institutes or organisations from the perspective of young researchers. Their member actively joined the Science Forum for Latin America and the Caribbean (CILAC 2018), European Commission Open Science Policy Platform, African Open Science Platform (AOSP). Public could contact them at Facebook and Twitter to understand their open science activities and involvement.

Apart from improving open science, larger communities aimed at building connections among stakeholders. Communities such as Innovations for Poverty Action (IPA) act as a platform to link up funders, journals and research groups. Guidelines and manuals on open science mandates were constantly updated and established to facilitate through the participation of these communities. Researchers and research groups gained supported in these communities by funders and journals to perform research under the framework of open science. In terms of connections, communities at district levels had an important role in connecting participants with national and international open science communities. Open Science Community (OSC) has numerous small communities such as Open Science

Community Utrecht, which helped research groups, no matter big or small, to have access to open science and larger open science communities.

Communities also contributed through direct ways to facilitate open science research. Some communities such as Open Science Community Utrecht shared replications and published findings under pre-registration, while others such as IPA centralisation and curation of datasets of research and make data visible and accessible to others. The data was useful for other researchers to take as reference, further analysis and replicate in future research. Provided with access to datasets and codes, researchers met less obstacles in performing replications and analysis. Researchers at the same time shared own datasets and codes using the community as a platform to let others continue the research.

Some communities encouraged open science publications through providing services to researchers. Traditional publishing to journals was expensive and time-consuming. The peer-review process may not be transparent as well. Therefore, some communities provided services on pre-prints and post-prints services which is transparent in recommendations, review and comments. This lowers the fee to publish research results openly and allow public exchange of research ideas. Peer Community In gathered researchers from science discipline to evaluate and recommend pre-prints and post-prints within the community, so as to assure the integrity and validity of the scientific process freely and transparently while providing free access to new discoveries efficiently and quickly. PubPub contributed by allowing an alternative publishing platform than traditional journals. It aimed to reach more research paper readers by providing affordable online publishing, disseminating and feedback receiving of a research paper. It was open and allow exchange of ideas and feedback from other researchers.

Communities were supported by funders, scholars and other communities. Fundings allowed communities to operate and support publication of research about open science or research done using the framework of open science. Most communities are non-profit organisations, while some organisations that focused on doing open science research receive funding from European Commission or government funders (Bahlai, et al., 2019). There are funding agencies to support grants and funding for research, while crowdfunding could be another way to gather funds (OECD, 2015). For instance, Mozilla established Mozilla’s Open Science Mini-Grants to fund and support open science research in biomedicine and artificial intelligence. Scholars also played an important role in supporting the communities. On top of performing pre-registration and replications, they can spend time peer-reviewing and commenting on others' work in communities, so as to construct knowledge in a collaborative and transparent manner. Communities support each other since they serves different purposes and goals, which help each other throughout the publication process and spreading the idea of open science. Communities would educate the public and academics about open science and its practices, facilitate open science by providing communication platforms and aid the publication reviewing and research sharing.

The coordination between and within communities

The way the coordination built between communities is a combination of traditional pattern and contemporary evolution. Historically, the lay/expert collaboration, which means the partnership between researchers and activists, plays an irreplaceable role in the democratizing science to come out with doubt with existing science and spread out new knowledge (McCormick, 2009). For example, the Communities for Nuclear Information (CNI) was formed by American scientists and advocate John Gofan with four Nobel Laureates in board to translate expert knowledge to the public. John Gofan and Nobel Laureates are the sole experts in the community, their presence could provide credibility to the community because of their authorized expert identity and thus they are able to broadcast new scientific argument to the lay persons.

To form the traditional pattern, not only celebrity scientist, or more precisely, scientist leader, is needed, but the participation of relative institutions and facilities is also critical. For example, Tilburg Universities has its own open science community with its website https://www.tilburguniversity.edu/research/open-science-community. Though it's mainly for the employees and students to talk about open science practices, other people are also welcome. Just by following its twitter account, people could join the community and enjoy the 3-O: Open access, open education, and open data, which means green self-publishing, open educational materials and open data are all available in the community. Academic libraries have long been regarded as the peak of traditional way of science. But with the passage of time, Academic libraries continue to be involved as key players in the open science through advocacy, building institutional data repositories and serving as hubs for scientific collaboration among others (Ogungbeni et al, 2018). The Lund University Libraries provide Lund University Library in Sweden is one of the best examples of libraries are taking towards open access (OA). Librarians created the Lund University OA Archive, LU, containing about 170 full-text OA articles to date and maintain the Directory Open Access Journal (DOAJ), which lists all known OA journals (about 1,200 to date) and allows many of their full texts to be searched and harvested (Bosc & Harnad, 2005).

The coordination is not just built in terms of communities and institutions. Researchers themselves are also highly enthusiastic in communication with each other. A famous example is the Peer Review's Openness initiative. Researchers agreed with that initiative will offer comprehensive reviews of manuscripts only when five requirements are fulfilled. Another prominent example is the Registered Reports (RR) initiative, which is community-led that highlights the data sharing within communities. With the worried with little incentive for researchers to join the coordination, the financial subsidy has already been available in some newly generated website. The launch of Science Open Review, endeavouring to increase both the quality and quantity of peer-reviews, provides reputation rewards for the reviewer and boost the transparency and openness during the process of peer-reviews (Aarssen & Lortie, 2012). All of these enable it to become a communication platform for both producers and reviewers. And the possible publishers could also get the articles that have already experienced peer reviewing and get ready to be published. A benign ecosphere for science has gradually come into being with lots of platforms like that growing up.

With the rapid development of technology and expansion of the internet, the coordination between communities have gone beyond the sole scientist-leading pattern and specific institutions, it has also increased in the social media sites. The social media tools like Twitter and blog have become another battlefield where experts and public audience exchange their opinions and disseminate new academic research work new knowledge. And surprisingly, twitter has been found as relative effective tools for open science promotion, with a study stated that twitter citations were significantly faster than citations in the traditional media, as 39% of citations in the sample referred to articles less than 1-year-old, and 15% of the citations referred to articles that were published on the same day (Priem & Costello, 2010). Social media can also be utilized by individual researcher to promote the sharing of their research data, an important software for science. For example, the Open data science will often post some journal with its data in the twitter to foster the exchange of innovative ideas in science. But to date, there has not been any study examining the relationship between primary data and social media use (Zhu, 2019).

Luckily, today, the coordination between communities is not limited to software but also expand to the hardware due to technology development. Open science hardware (OSH) demonstrates access to the professional, standardized, and expensive lab instruments that are used to be monopolized by big scientific institutions and universities. Now it's not only possible for grassroots to get in touch with the first-hand scientific news and knowledge, but also possible for them to conduct the first-hand scientific research practice and get the first-hand scientific research experience, which may significantly help then better understand the scientific world.

What's more, now with the assistance of the internet, the audience begins to have access to the latest research outcome via the internet. One prominent example is the ScienceOpen, where people can find 60 million publications from 26 million authors. It collaborates different types of communities from libraries like Open Library of Humanities to presslike UCL PRESS and online learning network like OPEN SCIENCE MOOC. It provides researchers a DIY platforms for the launching of open access journal and/or open access press.

Achievements and Challenges

Grassroot movements emerged in distinct forms which can vary from legitimate to illegitimate, individuals to organization-supported. Taking a look at the following examples might help revealing the reality of operation, degree of contribution of an open science tool or community and relative influence to scholars. In concern of improvement in accessibility, SCI-HUB and Unpaywall are well-known engines for free access to research papers despite their difference in legitimacy and mode of functioning.

One of the difficulties encountered by SCI-HUB is the unstable financial sources. Founded by Alexandra Elbakyan alone without the backup of a team in 2011 in Kazakhstan in response to costly research paper, SCI-HUB is a hacking engine that provides free access to both graits and toll papers (Elbakyan, 2019). Since archiving of some openness-exemplified journals began to charge in 2015, SCI-HUB's operation targets at bypassing paywalls instead of inclusion of all literature (Himmelstein et al, 2018). Despite the lack of funding at the beginning (Elbakyan, 2019), to sustain the operation and management of the engine, SCI-HUB received donations from its users primarily via Bitcoin to avoid banking blockades or government seizure of funds (see figure 1). Living on donations for years, Himmelstein et al (2018) suggested that SCI-HUB kept expanding its database and coverage of research papers which has maintained an annual growth of 88% - from reliance on Library Genesis repository to having its own storage of over 56 million articles from the corpus of scholarly literature, which equals to about 69% of all papers.

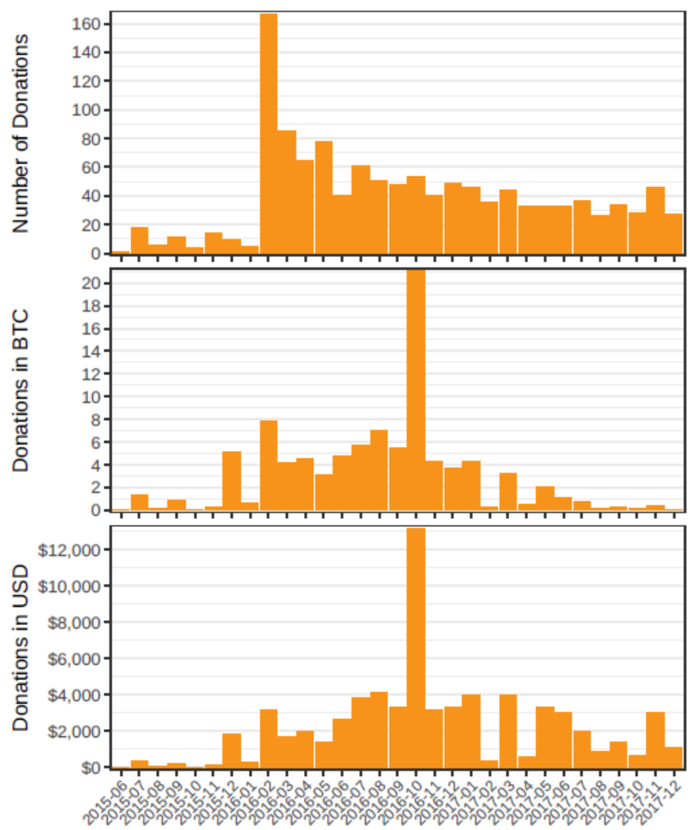


Figure 2: Bitcoin donations to Sci-Hub per month. *For months since June 2015, total bitcoin donations (deposits to known Sci-Hub addresses) were assessed. Donations in USD refers to the United States dollar value at time of transaction confirmation (Himmelstein et al, 2018)*

The success in overcoming the constraint of paywall has significantly increased the degree of open access in science which is evident from the dramatic increase in citation of toll papers and enormous number of download events on SCI-HUB (see figure 2, 3). Himmelstein et al (2018) suggested that with the exclusion of citations to articles in open access journals, 96% of citations since 2015 were presented in SCI-HUB repository and the accessibility of toll articles has already surpassed that of the University of Pennsylvania, a leading research university. Bohannon and John (2016) also mentioned that the number of Sci-Hub downloads increased from 42 million in 2015 to 75 million in 2016. Traditionally, to access papers without payment, scholars could either send an email or write a letter to the authors to ask for a reprint of the manuscript or search online via Google Scholar for limited amount of non-paywalled articles. It is time-consuming and some old papers might not be available due to loss of contact to authors.

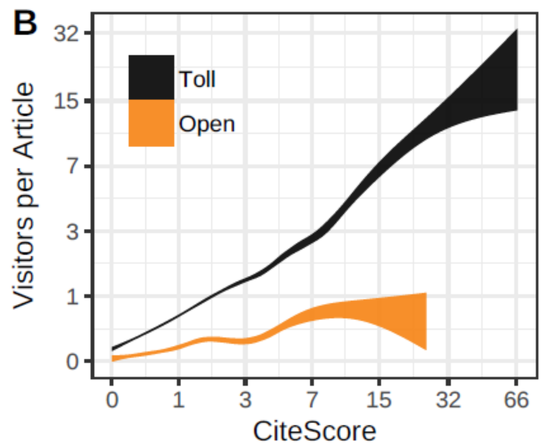


Figure 2: The association between 2015 CiteScore and average visitors per article is plotted for open and toll access journals *(Himmelstein et al, 2018)*

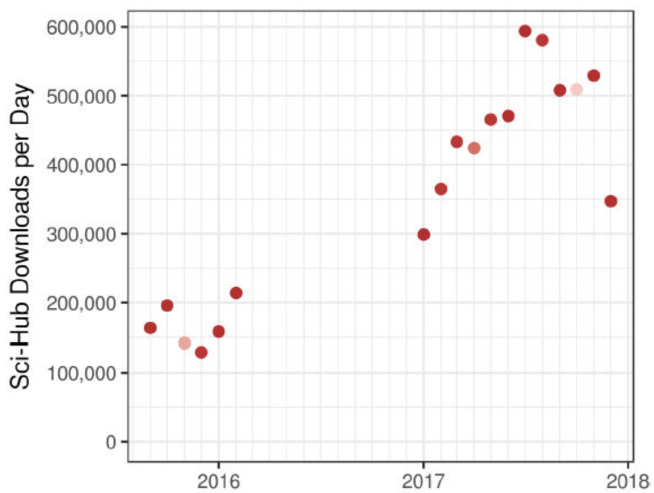


Figure 3: Downloads per day on Sci-Hub for months with access logs*. The number of articles downloaded from Sci-Hub is shown over time. Sci-Hub access logs were combined from two releases: covering 27,819,963 downloads from September 2015 to February 2016 and covering 150,875,862 downloads from 2017. The plot shows the average number of downloads per day for months with data. (Himmelstein et al, 2018)*

Under the current system, however, there is a trade-off between open access and legal consequences. Both Elsevier and the American Chemical Society (ACS) filed a lawsuit against Sci-Hub in 2015 and 2017 respectively in response to copyright infringement. Judged by the US court, SCI-HUB lost both lawsuits but did it not suffer huge monetary and systematic damages because of the absence of assets and hosting root in US. Yet, SCI-HUB has lost its original domain name and its domain has been cycling since then (Chawla, 2017; Himmelstein et al, 2018). Even though SCI-HUB remains illegitimate, its robust impacts to the transformation of open science is inevitable.

A legal alternative could be Unpaywall though there is limited accessibility. Unpaywall aims at providing access to freely available scholarly articles (Piwowar & Priem, n.d.). Comparing to SCI-HUB, Unpaywall is less influential in expansion of open access due to the fact that over 70% of the papers are locked or paywalled (Himmelstein et al, 2018). Relatively it is less popular among users concerning the fact that it only supports 200.000 active users (Priem, 2019). The difference between SCI-HUB and Unpayway has highlighted the controversy of open science - whether it shall be criticized as taking others effort for granted or the solution to comprehensive exchange of scientific advancements.

In addition to individuals' contributions in open science, advocates of open science have formed their own communities and some of them are even supported by universities. One of the organized communities is the UK Reproductivity Network (UKRN). As aforementioned, reproductivity is a challenge in science and UKRN is a researcher-led organization focuses on investigation of factors for robust research. Led by a steering committee composed of four researchers from distinct universities namely University of Bristol, University of Oxford, University of Cardiff and University of Edinburgh, UKRN is more like an administrative center that provides coordination among researchers (University of Bristol, 2019). Supported by the committee, two groups of the network are directly engaged in replication which are Local Network Leads and Stakeholders. Local Network Leads is made up of researchers who are also representatives of each university. While receiving training and advice from the steering committee, the local network aims to provide support within universities as well as promotion of high quality research for instance they assisted "Reproducibilitea journal clubs" in papers review (Hunter, 2019). For Stakeholders, it is formed by representatives of research-related organisations which includes but not limited to funders and publishers. A mechanism for the exchange of ideas is established with the integration of different shareholders of open science - UKRN can obtain feedback from research communities while ensuring its initiatives are aligned with strategy and activity of the stakeholders (University of Bristol ,2019).

The case of UKRN has pinpointed both top-down and bottom-up influence of grassroot communities in expansion of open science. From a top-down perspective, the professionals have taken initiatives in improving the quality of science by doing research on research and the concept of open science is thus expanded from researchers to other stakeholders such as publishers. From the bottom-up side, the response from those influenced by expansion of open science helps consolidating the interactive mechanism in open science.

Outlook

As science is opening up, the future would only see scientific research become increasingly accessible and transparent for the public. Even though progress has been made in the past decades, there is still a long way to achieve truly public and transparent research. Improvements are still warranted in multiple areas of the open science domain, and possible advancements could bring the scientific community to a new realm.

In terms of technological advancement, information and tools used in scientific research would be more easily available with the aid of the new internet services. Cloud Storage is one of the new technologies that could transform data transparency in scientific research. Cloud services dedicated to open science network allows the community to retrieve supporting data from studies, improving the credibility of studies. Businesses are actively developing cloud services for the scientists, with the aim to help with fostering open science principles (Business Network, 2019). Recent year also seen a number of internet platforms designed for the scientific community to publish their findings, instead of traditional journal publishing, which can allow researchers to include all their information and data in their publishing, and allowing scientists from all over the world to review and replicate. By establishing an online community for verification of information, findings published within the community have more reproducibility and credibility (Mirowski, 2018). Technological progresses are thought to be able to break through the limits of space and time, allowing unrestricted access of information with low cost, facilitating the exchange of information between scholars and public.

Challenges for open science is still apparent in the future. The biggest deterrent for scientists to disclose their data collected and methodologies are the concerns of incremental costs (Allen et al, 2019). Funding for pieces of research are limited and transparency in research would incur extra cost in preparing and finding platform to release the extra data. Therefore, the scientific community finds it reluctant to make investment to develop advantageous tools and technologies to aid in the progress of open science, even when arguments for open science suggests that long term efficiency would be enhanced. Regardless, the challenges can be mitigated if an incentive structure is in place, to reward or accordingly to open science principles (Ali-Khan et al, 2018). Organizational challenges are other obstacles anticipated for the future of open science. For the moment, researches and studies are taken place within different institutional networks, with varying regulations and guidelines. Even in the same institution, different aspects of science have varying standards of framework and culture, which is difficult to reconcile. Openness of the research under these variables could be hard to come to a single uniform platform and framework for disclosure guidelines and input output presentation methods. Nevertheless, the barriers over these differences are possible to overcome, if a clear legislation on an international institutional level is achieved.

Open Science is often regarded as the future for the scientific community. With ever-progressing endeavors, the scientific community is closer than ever to the ideal openness put forward by the theorist years ago. With the world growing closer and closer together online, the scientific community is also moving into a new dynamics. Credibility and transparency is going to be the new focus of research and findings published in the future. The discussion for open science must keep going to find the appropriate framework and guidelines for the open science principles, taking us into a truly open future.

Quiz

Write 10 easy quiz 4 multiple choice questions to ensure reading and understanding of this chapter:

1. What is meant by the “replication crisis”?
    a. **Scientists are unable to replicate each other’s work**

   b. Scientists are fabricating false evidence to create new findings
   c. Scientists are unable to make innovative research
   d. Scientists are replicating each other's work only
2. Why were replications not a major focus of the scientific community before 2011?
   a. Replications are hard to complete
   b. Replications holds little scientific importance
   c. Replications are less rewarded in academia than innovation
   d. Replications requires permission from the original researcher
3. What is not part of the Open Science principles?
   a. Rigority
   b. Reproducibility
   c. Accumulation of Knowledge
   d. None of the above
4. What gives credibility to a research?
   a. Transparency in research process and Open Access in Publication
   b. The reputation of the researchers
   c. Positive comments from the scientific community
   d. Similarity between the work of researchers and peers
5. What is the importance of a pre-registration?
   a. It allows the public to understand the research in simple terms
   b. It prevents researchers to change the hypotheses and methodology in the middle of the research
   c. It provides a framework for anyone to replicate the research
   d. It allows flexibility in the research by providing different options for the researchers
6. What is not the benefits of building communities for open science?
   a. Open Science are promoted and advocated
   b. It allows more research to be done in a more efficient way
   c. It allows academics to exchange ideas on open science
   d. The above are all benefits of building communities for open science
7. What is Open Science Hardware?
   a. A tool allowing individual communities to access professional, standardise lab instruments
   b. A tool for people to access open science information and each other's work
   c. An instrument aims to guide scientists to follow open science principles in their research
   d. An add-on for scientists to run open science applications on their computers
8. How does journal helps with open science application?
   a. It helps solidify policies within communities
   b. It usually requires data openness for research to be published
   c. It encourages authors to be more willing to disclose their data
   d. All of the above
9. How did SCI-HUB help with the grassroot movement?

a. It uses blockchain technology to store data collected from research permanently and openly
b. It is an open resource library for the researchers to look for unreplicated research
c. It is a hacking engine to provide access for graits and toll papers
d. It is an online platform for scientists to review each other's research

10. What is not the influence of grassroot communities on the open science movement?
    a. They help take the initiatives to practice and advocate open science principles
    b. They provide responses for the open science principles
    c. They help create the standards for open science for all the scientific community to follow
    d. They create interconnected communities to exchange opinions over each others work

References

Aarssen, L. W., & Lortie, C. J. (2012). Science Open Reviewed: An online community connecting authors with reviewers for journals. *Ideas in Ecology and Evolution*, *5*, 78-83. 10.4033/iee.2012.5b.16.f

Ali-Khan, S., & Jean, A. (2018). Identifying the challenges in implementing open science. *MNI Open Research*. Retrieved from https://mniopenresearch.org/articles/2-5

Allen, C. & Mehler, D. (2019). Open Science Challenge, benefits and beyond. *PLOS Biology*. Retrieved from https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000246

Bahlai, C., Bartlett, L. j, Burgio, K. R., Fournier, A. M. V., Keiser, C. N., Poisot, T., & Whitney, K. S. (2019, June 14). Open Science Isn't Always Open to All Scientists. Retrieved from https://www.americanscientist.org/article/open-science-isnt-always-open-to-all-scientists.

Baker, M. (2016). Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help. *Nature*, *533*(7604), 452-455. https://doi.org/10.1038/533452a

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425. doi:10.1037/a0021524

Bohannon, J. (2016). Who's downloading pirated papers? Everyone. *Science*, *352*(6285), 508-512. doi: 10.1126/science.352.6285.508.

Bosc, H., & Harnad, S. (2005). In a paperless world a new role for academic libraries: providing open access. *Learned Publishing,* 18, 95-99. https://doi.org/10.1087/0953151053585028

Bressan, P. (2019). Confounds in "failed" replications. Frontiers in psychology, 10, 1884. https://doi.org/10.3389/fpsyg.2019.01884

Bueno-DeLaFente, G. (2016). Challenges and Strategies for the success of Open Science. Foster. Retrieved from https://www.fosteropenscience.eu/content/challenges-and-strategies-success-open-science

Business Network. (2019). Why Open Science is the Future (And how to make it happen). Science Business. Retrieved from https://sciencebusiness.net/report/why-open-science-future-and-how-make-it-happen

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... & Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. Science, 351(6280), 1433-1436. doi: 10.1126/science.aaf0918

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637-644. doi:10.1038/s41562-018-0399-z

Caplan, A. L., & Redman, B. K. (2018). Reproducibility. In *Getting to Good: Research Integrity in the Biomedical Sciences*. (pp. 95-111). Cham: Springer International Publishing.

Chawla, D. S. (2017). Court demands that search engines and internet service providers block Sci-Hub. *Science*, Retrieved from https://www.sciencemag.org/news/2017/11/court-demands-search-engines-and-internet-service-providers-block-sci-hub

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., … Schulte-Mecklenbeck, M. (2018, November 16). 7 Easy Steps to Open Science: An Annotated Reading List. https://doi.org/10.31234/osf.io/cfzyx

Elbakyan, A. (2019, March 31). Sci-Hub and Alexandra basic information [Blog post]. Retrieved from https://engineuring.wordpress.com/category/sci-hub/

Fidler, F., Chee, Y. E., Wintle, B. C., Burgman, M. A., McCarthy, M. A., & Gordon, A. (2017). Metaresearch for evaluating reproducibility in ecology and evolution. *BioScience*, *67*(3), 282-289. doi: 10.1093/biosci/biw159

Gray, J. (2010). *Launch of the Panton Principles for Open Data in Science and 'Is It Open Data?' Web Service.*

Retrieved from

https://blog.okfn.org/2010/02/19/launch-of-the-panton-principles-for-open-data-in-science/

Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., Greshake Tzovaras, B., & Greene, C. S. (2018). Sci-Hub provides access to nearly all scholarly literature. *eLife*, 7, e32822. doi:10.7554/eLife.32822

Hunter, N. (2019, March 15). Improving the way research is done: the UK Reproducibility Network [Blog post]. Retrieved from https://www.linkedin.com/pulse/improving-way-research-done-uk-reproducibility-network-natalie-hunter

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Johnson, V. E., Payne, R. D., Wang, T., Asher, A., & Mandal, S. (2017). On the reproducibility of psychological science. *Journal of the American Statistical Association*, *112*(517), 1-10. https://doi.org/10.1080/01621459.2016.1240079

Klein, O., Hardwicke, T.E., Aust, F., Breuer, J., Danielsson, H., Mohr, A.H., IJzerman, H., Nilsonne, G., Vanpaemel, W. and Frank, M.C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1). http://doi.org/10.1525/collabra.158

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?. *American Psychologist*, *70*(6), 487. doi: 10.1037/a0039400

McCormick, S. (2009). From “Politioco-Scientists” to Democratizing science Movements. *Organization & Environment*, 22(1), 34-51. https://doi.org/10.1177/1086026609333419

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, *48*(2), 171–203. https://doi.org/10.1177/0306312718772086

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., ... & Vanpaemel, W. (2016). The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. *Royal Society Open Science*, *3*(1), 150547. https://doi.org/10.1098/rsos.150547

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., Simonsohn, U., Wagenmakers, E. J., Ware, J. J. and Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Murray-Rust, P., Neylon, C., Pollock, R., Wilbanks, J. (2010). *About Panton Principles.* Retrieved from https://pantonprinciples.org/about/index.html

Murray-Rust, P., Neylon, C., Pollock, R., Wilbanks, J. (2010). *Panton Principles, Principles for open data in science.* Retrieved from https://pantonprinciples.org

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., Crowston, K. (2012). The future of citizen science: emerging technologies and shifting paradigms. *Ecol Environ*, *10*(6), 298–304. https://doi.org/10.1890/110294

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422-1425. doi: 10.1126/science.aab2374

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615-631. https://doi.org/10.1177/1745691612459058

OECD (2015-10-15), “Making Open Science a Reality”, *OECD Science, Technology and Industry Policy Papers, No. 25*, Paris: OECD Publishing. http://dx.doi.org/10.1787/5jrs2f963zs1-en

Ogungbeni, J. I., Obiamalu, A. R., Ssemambo, S., & Bazibu, C. M. (2018). The roles of academic libraries in propagating open science: A qualitative literature review. *Information Development*, *34*(2), 113-121. https://doi.org/10.1177/0266666916678444

Open Definition. (n.d.). *The Open Definition.* Retrieved from https://opendefinition.org

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*(6), 657-660. https://doi.org/10.1177/1745691612462588

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). doi: 10.1126/science.aac4716

ORION. (n.d.). *What is Open Science?* Retrieved from

https://www.orion-openscience.eu/resources/open-science

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science*, *7*(6), 528-530. https://doi.org/10.1177/1745691612465253

Piwowar, H & Priem, J. (n.d.). Unpaywall: Frequently Asked Questions. Retrieved from https://unpaywall.org/faq

Priem, J. (2019, February 4). Unpaywall extension adds 200,000th active user [Blog post]. Retrieved from https://blog.ourresearch.org/unpaywall-200k-users/

Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceeding of the Association for Information Science and Technology*, *47*, 1-4. https://doi.org/10.1002/meet.14504701201

Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H. (2016). The academic, economic and societal impacts of Open Access: an evidence-based review. *F1000Research*, *5*.

University of Bristol. (2019, October 17). UKRN Terms of Reference. Retrieved from https://www.bristol.ac.uk/psychology/research/ukrn/about/tor/

Schooler, J. W. (2014). Metascience could rescue the 'replication crisis'. *Nature News*, *515*(7525), 9. doi: 10.1038/515009a

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

Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual review of clinical psychology*, *15*, 579-604. https://doi.org/10.1146/annurev-clinpsy-050718-095710

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632-638. doi: 10.1177/1745691612463078.

Zhu, Y. (2019). Open-access policy and data-sharing practice in UK academia. *Journal of Information Science*, 25, 1-12. https://doi.org/10.1177/0165551518823174

7 - Scaling up - movement grows

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Man Wai	Lai	https://www.researchgate.net/profile/M_Wai_L	osf.io/nwr8d	u3547226@connect.hku.hk	yllmy99@yahoo.com.hk	3035472264
Zhi Ying (Cherry)	Kwong	https://www.researchgate.net/profile/Cherry_Kwong2	osf.io/vqhr6	kzcherry@connect.hku.hk	cherry.kwong@yahoo.com.hk	3035474341
Yuyi	Chen	https://www.researchgate.net/profile/Yuyi_Chen2	osf.io/t7g8j	u3546763@connect.hku.hk	yuyi.19329@gmail.com	3035467635
Hei Ting	Wong	https://www.researchgate.net/profile/Hei_Ting_Wong2	osf.io/9aznu	u3547024@connect.hku.hk	wongasteria@yahoo.com	3035470242
Tsz Yan	Chow	https://www.researchgate.net/profile/Tsz_Yan_Chow2	https://osf.io/xumns/	u3546593@connect.hku.hk	katiechow567@gmail.com	3035465936
Yee Yan	Chan	https://www.researchgate.net/profile/Chan_Yee_Yan	https://osf.io/gsutr/	u3548199@connect.hku.hk	eileenchan@msn.com	3035481992
Myrte Louise	Druyvesteyn	https://www.researchgate.net/profile/Myrte_Druyvesteyn	https://osf.io/hfqse/	mldruyv@hku.hk	m.l.druyvesteyn@gmail.com	n.a

Managerial summary

The open-science movement is growing constantly in recent years in a variety of domains of science. Journals and researchers set new standards for article submission including promoting the sharing of data, materials or research plans set before it was conducted (preregistration) so as to give more incentives to researchers to go open-science. And these standards were gradually being applied in many journals. There were also a range of studies that recorded improvements in open-science practices over the years. The following chapter will introduce you the reasons for open-science needing to grow, how it grew in the reality, and by how it could grow.

In depth report

In the last chapter, we introduced the replication crisis and how it has led to the development of the grass-root movement of open science. In Chapter 7, we will take a closer look at the scale of growth of this movement, how it has developed into mainstream psychology and the challenges ahead.

The academic reward system does not promote open practices but instead values innovation. This may have undermined the importance of verification practices. For example, Franco, Malhotra, and Simonovitsb stated that more often than not, statistically significant results are published instead of the null results (as cited in Nosek et al., 2015). Yet results from all experiments carried out are important for us to understand the evidence present for theories more accurately.

As researchers now lack incentives to share their work in a more transparent way (even though many know that there is a need for open science), the Transparency and Openness Promotion (TOP) Guidelines Committee suggested that a new standard for journals' procedures and policies for publication can be adapted to serve as incentives (Nosek et al., 2015). Members of the TOP Guidelines Committee consist of scholars specialized in different fields of science. The committee is involved in evaluating the adoption and effectiveness of the guideline to improve its quality, as well as its interdisciplinary applicability (“TOP Guidelines,” 2019). All versions of TOP guidelines are available on the Centre for Open Science (COS) website as the committee is also sponsored by COS. Journals can choose which version to adopt as well as to what degree will they follow the guidelines and this allows custom-made regulations and minimize the difficulty of implementing these guidelines (Aalbersberg, 2019).

The TOP guidelines, when well-followed, should largely increase the transparency, openness, and reproducibility of articles. The guidelines basically consist of seven elements:

1) Citation standards: citing data, code, and any important materials in the article,
2) Data transparency: data are available in a trusted repository such that analyses can be reproduced,
3) Code transparency: codes/analytic methods are available in a trusted repository such that analyses can be reproduced,
4) Research materials transparency: important materials used are accessible through a trusted repository,
5) Design and analysis transparency: journals require design transparency to a certain standard for review and publication,
6) Pre-registration of studies: submitting pre-registration reports before carrying out the experiments,
7) Pre-registration of analysis plan: pre-registration reports should include an analysis plan such as statistical tests that will be performed,
8) Replication: journals encourage the submission of replication studies or submission of Registered Reports (which contains peer review before seeing the results)

For each element, the guideline provided 4 levels such that journals can be well-informed before making a decision to follow the guideline as well as to which level in each element will they follow according to their needs (Nosek et al., 2015). Journals can also seek support from the information commons and support team from COS or the TOP guidelines committee about the selection and adoption of standards or give suggestions to the guideline.

To provide more incentives for researchers to go open science, COS also has the Open Science Badges system. These badges acknowledge the readers that relevant content (such as data, codes, and pre-registrations) of an article is constantly available in a location, which will be further discussed below. A total of 66 journals that issue open science badges is also available on their website (cos.io/our-services/open-science-badges/). Take Psychological Science as an example, it issues badges for “open data,” “open materials,” and “preregistration” according to Eich (as cited in Nosek et al., 2015).

It is more and more necessary to carry out Open Science. Firstly, Open Science provides efficiency for researchers to work on their research, saving much of their research times which often take them years. Under Open Science, researchers can get greater access to scientific inputs by others, such as the research data (OECD, 2015). It can reduce the costs of creating new data again as the data have been generated by the other authors and open for being used. Collaboration can then be achieved to work for making the research better with collaborative efforts, according to Munafò et al. (2017). Open Science also helps to monitor the quality of the research. This is because the scientific outputs by the researchers are open for access by anyone, welcoming evaluation and replication by the others (OECD, 2015). As stated by Munafò et al. (2017), there are peer reviews, reproducibility checks, some protocol checklists and more solutions developed so as to keep track of the quality of the research authors are working and have worked on (See Table 7.1). As could be monitored through various Open Science methods as mentioned just now, researchers would less likely to risk their integrity in these open research and try to be as rigorous as they can to keep the information accuracy in their work.

The seminal articles serves as a reminder to researchers and journals that there is a replication crisis going on and to deal with the problem, different parties must work together to increase the transparency, openness, and reproducibility of science. Although not all of the work submitted to journals can instantly attain level 3 in the TOP guideline in all standards, at least various journals (63 journals until now) adopted the badges system that encourages open science ("Open Science Badges", 2019). More recent studies focus on the potential challenges when adopting these standards as well as how researchers are also working on those challenges to improve open science by giving suggestions to the standards set for journals. Tackett et al., (2017) pointed out in their study the challenges that clinical psychology science may be facing in terms of replicability as well as recommendations to deal with the issue. One particularly important challenge is that there is an unclear definition of replicability in the clinical psychology field as clinical trials may differ depending on the client’s situation or there is a lot of flexibility in applying the manualized treatment. The authors suggested that expert groups or consortia need to collaborate more to set up standardize replicability principles (Tackett et al., 2017). The funding available also often affects the sample size of clinical research and thus there is no guarantee of consistent sample sizes throughout different replication studies. Large clinical samples may also be difficult to collect since there are simply not that many samples available. To deal with these two challenges, the authors proposed that encouraging cross-site collaboration, as well as using the same measurements across labs may help to solve the problem of sample unavailability or insufficient funding (Tackett et al., 2017). Another study by Dickersin and Mayo-Wilson (2018) also attempts to give solutions to the current problems. Dickersin and Mayo-Wilson basically stated that the rules that we have now on study registration, research protocol, and data collection are unclear. Dickersin and Mayo-Wilson also mentioned that there is a need to standardize those rules and make them clear and applicable to different fields of science or different study designs. They also suggested that unpublished findings should be made available on study registers. So as to avoid publication biases. It is clear that different parties are now working together to improve the standards of open science in journals/articles.

Table 7.1.
A manifesto for reproducible science.

What to improve		Potential solutions
Research Method	Get rid of cognitive bias	Blinding
	Improve methodological training	Rigorous training and education on statistics and research methods for researchers
	Collaboration science	Multi-site studies Collective data collection
Reporting	Promote pre-registration	Open Science Framework Registered reports
	Improve reporting quality	Reporting protocol checklist
	Eliminate conflict of interest	Exclude financial and non-financial conflicts Disclose conflicts of interests
Reproducibility	Open Science and transparency	Open data and materials Pre-registration
Evaluation	Peer review diversity	Preprints Peer review on pre- and post- publications
Incentives	Reward quality, reproducible and open-science works	Badges Funding replication studies Open-science work promotion

Note. Adapted from “A manifesto for reproducible science” by M. R. Munafò et al., 2017, Nature human behaviour, 1(1), 3. Copyright 2017 by M. R. Munafò et al.

Adoption Rate

The above reviewed the potential threats to replicability and respective solutions. The following section will look into current adoption of open science in several aspects, namely registered reports, data and code sharing, badges, pre-registration, pre-prints and use of OSF.

Registered Reports

A recent survey by Paluck et al. (2018) covered two parts: (1) trends in open science practices; and (2) a survey of researchers' awareness, attitudes and behavior towards (i) posting data and code online, (ii) posting study instruments, and (iii) and pre-registration. Data was obtained from State of Social Science (3S) Survey with a representative sample of elite social science researchers in economics, political science, psychology, and sociology. In specific, a random set of authors who published within 2014-2016 in 10 most cited journals and PHD students from top 20 North American departments of each discipline were drawn with monetary incentives.

For part 1, there has been a rapid rise of adopting open science practice from 49% before 2010 to 84% in 2017. Among the three types of behaviors, posting data and code online was most widely adopted, followed by posting study materials, with mild increase in pre-registration from 1996 to 2017 (see Figure 7.1). Performance of experimental and quantitative/non-experimental designs outperformed that of qualitative/theoretical designs in pre-registration (see Figure 7.2). Nevertheless, awareness and stated support of qualitative research remained at a high level, let alone the underperformance. Notably, equipment or techniques of measures in experimental designs were better quantified for reproducibility.

For part 2, in general, attitudes/ support were more well-received than actual behaviors across all disciplines (see Figure 7.3). An underestimation of general adaptation of open science was found by most scholars. Both bottom-up and top-down structural factors accounted for such increase (i.e., authors networking and technical innovations).

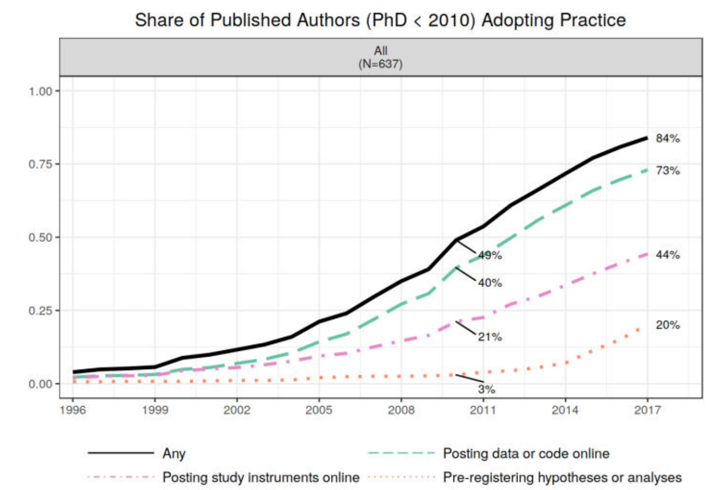


Figure 7.1. Open science practice adopted by published authors.
Noted. From Paluck et al. (2018).

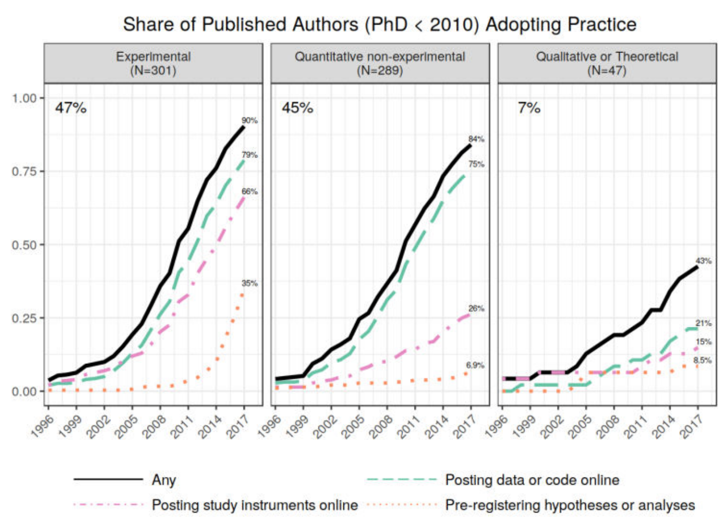


Figure 7.2. OS Practice adopted in three research types.
Noted. From Paluck et al. (2018).

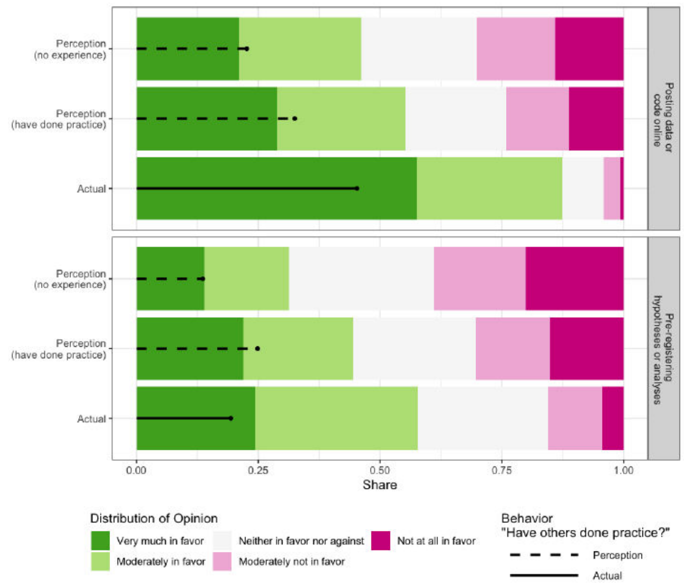


Figure 7.3. Perceived versus actual opinions/ behaviors.
Noted. From Paluck et al. (2018).

Data and Code Sharing

Several publishers and funders (e.g., PLOS) applied policies like Data Availability Statements to promote reproducibility of research data, with increased compliance (see Figure 7.4; Federer et al., 2017). Nevertheless, of those who complied, only about 20% indicated the data was in repository as preferred while most of the others simply stated in article/ supplementary (see Figure7.5), reflecting the needs of more stringent policies.

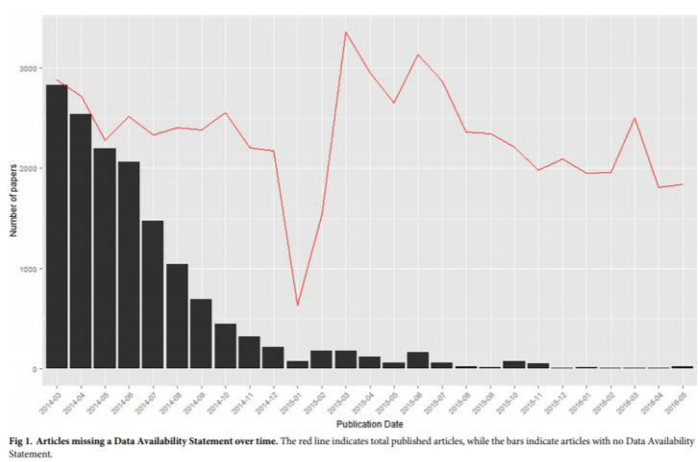


Figure 7.4. Articles missing a Data Availability Statement over time.
Noted. From Federer et al. (2018).

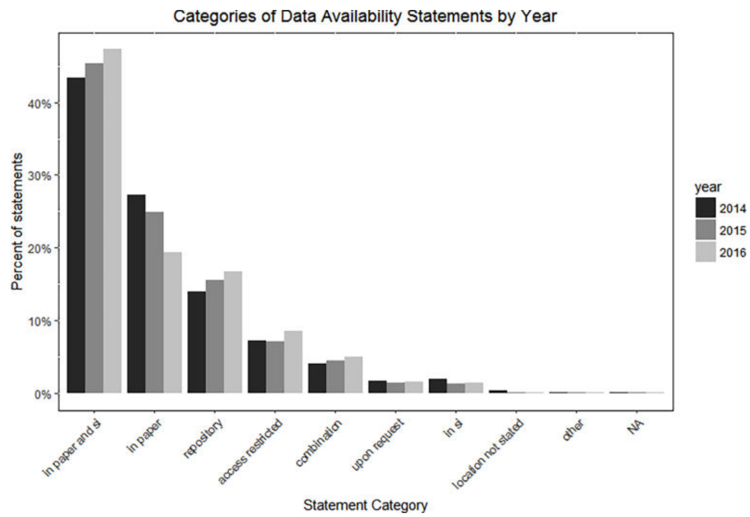


Figure 7.5. Distribution of statements across categories by year.
Noted. From Federer et al. (2018).

The introduction of two types of badges, namely Open Data badges and Open Materials badges, accelerated reported open data in Psychological Science (PSCI) from less than 3% to 39% from 2014 to 2015(see Figure 7.6; Kidwell et al., 2016). Meanwhile, the general higher rate of sharing may account for the limited increase in open materials (see Figure 7.7). There was an increased use in independent repositories which guaranteed better preservation of data. Nevertheless, the reported open data or material did not guarantee correct, usable and complete data, as well as materials, which were more prevalent in the presence of badges (see Figure 7.8 & 7.9).

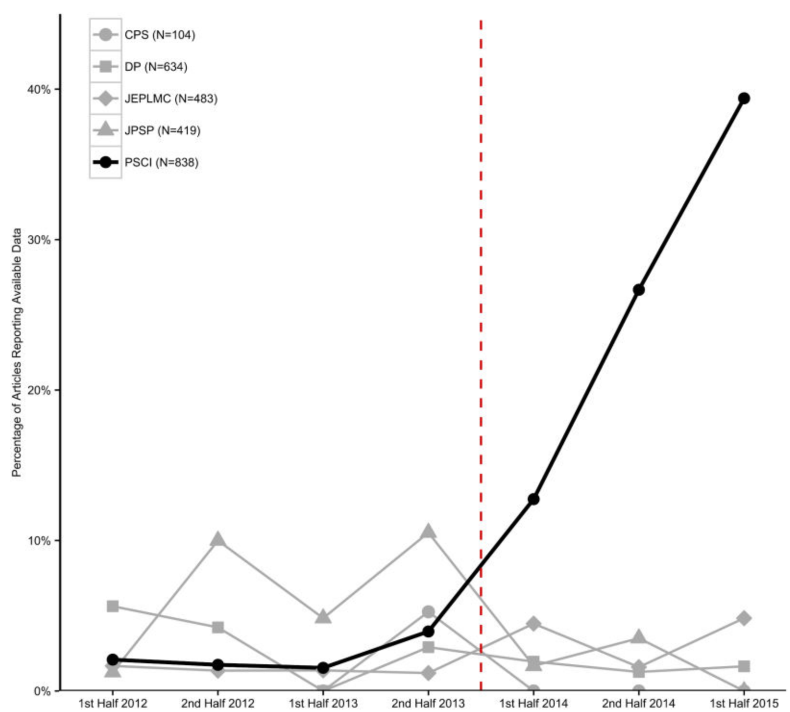


Figure 7.6. Reportedly available data.

Noted. From Kidwell et al. (2016). Darker line indicates PSCI; red dotted line indicates badges being introduced in PSCI, but not in other journals.

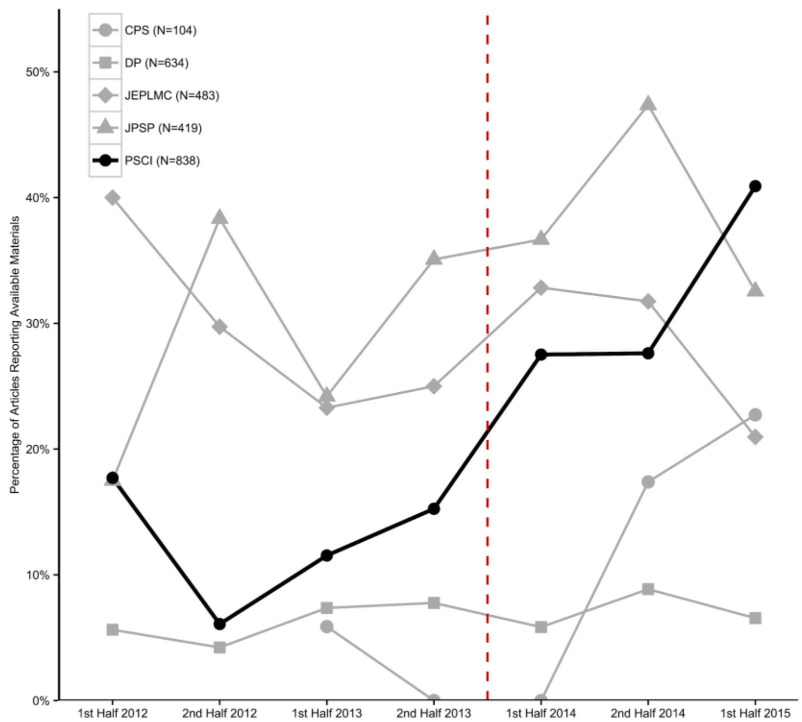


Figure 7.7. Reportedly available materials

Noted. From Kidwell et al. (2016). Darker line indicates PSCI; red dotted line indicates badges being introduced in PSCI, but not in other journals.

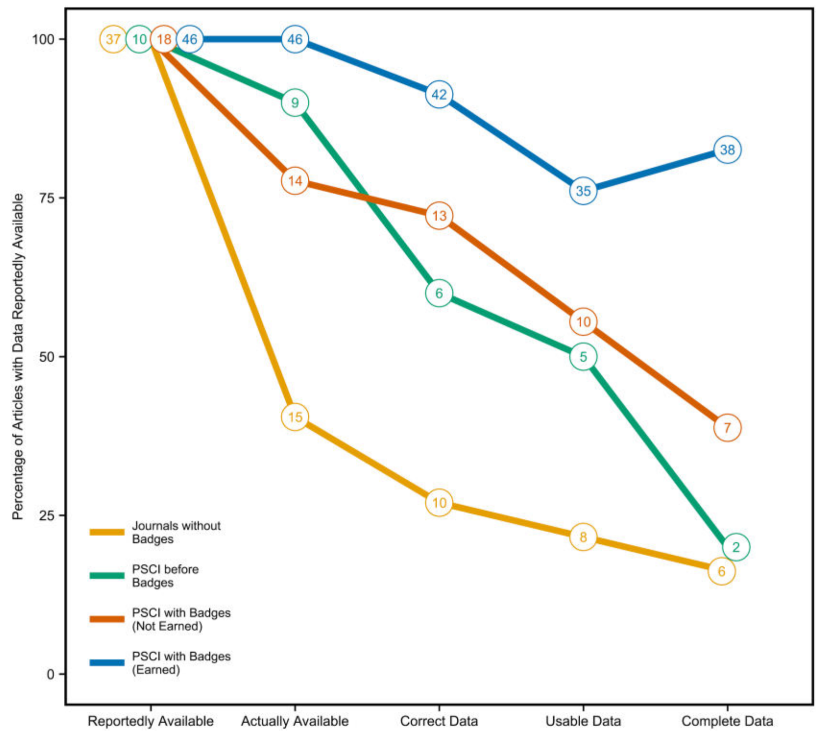


Figure 7.8. Actually available, correct, usable, and complete data.

Noted. From Kidwell et al. (2016).

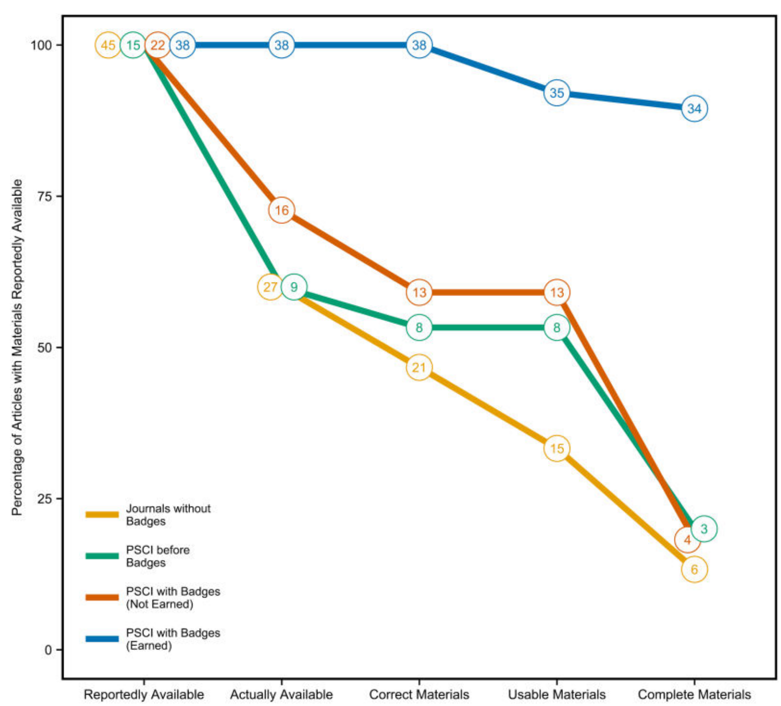


Figure 7.9. Actually available, correct, usable, and complete materials.

Noted. From Kidwell et al. (2016).

Badges

As mentioned, under TOP guidelines, badges are awarded for those that conduct rigorous, transparent and reproducible research, acknowledging their contribution to open science. This is a good incentive for scholars to enhance their efficiency and effectiveness of their work (Munafo, 2017).

Journals facilitate informed decision-making by defining multiple levels and standards, acknowledging the variation in research transparency. While some standards may not be relevant for a journal, different operations, norms and expectations are included when deciding. Journals with high visibility such as American Economic Review adopted strong policies requiring data sharing, but few psychology journals are comparable (Nosek et al, 2015).

There are multiple badges- 'Open Data Badge' and 'Open Materials Badge' for articles archiving data and materials publicly, that are mentioned above; 'Preregistered badge' or 'Preregistered+Analysis Plan badge' are also rewarded to preregistration research. Journals may start with a standard that rewards adherence . For example, according to Psychological Science, around 25% of accepted articles earned at least one badge in the first year of operation (Nosek et al, 2015).

According to Kidwell et al (2016), without badges, the paltry percentage of reported sharing is only a gross exaggeration. In fact, less than 1% of the articles had publicly accessible data. 6 of 37 articles from journals without badges and 2 of 10 articles from PSCI before badges reported available data had accessible, correct, usable and complete data (Kidwell et al, 2016). Although TOP guidelines provides a clear standard for readers, to further increase the uptake of open data, materials and preregistration by recognising the participated authors, the badges system were developed shortly afterwards.

In sum, badges were associated with a dramatic increase in reported sharing and, a dramatic increase in actual data sharing. A review by Rowhani-Farid et al (2017) showed that this badging program is the only evidence-based incentive program linked to increased data sharing. As mentioned, Kidwell et al. (2016) reported that, before the badge system began, authors publishing in Psychological Science were unlikely to make their data publicly available , like other authors from other leading journals (about 3% of authors did so). After a year, almost 40% of authors in Psychological Science posted their data, whereas other journals had no increase (Nelson et al, 2018).

Pre-registration

There are three types of pre-registrations namely the unreviewed pre-registrations, reviewed pre-registrations (registered reports) and registered replication reports (RRR) respectively. Researchers could create a detailed description of their plans and this record could be shared with reviewers, editors as well as other researchers for unreviewed pre-registrations (Lindsay, Simons & Lilienfeld, 2016). While registered reports, researchers submit detailed proposals to journals before conducting studies (Lindsay et al., 2016). There are currently 209 journals e.g. AAS Open Research, Acta Psychologica, BMC Biology, Collabra etc adopt the registered report publications that include peer review prior to the research outcome and offerance of acceptance for the reports surviving the pre-study peer reviews (COS, 2019). For RRR, it refers to direct replication of an original finding (Lindsay, et al., 2016).

In recent years, the present culture focuses on means (reasoning biases and misuse of statistics), motive (publication) and opportunity (commitment to predictions) in order to make pre-registrations the norm for the rigor and robustness of research practices (Nosek, Ebersole, DeHaven & Mellor, 2018).

In terms of means, education modules and resources for effective pre-registrations are established to facilitate the planning of pre-registrations (Nosek et al., 2018), such as the online courses by Coursera, criteria for pre-registration badge credentials and pre-registration templates.

In terms of motive, existing culture has changed to incorporate stronger incentives for research rigor and reproducibility e.g. pre-registration is necessary to be published in journals that adhere to the International Committee of Medical Journal Editors policy according to the US law for clinical trials. Besides, thousands of journals and funders are subjected to the Transparency and Openness Promotion (TOP) Guidelines that define the transparency and reproducibility guidelines including pre-registrations (Nosek et al., 2018). Additionally, incentives for pre-registrations for the publications are promoted. For instance, the Preregistration Challenge offers 1000 $1000 awards to researchers who publish the preregistered study results by the following steps: plan the study and analyses, submit the plan on OSF, write according to this plan, submit the manuscript for publication to eligible journals e.g. Academic Psychiartry and lastly submit the published article to confirm completion in this challenge (Nosek et al., 2018; COS, 2019). Registered Reports is another model that adopted by dozens of journals which the acceptance of paper is ensured based on the sufficiently important question and methodology of high quality as reviewed by peers before the observation of research outcomes (Nosek et al., 2018).

In terms of opportunity, there are large amount of domain-specific and domain-general registries for researchers from various disciplines to carry out pre-registrations. For instance, the Evidence in Governance and Politics (EGAP) Registry is responsible for researchers in the fields of economics and political science to submit their pre-registrations reports (Nosek et al., 2018).

According to Nosek & Lindsay (2018), the growth of pre-registration indicated with the rise of journals advising pre-registration by offering registered reports and badges as well as the total number of OSF registrations is skyrocketing. The amount of registrations is doubling each year and a dramatic growth is shown from 38 registrations in 2012 to 12090 registrations in 2017 within 5-year period.

Pre-prints

Pre-prints refer to the version of research paper prior to the peer review and journal publication (Balaji & Dhanamjaya, 2019). It is addressed by MDPI (2019) that the reasons of pre-prints include ensuring the research is visible as early as possible, getting feedback from peer researchers, avoiding to wait for peer review before the work is publicly available and making the early results citable, which benifits the authors in carrying out grant applications. Since the establishment of arXiv which is a preprint repository since 1991, various repositories for other disciplines have started to develop (see Table 7.2). Besides, the study of Balaji et al., 2019 also addressed the growth of preprints in the discipline of life science from 2007 to 2018 (see Figure 7.10). The arXiv q-bio from the arXiv has published pre-prints since September 2003, bioRxiv was launched in November 2013 and PeerJ Preprints covering medical, environmental and biological sciences was launched in April 2013 (Balaji et al., 2019).

Table 7.2.

Growth of preprint repositories, 1991-2018.

S. No.	Name of Preprints	Subject/Disciplines	Year Established	No. of Records as on 28 July 2018	Website
1	arXiv	Natural Sciences, Engineering, Economics, Finance and Computing	1991	1,421,596	https://arxiv.org
2	RePEc	Economics	1992	2,600,000	https://ideas.repec.org
3	SSRN	Social Sciences	1994	810,845	https://www.ssrn.com/en
4	E-LIS	Library and Information Science	2003	20,390	http://eprints.rclis.org
5	bioRxiv	Life Sciences	2013	25,632	http://www.biorxiv.org
6	PeerJ Preprints	Biological, Medical, Environmental and Computing Sciences	2013	4129	https://peerj.com/preprints
7	OSF Preprints	Natural Sciences, Technology, Engineering and Social Sciences. Arts and Humanities	2013	3170	https://osf.io/preprints
8	MDPI Preprints	Natural, Engineering, Social Sciences and Arts and Humanities	2016	5095	https://www.preprints.org
9	ChemRxiv	Chemical Sciences	2016	9910	http://www.chemrxiv.org
10	ESSOAr	Earth Sciences	2018	149	https://www.essoar.org

Noted. From Balaji & Dhanamjaya (2019).

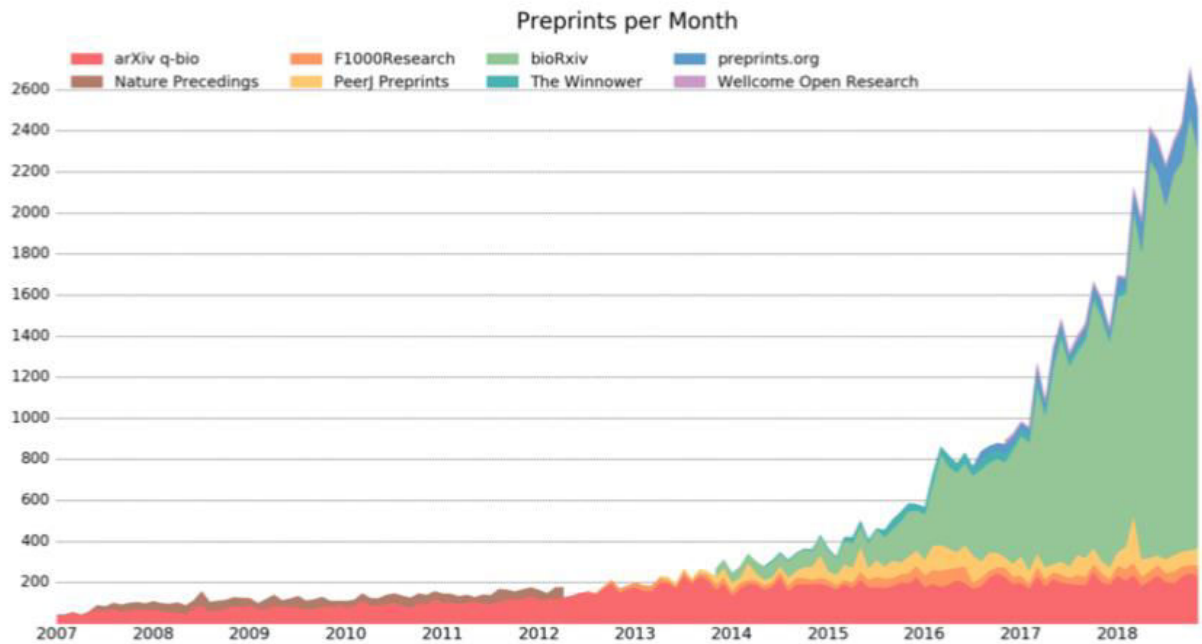


Figure 7.10. Monthly preprints added in November 2018 is 2509*.*

Noted. From Balaji & Dhanamjaya (2019).

The pre-print systems held on the Open Science Framework (OSF) which is developed by the Center for Open Science (COS) hosts 23 pre-print services . According to a research (Narock et al., 2019), 9 of these services are analysed based on the number of at least 100 manuscripts, English language medium, accessible through programmatic means and a domain/community focus (see Table 7.3). The distinct authors is a proxy for the frequency of usage for the services, i.e. a high percentage indicates more uptake of the pre-print services, from Table 7.3, LawArXiv has the lowest percentage (39%), while EarthArXiv has the highest percentage (71%).

Table 7.3.

Paper and Author counts for 9 COS preprint services.

Preprint System	Domain	Total Papers	Total Authors	Distinct Authors
PsyArXiv	Psychology	3534	12,439	7342
SocArXiv	Sociology	3034	5337	3017
LawArXiv	Law	905	1186	463
EarthArXiv	Earth and Planetary Science	567	2353	1659
EngrXiv	Engineering	362	961	664
MarXiv	Marine Science	324	960	627
LISSA	Library and Information Science	139	257	175
MindRxiv	Mind and Contemplative Practices	121	413	285
PaleorXiv	Paleontology	112	343	236

Noted. From Narock & Goldstein (2019).

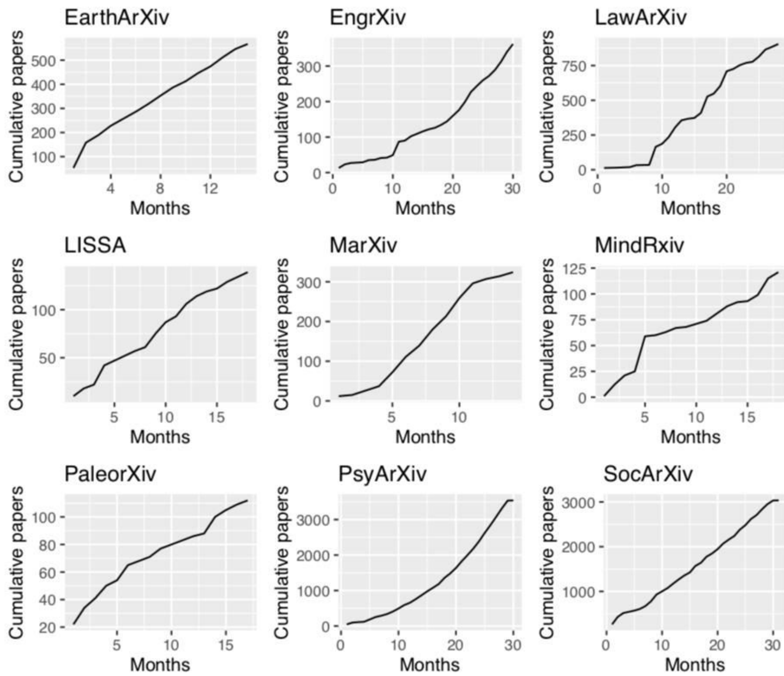


Figure 1. Cumulative paper submissions for the 9 services, data is recorded monthly. Note that values on both the X axis (Cumulative paper submission count) and Y axis (months since founding) are different for each panel. Each service will vary based on discipline specific paper production rates.

Figure 7.11. Cumulative paper submissions for the 9 services.
Noted. From Narock & Goldstein (2019).

According to Figure 7.11, the growth of the 9 services are growing in linear manner as illustrated by the plots of the cumulative manuscripts.

In addition, a scholar addressed that preprint servers are managed through four key approaches (see Figure 7.12) recently and it is noted that Peer J of publisher-supported preprints has closed down. Standalone preprint servers e.g. bioRxiv have developed own technical solutions and operate independently of other parties while other standalone preprint servers e.g. ChemRx-iv use third-party technology and technical infrastructure such as the Figshare infrastructure (Manista, 2019). Publisher-supported preprints refer to the posting of preprints as part of the publication workflow e.g. PeerJ while there are as well publishers posting preprints to preprint servers such as the PLOS posting preprints to bioRxiv (Manista, 2019).

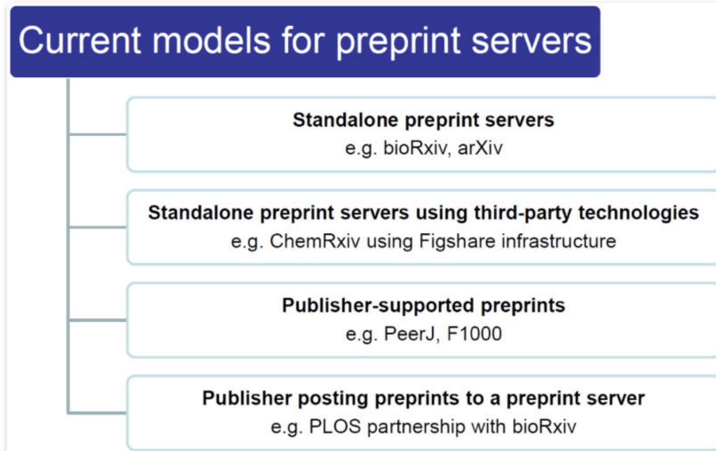


Figure 7.12. Models for preprint servers.
Noted. From Manista (2019).

Use of OSF

Center for Open Science (OSF) was found by Brian Nosek and Jeffrey Spie in 2013 as an open source software project to advocate openness, integrity, and reproducibility of scientific research. According to a crowdsourcing project by Nosek (2019), OSF adoption was taken as a proxy engagement of open-science which has grown in the past 6 years. Among disciplines, social-behavioral sciences, including psychology departments which accounted for 69%, were one of the dominant OSF users. Thus, OSF application was at the early majority phase based on Rogers' (1962) model of innovation diffusion (see Figure 7.13). Furthermore, social psychology was found already in late majority with almost complete OSF registration among assistant professors. Variations among countries were noted, with UK, Netherlands, and Australia outperforming US, Canada, and Germany (see Table 7.4).

OSF users have spun from 371 in 2013 to 6700 in 2014, and further doubled to 13,700 in 2016 and finally reaching 100,000 in 2018. Its user base has grown by 50,000 in only 4 years and another 4 months for the recent 50,000 users (see Figure 7.14; Pfeiffer, 2019).

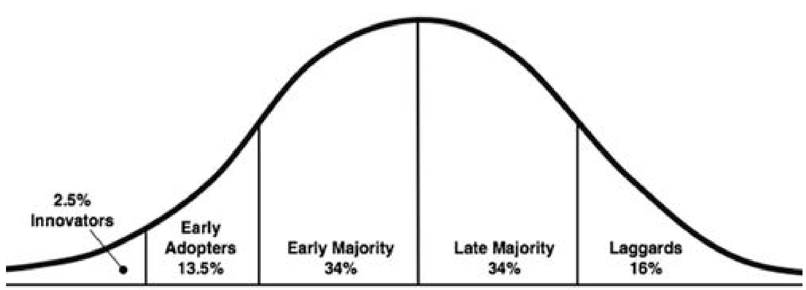


Figure 7.13. Rogers' (1962) classic diffusion of innovation model.
Noted. From Nosek (2019).

Table 7.4.
Adoption of OSF across nations.

Country	N	No	Yes
Norway	23	48%	52%
United Kingdom	143	52%	48%
Netherlands	210	55%	45%
Australia	170	57%	43%
France	28	64%	36%
United States	747	65%	35%
Canada	135	70%	30%
Germany	543	71%	29%
Serbia	72	82%	18%

Noted. From Nosek (2019).

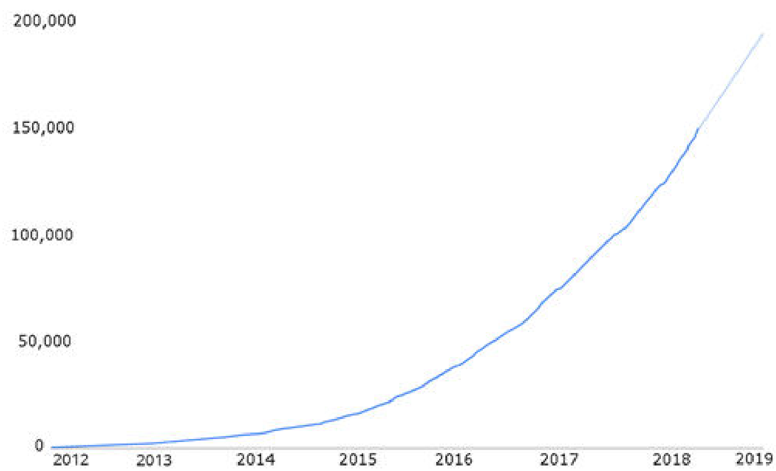


Figure 7.14. Recorded and prospective growth of OSF users from 2012 to 2019.
Noted. From Pfeiffer (2019).

Reasons for Growth of Open Science

We have understood the increasing trend of the usage of open science, and now we come to investigate the reasons behind.

Growth of the Internet

Talking about the reinforcement, the growth of the internet has definitely helped with the boost of openscience. The number of websites in the world increased to more than 1.8 billions in 2017 (Relojo, 2017). With the increase in websites, blogs as another form of website, social media networking sites, podcasting sites rise too. The open discussions on sciences are enabled and have been increasing quite rapidly at these sites since around the 2010s (American Psychological Association, 2012). For instance, the topics for scientific blogs are expanding and also widely ranged (See Table 7.5) (Nicolas, Bai & Fiske, 2018); so as scientific podcasts (See Figure 7.15) (MacKenzie, 2019). Social media provides frequent news update about scientific articles regarding the topic.

Table 7.5.
Topic proportions of scientific blogs.

	Mean	SD		Mean	SD
Statistics	0.21	0.17	**Clinical**	0.06	0.09
Bayesian	0.08	0.09	**Neuroscience**	0.05	0.05
General	0.04	0.04	**Social Science**	0.05	0.04
Software	0.04	0.04	Social Science	0.04	0.03
Regression	0.02	0.02	Demographics	0.01	0.01
Visualization	0.02	0.03	**Fraud**	0.03	0.02
Frequentist	0.01	0.02	**Teaching**	0.03	0.02
Replication	0.19	0.15	**Psychology: Other**	0.02	0.02
Science Communication	0.15	0.08	**Other Content**	0.01	0.01
Research Findings	0.12	0.05	Nutrition	0.01	0.01
Theoretical	0.07	0.05	Law	0.00	0.00
Experimental	0.05	0.05	**Unidentified**	0.08	0.07

Note. Topics and the mean and standard deviation of their share of words per post (at the blog level, averaged across each blog's posts). Non-bolded topics are subtopics of the bolded topics above them.

Noted. From Nicolas, Bai & Fiske (2018)

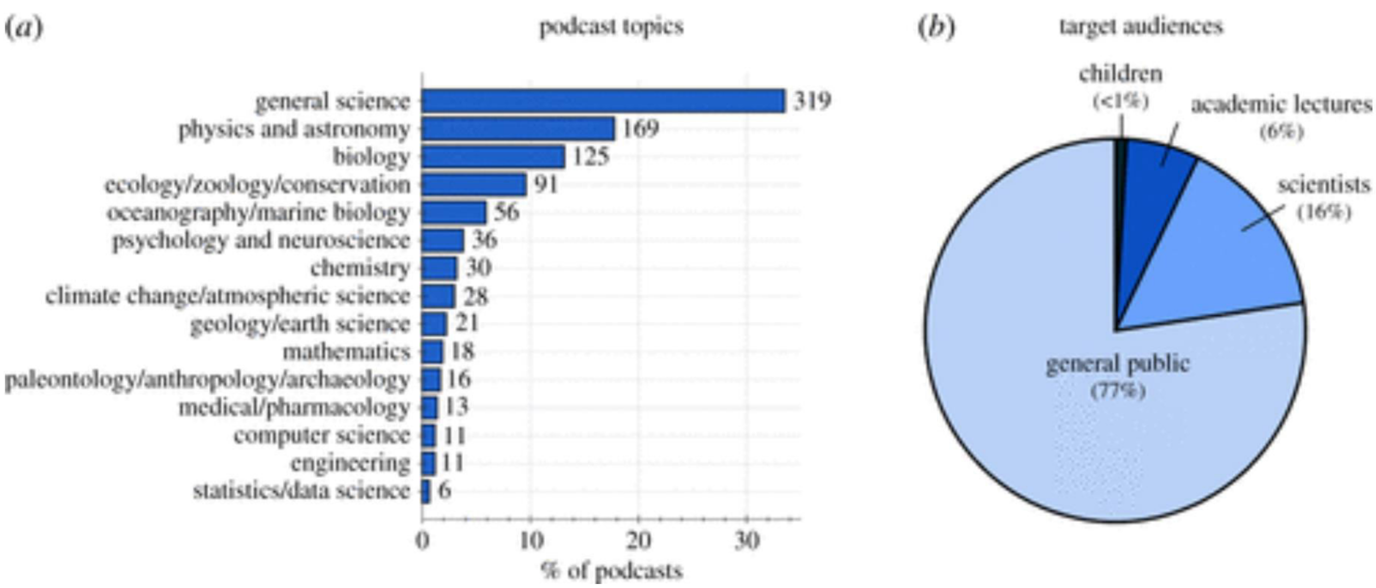


Figure 7.15. Coverage proportion of topics in scientific podcasts.
Noted. From MacKenzie (2019).

These websites with scientific contents are easier to access onced published online. For instance, blogs, podcasts on Youtube or social media sites, are mostly free of charge and easily found on search engine optimisation (i.e. Google) (Steenstrup-Duch, 2018). Blogs, social media posts and podcasts, which aim to let people have a glimpse and brief understanding of the scientific topic, they are easier, faster and livelier to understand than the very detailed and advanced phrases in the journal articles (Relojo, 2017; Feldman, 2016). Hence, these websites are good teaching tools to teach the general public who is not familiar with the scientific topic (Clay, 2008). What is more, it is easier to access crowdsourced science which saves much time for researchers or other people to collect. There are growing crowdsourcing websites for researchers to access public scientific information and data easily, such as Amazon's Mechanical Turk (for collecting large amount of behavioral data globally) and The Center for Open Science launched the Open Science Framework (provides sources of preprints and open public research materials online). Human interaction data can be got through Facebook and Twitter(Voytek, 2017). People would increase their visit to these websites with scientific contents with these cheaper-cost and low-effort incentives.

From the perspective of content creators, inputting scientific contents in blogs, social media and podcasts are ever easier than writing journal articles, as unlike the journal articles, they do not need formal structures nor reporting format, nor have to be reviewed rigorously, that authors can input anything relevant in a natural way to organise thoughts. In terms of feasibility, blogging and filming technology becomes more and more simple (Clay, 2008; MacKenzie, 2019). Furthermore, figure 7.16 also demonstrated to reach people in the community of that particular field, so to expand one's expertee network, the scientific content should be spread through corresponding channels and parties on the internet (Steenstrup-Duch, 2018). For example, blogs can increase the visibility of blogged research to the research business or community; and contents linked to social media can impact the general public. Thus this leads to the increase of number of uploading blogs, social media and podcasts so as to spread the scientific knowledge wider.

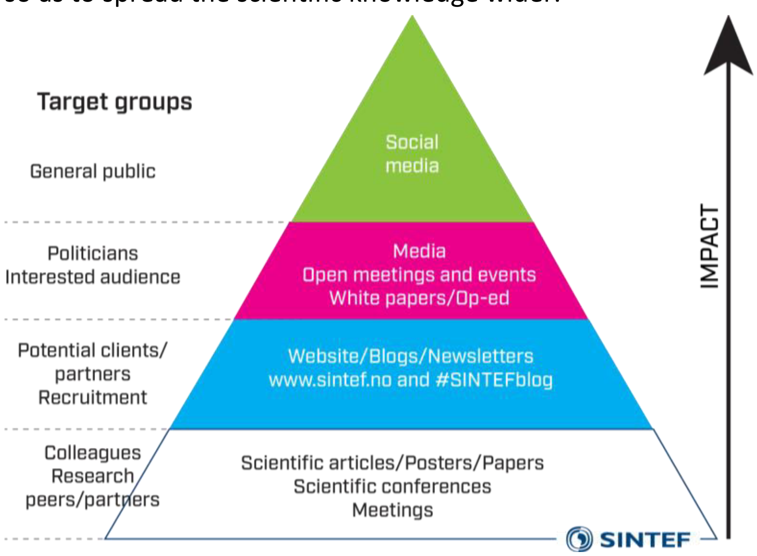


Figure 7.16. Channels for increased research impact.

Noted. From Steenstrup-Duch (2018).

With these reinforcements of the growth and delivery of the scientific content websites, it is more convenient to learn the science online and improve the science collectively on a platform called the "Internet". The accessibility of the scientific knowledge blasts at all levels, leading to the growth of open science.

Growth of Public Science

The emergence of public science plays a prominent role in boosting open science. Since scientists realised there is a low participation on scientific researches and discussions by the general public in 2017, there has started to increase science outreaching towards and science communication with the public (Leeming, 2017). For instance, there have been more and more universities (to name some: Utrecht University, University of Hong Kong, University of New England, Stanford University, Vanderbilt Univerity, etc.) and science institutes (e.g. International Centre for Theoretical Physics, Association for the Sciences of Limnology and Oceanography, etc.) doing outreach to secondary and primary school students. On the other hand, there are funds to reward scientists who successfully did the outreach, attaching them more of them to initiate the public science (National Science Foundation, 2015). These can explain the growth of public science.

Public outreach includes public presentations about scientific issues, public science workshops and big science events, that aim to promote science. In this way, the general public can have their awareness on the importance and the understanding of science raised, because the scientists actively reaching them (ASLO, 2019). As more people can get in touch with these sciences which are reached to them, this may increase the chance of attracting their interest and support to scientific researches. At the same time, since many people are watching over these public outreaches, scientists have to make sure their studies are high quality, such as being reproducible and credible, before they have these studies outreached to the public. The public science action gives incentive to the scientists to follow the open science practice to uphold the research quality and transparency, gaining the trust and support from the general public audience.

Conclusion

In this chapter, we have come across why we need to practice open science, such as to uphold the quality of the scientific work; and how open science grows, for example, using pre-registration, preprints, etc, with the reinforcement of the internet growth and public science growth. These all contribute to the drastic increase of the use and the understanding of open science, in favour of keeping the credibility and usability of science.

Quiz

Write 10 easy quiz 4 multiple choice questions to ensure reading and understanding of this chapter:

1. What is the best way to promote scientific knowledge to the general public?
   a. Writing scientific blogs
   b. Posting scientific news on social media
   c. Making scientific podcasts
   d. Announcing scientific news through media

2. Which of the following is true about Open Science Badges?
   a. Open Science Badges are awarded to articles to acknowledge open science practices such as open data, open materials, or preregistion
   b. Open Science Badges are used to rate which articles performed best in open science practice
   c. Open Science Badges are useless in boosting data sharing
   d. Open Science Badges are awarded by the Centre for Open Science

3. According to Rogers' (1962) model of innovation diffusion, which stage was OSF application at by 2019?
   a. Early adopters
   b. Early Majority
   c. Late Majority
   d. Laggards

4. Which types of open science practice is most frequently adopted by published authors from 1996 to 2017?
   a. Posting data or code online
   b. Posting study instruments online
   c. Pre-registering hypotheses or analysis
   d. not making anything public

5. Which of the following campaigns offer one thousand $1000 awards to researchers who publish the results of pre-registered study?
   a. the Transparency and Openness Promotion (TOP) Guidelines
   b. the Evidence in Governance and Politics (EGAP) Registry
   c. Preregistration Challenge
   d. Psychological Science (PSCI)

6. Which of the following is the first established preprint repository in 1991?
   a. arXiv
   b. bioRxiv

c. SSRN
d. PeerJ Preprints

7. What effect do badges have for Open Science after its implementation in 2015?
   a. Open data in Psychological Science (PSCI) doubled
   b. Open data in Psychological Science (PSCI) increased 10 times
   c. It does not have significant effect
   d. Open data in Psychological Science (PSCI) increased 15 times

8. Which of the following has become a powerful tool to collect a larger amount of research responses?
   a. Facebook
   b. SciStarter
   c. Mechanical Turk
   d. Netflix

9. Which of the following is NOT a predisposition that social media can promote science successfully?
   a. Free of charge
   b. Easy and interesting content
   c. Able to educate the public with meticulously reviewed scientific knowledge
   d. Able to enhance public involvement in scientific research

10. Which of the following is NOT an example of public science?
    a. Scientific research done by universities
    b. Science presentations
    c. Outreaching
    d. Science workshops

References

Aalbersberg, I. J., Appleyard, T., Brookhart, S., Carpenter, T., Clarke, M., Curry, S., … Vazire, S. (2018, February 15). Making Science Transparent By Default; Introducing the TOP Statement. https://doi.org/10.31219/osf.io/sm78t

Abraham, W. T., & Rusell, D. W. (2008). Statistical Power Analysis in Psychological Research. *Social and Personality Psychology Compass,* 2(1):283-301. DOI: 10.1111/j.1751-9004.2007.00052.x

Albright, A. (2017, July 26). A rising tide lifts all podcasts. *The Little Dataset*. Retrieved from https://thelittledataset.com/2017/07/26/a-rising-tide-lifts-all-podcasts/ .

American Psychological Association (2012). Psychology's growing library of podcasts. *American Psychological Association, 43*(1). Retrieved from https://www.apa.org/monitor/2012/01/podcasts

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-Size Planning for More Accurate Statistical Power: A Method Adjusting Sample Effect Sizes for Publication Bias and Uncertainty. Psychological Science, 28(11), 1547-1562. doi:10.1177/0956797617723724

ASLO. (2019). Science outreach in action. *Association for the Sciences of Limnology and Oceanography.* Retrieved from https://www.aslo.org/science-communication/science-outreach-in-action/ .

Bakker, M., Hartgerink, C. H., Wicherts, J. M., & Van der Maas, H. L. (2016). Researchers' Intuitions About Power in Psychological Research. Psychological Science, 27(8), 1069-1077. DOI:10.1177/0956797616647519

Balaji, B. P., & Dhanamjaya, M. (2019). Preprints in Scholarly Communication: Re-Imagining Metrics and Infrastructures. *Publications*, *7*(1), 6. DOI: 10.3390/publications7010006

Center for Open Science. (2019, November 8). Registered Reports: Peer review before results are known to align scientific values and practices. Retrieved from https://cos.io/rr/#RR.

Clay, R. (2008). Meet psychology's bloggers. *American Psychological Association, 39*(11), 34. Retrieved from https://www.apa.org/monitor/2008/12/bloggers

Cohen, J., & Wolman, B. B. (Ed.) (1965). Handbook of clinical psychology. *New York: McGraw-Hill.*

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. *New York: Academic Press.*

Cooper, H., & Findley, M. (1982). Expected effect sizes: Estimates for statistical power analysis in social psychology. *Personality and Social Psychology Bulletin*, 8, 168-173. DOI: 10.1177/014616728281026

COS. (2019). More About the Preregistration Challenge. Retrieved from https://cos.io/our-services/prereg-more-information/.

Dickersin, K., & Mayo-Wilson, E. (2018). Standards for design and measurement would make clinical research reproducible and usable. *Proceedings of the National Academy of Sciences*, *115*(11), 2590-2594. https://doi.org/10.1073/pnas.1708273114

European Commission. (2019, May 11). Data on open collaboration. Retrieved from https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/data-open-collaboration_en

European Commission. (2019, May 11). Facts and Figures for open research data. Retrieved from https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en

Federer, L.M., Belter, C.W., Joubert, D.J., Livinski, A., Lu, Y-L., Snyders, L.N., et al. (2018). Data sharing in PLOS ONE: An analysis of Data Availability Statements. PLoS ONE 13(5): e0194768. https://doi.org/10.1371/journal. pone.0194768.

Feldman, G. (n.d.) Open science and replication crisis. Retrieved October 24, 2019, from http://mgto.org/open-science/#crisispodcasts .

Guilford, J. P. (1956). Fundamental statistics in psychology and education. *New York: McGraw-Hill.*

Haase, R. F., Waechter, D. M., Solomon, G. S. (1982). How significant is a significant difference? Average effect size of research in Counseling Psychology. *Journal of Counseling Psychology,* 29, 58-65. https://doi.org/10.1037/0022-0167.29.1.58

Kidwell, M.C., Lazarević, L.B., Baranski,E., Hardwicke,T.E., Piechowski, S., Falkenberg, L.S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.M., Fiedler, S., Nosek, B.A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. https://doi.org/10.1371/journal.pbio.1002456

Kraemer, H. C., & Thiemann, S. (1987). How many subjects? Statistical power analysis in research. *Beverly Hills, CA: Sage.*

Kraemer, H. C., Gardner, C., Brooks, J. O. III, & Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychological Methods*, 3 (1), 23–31. https://doi.org/10.1037/1082-989X.3.1.23

Leeming, J. (2017, April 14). Increased communication and outreach efforts require changes in the structure and culture of academic science, says Nicole Forrester. *Naturejobs*. Retrieved from http://blogs.nature.com/naturejobs/2017/04/14/the-next-generation-of-science-outreach/ .

Mackenzie, L. E. (2019). Science podcasts: Analysis of global production output from 2004 to 2018. *R. Soc open sci,* 6. https://doi.org/10.1098/rsos.180932

Manista, F. (2019, May 22). *Are preprints paving the way to science in real time?* Retrieved from https://scholarlycommunications.jiscinvolve.org/wp/2019/05/22/are-preprints-paving-the-way-to-science-in-real-time/.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., Simonsohn, U., Wagenmakers, E., Ware, J. J. & Ioannidis, J. P. (2017). A manifesto for reproducible science. Nature human behaviour, 1(1), 0021. DOI:10.1038/s41562-016-0021

MPDI. (2019). Preprints' benefits to Authors. Retrieved from https://www.preprints.org/instructions_for_authors.

Narock, T., & Goldstein, E. B. (2019). Quantifying the growth of preprint services hosted by the Center for Open Science. *Publications*, *7*(2), 44. DOI: 10.3390/publications7020044

National Science Foundation. (2015). Perspectives on Broader Impacts. Retrieved from https://www.nsf.gov/od/oia/publications/Broader_Impacts.pdf .

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. *Annual Review of Psychology*, 69, 511-534. https://doi.org/10.1146/annurev-psych-122216-011836

Nicolas, G., Bai, X. & Fiske, S. T. (2018). Exploring research methods blogs in psychology: Who posts what about whom, with what effect. *Perspectives on Psychological Science*, 1-38. https://doi.org/10.1177/1745691619835216

Nosek, B. (2019). The Rise of Open Science in Psychology, A Preliminary Report. Retrieved from: https://cos.io/blog/rise-open-science-psychology-preliminary-report/ .

Nosek, B. A., & Lindsay, D. S. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, *31*(3).

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. Infancy, 22(4): 436–469. https://doi.org/10.1111/infa.12186

OECD. (2015). Making Open Science a Reality. *Organisation for Economic Co-operation and Development*, 18. DOI: 10.1787/5jrs2f963zs1-en

Open Science Badges. (2019). Retrieved 9 November 2019, from https://cos.io/our-services/open-science-badges/

Paluck, B.L., Birke, D., Christensen, G., Littman, R., Miguel, T., & Wang, Z. (2018). Open Science Practices are on the Rise Across Four Social Science Disciplines. Retrieved from: https://osf.io/kvbnh/

Pfeiffer, N. (2019). OSF: Origin, growth, and what's next. Retrieved from: https://cos.io/blog/osf-origin-growth-and-whats-next/

Relojo, D. (2017). Blog psychology: Insights, benefits, and research agenda on blogs as a dynamic medium to promote the discipline of psychology and allied fields. *Psychreg Journal of PsychologyI*, 70-75*.*

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

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist,* 44 (10): 1276-1284. https://doi.org/10.1037/0003-066X.44.10.1276

Rowhani-Farid, A., Allen, M. & Barnett, A.G.(2017). What incentives increase data sharing in health and medical research? A systematic review. *Research Integrity and Peer Review 2* (4). Retrieved from : doi:10.1186/s41073-017-0028-9.

Sedlmeier, P.,& Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? Psychological Bulletin, 105(2): 309-316. https://doi.org/10.1037/0033-2909.105.2.309

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science, 22(11), 1359-1366. Retrieved from https://journals.sagepub.com/doi/10.1177/0956797611417632

Steenstrup-Duch, A. (2018, November 23). Why I <3 science blogs. *Linkedin.* Retrieved from https://www.linkedin.com/pulse/why-i-3-science-blogs-anne-steenstrup-duch/ .

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., … Shrout, P. E. (2017). It's Time to Broaden the Replicability Conversation: Thoughts for and From Clinical Psychological Science. Perspectives on Psychological Science, 12(5), 742–756. DOI: 10.1177/1745691617690042

TOP Guidelines. (2019). Retrieved 11 December 2019, from https://cos.io/top/

Voytek, B. (2017). Social Media, Open Science, and Data Science Are Inextricably Linked. Neuron, 96(6), 1219-1222. DOI:10.1016/j.neuron.2017.11.015

8 - Changes in stats/science implementation

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Man Fung Morris	CHAN	https://www.researchgate.net/profile/Morris_Chan2	https://osf.io/chmj7/	morris69@hku.hk	morrismanfung@gmail.com	3035573068
Hoi Ching	CHAN	https://www.researchgate.net/profile/Chan_Hoi_Ching3	https://osf.io/u3q6r/	pchc31@hku.hk	phoebechan2000@gmail.com	3035574593
Tsz Huen	CHENG	https://www.researchgate.net/profile/Tszhuen_Cheng	https://osf.io/kefbx/	u3556639@hku.hk	tracyy.cheng.tc@gmail.com	3035566390
Ka Ki	LAU	https://www.researchgate.net/profile/Ka_Ki_Lau3	https://osf.io/jxbzn/	u3557784@hku.hk	kakilau.19990716@gmail.com	3035577844
Sze Man	LO	https://www.researchgate.net/profile/Sze_Man_Lo2	https://osf.io/4dnha/	u3556912@hku.hk	loszeman123@gmail.com	3035569122
Ming Tung	WU	https://www.researchgate.net/profile/Ming_Tung_Wu2	https://osf.io/yq75d/	u3555134@hku.hk	christywu728@gmail.com	3035551345
Myrte	DRUYVESTEYN	https://www.researchgate.net/profile/Myrte_Druyvesteyn	https://osf.io/hfqse/	mldruyv@hku.hk	m.l.druyvesteyn@gmail.com	n.a

Managerial summary

The origin of statistics theory can be traced back to the 18th-century advances in probability. Throughout the years, the world of statistics and science has been rapidly changed. From Francis Galton and Karl Pearson turning statistics into a stringent mathematical discipline used for analysis to the development of better design of experiments models, hypothesis testing and techniques for use with small data samples initiated by William Sealy Gosset. Today, the revolution of statistics and science is still ongoing and researchers start to concern about the credibility crisis. An increasing number of published studies in fields have been found to have credibility problems as time goes by, such as the published result failed to be reproduced by researchers. The following session would discuss the changes in statistics in terms of the emergence of “new statistics”, statistics tools, stats, websites and services, changes in process, and meta-analyses.

In depth report

New statistics and Bayesian statistics

New statistics is the statistical practice with the estimated effect size, confidence interval and meta-analysis in addition to merely *p* value in research studies (Cumming, 2014). The idea is not new while it had been mentioned in the 80's (Gardner & Altman, 1986). Before knowing the emerging of new statistics, one has to know about the demerits of the traditional null hypothesis significance testing (NHST).

Null hypothesis significance testing is an approach, proposed by Sir Ronald Fisher, to make inferences about the population from a sample, by comparing a *p* value with the statistical criterion (McGrath, 2011). So, what is a *p* value? A *p* value is a conditional probability that an extreme test statistic is encountered, assuming the null hypothesis is true (van Zyl, 2018). Several misconceptions were suggested, including that *p* value is the probability that an observed result is because of sampling error, that the null hypothesis is true or that the alternative hypothesis is true with the observed data (Kline, 2004).

Wagenmaker et al. (2017) wrote about a logical flaw of null hypothesis testing. Consider an argument known us *modus tollens*:

If P, then Q.

Not Q.

∴ Not P.

Fitting the logic flow of NHST would be like:

If the null hypothesis is true, then the *p* value is not likely to be extremely small.

The *p* value is extremely small.

∴ The null hypothesis is false.

While the flaw in the logic flow of NHST is not obvious in the above case, let's rewrite it with another example, just by changing the event but not the logic:

If Jonathan is a man, he is not likely to be a millionaire.

Jonathan is a millionaire.

∴ Jonathan is not a man.

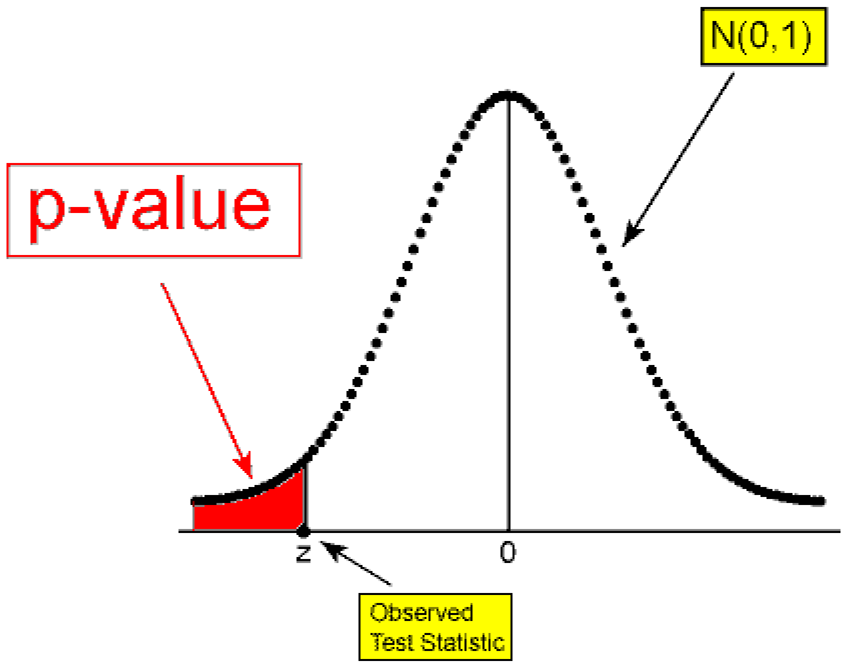


Figure 1. Normal distribution curve and *p* value.

It is clear that the logic flow here is faulty, while it involved the idea of likelihood. While putting this problem aside, Dian (as cited in van Zyl, 2018) also questioned the use of *p* to make inferences. *p* is the probability that certain data is observed given that the null hypothesis is true, $P(D|H_0)$, while D represents the data observed, H_0 represents that the null hypothesis is true, and | means "given" in conditional probability. While having a study, not limited to a psychological one, a researcher is interested in the likelihood of the null hypothesis being true with the given data, $P(H_0|D)$. Is $P(D|H_0)$ equal to $P(H_0|D)$? Readers having prior statistical knowledge would know the answer is NO, but still, an analogy is provided below:

P(being dead | being decapitated) is 100%, while none can survive without the head (and the brain). However, P(being decapitated | being dead) is extremely small, that many die because of other reasons while keeping their head with their body.

From the example above, it can be deduced that $P(D|H_0)$ is not equal to $P(H_0|D)$. While $P(H_0|D)$ is what scholars would like to know but $P(D|H_0)$ is used to make inferences on the hypothesis, there is a fundamental problem in NHST.

While doing statistical inference, we are only allowed to reject the null hypothesis when the *p* value is smaller than the statistical criterion or we fail to reject the null hypothesis. No matter how large the *p* value is, it is not sufficient to provide support to the null hypothesis (Wagenmakers, 2007). The *p* value is described as asymmetric as it can only be used to reject the null hypothesis and support the alternative hypothesis but not vice versa (Dienes, 2016).

The biggest problem of *p* value in causing the credibility crisis is that researchers misuse and have misconceptions about *p* value and . Goodman (2008) wrote about 12 misconceptions of the *p* value which will lead to errors while interpreting the results. *p* value is also susceptible to optional stopping, multiple testing or post-hoc analysis by researchers (Dienes, 2016). Ioannidis (2005) showed that many of the publications are false due to the heavy reliance on NHST. The problems of NHST and *p* value have led to the emergence of new statistics and Bayesian statistics.

Effect size is the amount of anything of interest, which can be presented with a standardised unit (e.g. Cohen's *d*) or being unit free (*r*; Cumming and Fidler, 2009). While effect size allows comparison and shows the strength of an effect, Cumming (2014) emphasised that it has to be interpreted with contexts, especially when it is related to clinical interest.

Fidler and Loftus (2009) argues that confidence intervals (CI) should be reported in addition to the *p* value. It was shown that with the CI provided, readers were clearer about that failing to reject H_0 does not mean supporting it and the uncertainty of the results were easier to be understood with the CI provided. While the *p* value itself does not give the strength of an effect (which the opposite is always a misconception; Goodman, 2008), the CI can visually provide a better idea of the certainty of the conclusion (Cumming, 2014).

Meta-analysis provides cumulative information, which is the extension of the quantitative information by effect size and confidence interval (Cumming, 2014). With a Forest plot, confidence intervals of effect sizes are shown, and readers and the academia can get a bigger picture of an effect. The idea of meta-analytic thinking is emphasised (Cumming & Finch, 2001), that readers should not focus on the result of one empirical study but also review the entire literature holistically. This kind of thinking is a new way of treating data that can reduce the faulty belief in academia by replication. Mega-analysis is a way of doing analysis by integrating findings in the discipline, pooling all the relevant data to achieve new knowledge and conclusion (Smith, 1982). It requires open science and coordination among scholars. (Bekkers, 2016). Increasing the transparency of data and more effective collaboration of researchers would help with the credibility crisis and boost development in academia.

Despite the adding of estimated effect size and confidence interval, the problem of traditional NHST still exist. Bayesian inference is introduced to the psychological field, trying to solve the problem of NHST.

In Bayesian statistics, a Bayesian factor is calculated from the prior priority of an event and the observed likelihood of the event. It uses the idea of conditional probability, calculating the posterior probability from the prior probability and the observed likelihood. It makes use of the previous research result and its advantages were discussed below. With NHST, one can make a dichotomous conclusion, whether rejecting H_0 or not. However, Bayseian statistics provides information more than merely rejecting the null. Comparing to an all-or-nothing dichotomous decision made in NHST, a continuous posterior distribution from Bayesian inference can give a full information about the data and the certainty in making conclusion (Kruschke & Liddell, 2017). A *p* value can only make a two-way distinction while a Bayesian factor can make a three-way distinction, having the power to support H_0.

One of the problems of *p* value is that it cannot support the null. Bayesian statistics involves the calculation of the probability of a model, including whether the null hypothesis or an alternative hypothesis has a higher chance to be true with the given data. Note that the probability of the hypotheses being true with the given data, $P(H_0|D)$, which is what researchers should be interested in is calculated. While the posterior probability function shows a high chance of H_0 to be true, one can conclude that there is no effect. Bayesian is said to make three-way distinction as there would be three possible results, H_0 is supported, H_1 is supported, or failing to make a conclusion (Diene, 2016).

Dienes (2016) also showed that Bayesian inference is immune to researchers' intention to perform optional stopping or post-hoc analysis. While misuse of NHST caused the credibility crisis, changing the way of doing statistics without changing the way researchers doing research may help the academia from faulty articles.

However, the adoption of Bayesian is costly. While the available programmes are free, it still takes researchers time and effort to learn and accept this new mode of doing statistics. While researchers are already busy with their projects, it is doubted if they are willing to spend extra time to adopt the method, while using Bayesian statistics may not lead to great improvement in their career and NHST could provide similar appearing results.

Normile, Bloesch, Davoli and Scherr (2019) suggested ways to introduce new statistics to the classroom, teaching the future researchers to adopt a more accurate method in doing statistics. These include giving new statistics instruction in additional to NHST and the use of new-statistic-friendly software. Both new statistics and Bayesian inference are meant to improve the clarity and the accuracy of publications by changing the statistical aspect of research. Still, there is a long way to go while traditional null hypothesis significance testing has been used for a long period and the cost of adopting new yet better methods is not small.

Tools

In light of the prevalence of statistical reporting error, a multitude of tools has been introduced to check the accuracy and consistency of statistical test. It serves two functions: 1. Allow researchers to check their work before publishing. 2. To evaluate the published research for fraud, malpractice and data manipulation. GRIM and SPRITE are two simple tools that uses descriptive statistics to identify errors and investigate properties of research datasets (Brown & Heathers, 2016; Heathers, Anaya & Brown, 2018).

Granularity-Related Inconsistency of Means (GRIM) is a method of evaluating the accuracy of published research, which only a mean and a sample size are needed (Brown & Heathers, 2016). To start with, GRIM is a test that particularly works for data with strong granularity, which makes it a useful tool in the field of social sciences. It evaluates whether the reported means of integer data such as Likert-type scales are consistent with the given sample size and number of items. For example, the reported mean age of 20.95 would be impossible for a sample size of 12, as the smallest amount that the mean can change by in this situation is one-twelfth. The technique was tested with 260 recent journal articles, and around half of them have at least one reported mean inconsistent with the reported sample sizes and scale characteristics (Brown & Heathers, 2016). It is however, worth-noting that a great degree of errors is due to unintentional mistake, like typo.

Sample Parameter Reconstruction via Interactive Techniques (SPRITE) is a heuristic method for reconstructing plausible samples from descriptive statistics of granular data, allowing future researchers to gain insights into the possible distribution of item values in the original data set (Heathers, Anaya, & Brown, 2018). In a lot of cases, researchers are reluctant to share the original dataset due to different reasons, like lack of institutional approval or sensitive data. Nonetheless, checking the data is of paramount importance, especially when manuscript contains inconsistencies and even some substantial error that may affect the result fundamentally. SPRITE is a flexible tool that let readers gain more insights to the research by reconstructing potential discrete datasets, with the use of some basic information like the mean, the standard deviation (SD), the sample size, and the lower and upper bounds of the range of item values. SPRITE starts by generating a random sample of item values with correct mean an arbitrary SD. Randomly selected pair of values are then adjusted, with one being increased and the other being decreased, until the SD reaches the target value with the mean untouched. The results will then be plotted in the form of bar charts which can give a simple visual impression of the possible distribution. By using SPRITE, James and his colleagues (2018) revealed anomalies in a multitude of studies, for example a study about carrot intake of elementary school students turned out to have an unrealistically large maximum value, with at least 41 carrots at one sitting (Wansink, Just, Payne, & Klinger, 2012).

R package ‘Statcheck’, introduced by Nuijten and her colleagues (2016), is a programme that can automatically extract statistics from articles and recompute p values. The reason they find such package necessary is that the error rate of inconsistent p-value of published articles are exceptionally high (Nuijten, Hartgerink, Van Assen, Epskamp, & Wicherts, 2015).

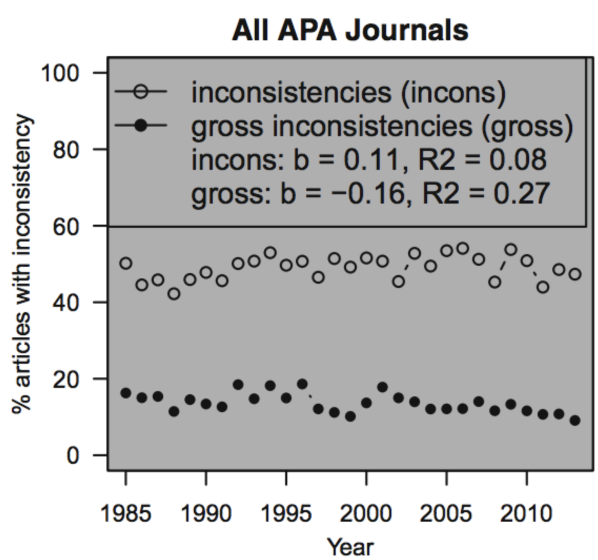


Figure 2. Percentage of articles with at least one inconsistency (open circles) or at least one gross inconsistency (solid circles)

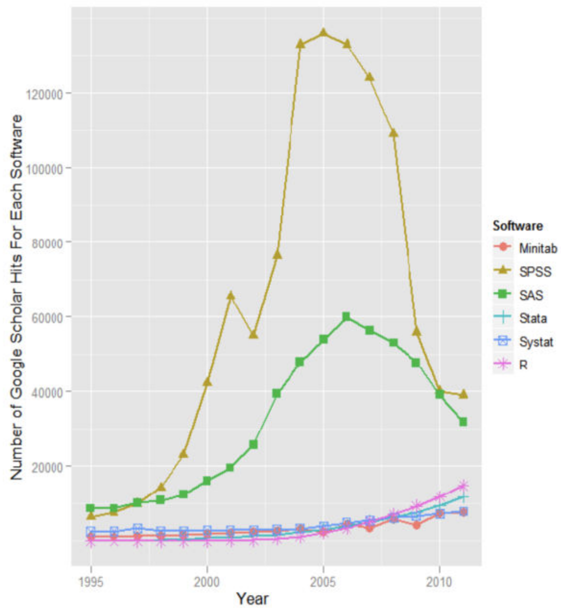
These studies showed that nearly half of all published empirical psychology consisted at least one inconsistent p-value. One in eight papers contained a gross inconsistency in which that may have affected the statistical conclusion. They also found out that prevalence of gross inconsistency is higher in p-values reported as significant than p-value reported as insignificant, which implies a systematic bias. These inconsistent results may lead to wrong substantive conclusions and affect meta-analyses. The R package 'statcheck' is created to recompute p-value in a much more effective and less error-prone way.

Stats

JASP/JAMOVI, the decline of SPSS/SAS

Jamovi is a free and open software which offers graphical user interface (GUI), simplifying some aspects in the R program. It also combines functions from other programming software such as SPSS (Statistical Packages for Social Sciences) and SAS (Statistical Analysis System). SPSS is an interactive statistical analysis including descriptive statistics, bivariate statistics, prediction for numerical outcomes and identifying groups, Geospatial analysis and R extension and python. In Jamovi, a bundle of statistical tests can be carried out, including t-tests, non-parametric tests, ANOVAs, and more. In addition, there is an R syntax, which users can make R syntax and copy to the R program. Data can eventually be run directly in R.

According to a comprehensive analysis by Robert A. Muenchen (2019) and some updated analysis by Lindeløv (2019), looking at everything about academic citations as well as the google trends, there is a sharp decline in SPSS in both google trends and scholar citations, while R seems to have an increasing trend, and is predicted to overtake SPSS by 2020 (Figure 2 & 3).



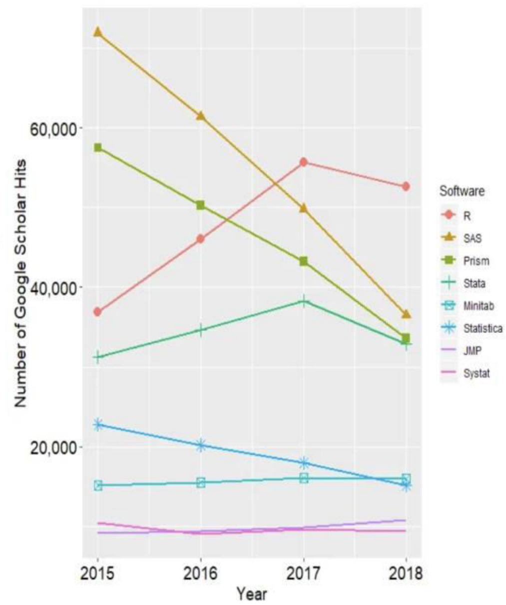


Figure 3 (left). Graph showing the number of Google Scholar citations for each software per year from 1995 to 2010.

Figure 4 (right). Graph showing the number of Google Scholar citations for each software per year from 2015 to 2018.

SPSS just simply feels old and unmaintained (Lindeløv, 2019). Some limitations of the SPSS that lead to a decline might include the incapability of giving a Cohen's d effect size when performing a t-test, which makes researchers unable to obtain how strong the relationship is between the two variables. Moreover, confidence interval on correlation coefficients as well as meta-analysis could not be carried out. SPSS also lacks the ability for reproducibility and flexibility. As R and Jamovi both take advantage of open source characteristic, users' code can be shared as well as to add or improve the shared items as a community (Hornick, 2018). In addition, SPSS is an expensive software which costs £79.13 per user per month. These all factors limit open science and in turn lead to a decline of users in SPSS and a surging of users in R (Lindeløv, 2019).

R/R studio revolution

R is a programming language for statistical computing and graphics for data manipulations, calculations as well as graphical display, which first appeared in 1993 and was designed by Ross Ihaka and Robert Gentleman. The reason it is called "R" is due to the fact that the first names of the two authors are started with "R".

R is usually labeled as an "environment" which provides a fully planned and coherent system for data analysis, as well as a place to implement a great deal of classical and modern statistical techniques. For example, the linear and nonlinear modeling, time-series analysis, and clustering, etc. Analyses done by using R are also reproducible. In fact, R is a modern implementation of the S (Scheme) programming language, which is also a statistical programming language developed by John Chambers, Allan Wilks, and Rick Becker in 1976. They differ fundamentally in their ability to maintain state information within functions (Ihaka & Gentleman, 1996).

R program is a highly collaborative program. While generating your own analysis, documentation, code, commentary, and metadata could be shared through an R Markdown file, which reduces the time with others for analysis. R Markdown therefore provides people a much easier for reproducible analysis.

R packages: ggstatsplot

R packages are a bundle of R functions that are stored under “library” in the R environment, and each package may include codes. Default standard set of packages can be found in R, yet further packages can be downloaded and installed later on. Details about each package can be found in the “DESCRIPTION file”, along with the information of the author, version, and the package dependencies, etc.

As we all know, statistics is fundamental in psychology. Finding all statistical packages in R could make it easier for data analysis. What we need to do is to input “help(package=stats)”, and all functions related to statistics will be shown in the R environment.

One of the packages is called “Ggstatsplot”, which is a package for data visualization using Grammar of Graphics created by Hadley Wickham. Users can generate graphs from the same components, including a data set, a coordinate system, and geoms (geometric objects). Geoms allow users to represent data points, in which they can use geoms to display data in the way they want to. Some examples of geoms include jitter, segment, boxplot, and histogram, etc. By creating graphics using statistical tests, data can be faster and more easily to be analyzed. For instance, to compare data between groups and within groups, we can input “ggbetweenstats” and “ggwithinstats” respectively to obtain violin plots. Other than violin plots, pie charts, scatterplots, correlation matrices, histograms, and dot-and-whisker plots could also be generated.

Shiny apps: p-checker / stimulations

According to the official page of R studio, Shiny is an open-source R package providing an elegant and strong web framework for developing web applications, by using statistical power of R. It also provides an outstanding and effective alternative on spreadsheets, reducing the time needed for statistical analyses.

A *p-value* is a conditional probability that an extreme test statistic is encountered, assuming the null hypothesis is true (van Zyl, 2018). An analyzer named *P-checker* allows us to check whether the *p-value* is correctly reported.

Websites and services

In light of the credibility crisis, a few changes have been made on how we do science. The following websites and services are more widely used when implementing scientific research, encouraging openness, transparency and mass collaborations between researchers.

As mentioned in previous chapters, the Centre for Open Science (COS) developed the Open Science Framework (OSF), an online platform which facilitates "open, centralized workflows" by capturing various stages of research (Foster & Deardorff, 2017, p. 203). It provides different services that aid the whole process (Foster & Deardorff, 2017).

One of the features of the OSF is that it promotes openness. The OSF will assign a particular uniform resource locator (URL) to each researcher, project and file to encourage information sharing. To help make the projects available to the public, the OSF can set archival resource keys (ARKs) and the digital object identifiers (DOIs) to different projects. Moreover, there is a version control which can keep track of the changes done on the files stored on the OSF (Foster & Deardorff, 2017).

Furthermore, it facilitates mass collaborations. After creating a workspace for a research project using an OSF account, it is easy for the users to add other people who are without accounts of the OSF beforehand to the projects. Those people will then be given links to register for their OSF accounts later (Foster & Deardorff, 2017). This provides convenience to the users and makes the projects collaborative.

To increase the transparency of the projects, the OSF allows registration of all kinds of projects. This implies that all projects can be created with timestamps which people cannot edit. Also, there will be records of any removal of content in a registered research project (Foster & Deardorff, 2017).

Another useful tool for doing preregistrations is the website AsPredicted. Preregistering a project means that researchers have to specify the "research design, hypotheses and analysis plan" before observing the data collected in research (Nosek & Lindsay, 2018). As the preregistrations done by AsPredicted can be private permanently unless the researchers choose to make them public, using AsPredicted is not a formal way to do preregistrations and it becomes difficult for the projects to be discovered by others. Nevertheless, it can safeguard the privacy of its users (Nosek & Lindsay, 2018).

When compared to the OSF, AsPredicted provides a kind of preregistration which is relatively easy and efficient as it only requires the researchers to answer 8 questions concerning their research project. On the other hand, the OSF provides various forms of preregistrations with varying level of requirements on the details of the study that the researchers have to provide (Veldkamp, Bakker, van Assen, Crompvoets, Ong, Soderberg, Mellor, Nosek, & Wicherts, 2017).

PsychFileDrawer.com is another website designed to deal with the replication crisis and encourage open science. People using this website can post the outcomes of their replication attempts concerning any psychological research studies onto the website. The replication attempts of particular research studies will be stored in an archive, no matter they are failed or successful (Diener & Biswas-Diener, 2019).

When researching in psychology, resources such as materials and manpower may not be circulated and used efficiently, slowing down the progress of the science community. To improve the efficacy of utilizing resources for psychological research, StudySwap is a tool which promotes crowdsourcing by enabling users to cooperate with others (McCarthy & Chartier, 2017). It allows researchers to tell others the resources they have and the resources they need respectively, as well as cooperating with other researchers in different projects. Then the resources of the whole psychological research community will be used more efficiently (Chartier, Riegelman, & McCarthy, 2018).

Besides, more and more researchers upload their preprints onto the preprint servers, disseminating prepublication rapidly. One of the most famous examples of preprint servers is arXiv. Nearly all papers in the field of high energy physics are included in this platform nowadays (Warr, 2003). Indeed, preprint servers have been created for various disciplines. For example, bioRxiv was developed for the biological sciences (Inglis & Sever, 2016). Social Science Research Network (SSRN) was designed for the social sciences, economics and law (Van Noorden, 2016).

Changes in process

As the advancement of technology and convenience brought to communication to people around the world, the way of researchers in doing science has changed. It has become more collaborative in different aspects of the research process.

Firstly, it is about the trend of sharing of data. Public sharing of primary data is not common even with the great potential in making scientific progress (Houtkoop, Chambers, Macleod, Bishop, Nichols, & Wagenmakers, 2018). Barriers deter researchers to share their data publicly is mainly attributed to fears and three non-fear considerations including: i) Popularity, which sharing data is not a common practice in their field. ii) Controllability, which researchers want to control whom they will distribute the data to. iii) Time-consuming, researchers consider that preparing data and learning how to share data online are very time-consuming (Houtkoop et al., 2018).

In order to change the phenomenon, both journals, researchers and funders initiated different campaigns for encouraging data sharing. From initiatives of journals, academic journals such as Royal Society journal and PLOS have applied stringent sharing policies, authors have to make their data being publicly available in recognized repositories (Houtkoop et al., 2018). Another effective encouraging project from journals is the badges project initiated by Centre for Open Science which provides badges for researchers following different kind of open practices, it caused a dramatic increase in the data sharing rate (Houtkoop et al., 2018). Besides journals, researchers take effort in promoting data sharing by collaboration. For example, the Peer Reviewers' Openness initiative which researchers refuse to review manuscripts without meeting five requirements, the public availability of data is included as one of the requirements (Houtkoop et al., 2018). Researcher funders also take an active role in data sharing promotion, funders such as the National Science Foundation and National Institution of Health require the provision of data sharing plan from applicants (Houtkoop et al., 2018). Both of those stakeholders in science have taken effort to raise the importance of data sharing nowadays.

Another change in process is about the version control and preprint. Version control is mean of management in different versions of a document, naming and distinction of series of draft document are involved in the process (University of Leicester, n.d.). From the version control, we can see the new trend of research process is about collaboration, researchers coming from different regions can work on the same research project or make amendments based on the previous file by the sharing of data and code online. Thus, version control is necessary to help tracking authorship and changes. As preprint which means a research paper shared before peer review has become more popular (AJE Scholar, n.d.), it also shows more collaborative change in process as version control. Comparing with peer review which consist of only two or three researchers, preprint increases feedback received by author since many researchers can discover your work and thus give comment on it. It is beneficial to researchers to improve their work.

To conclude, the changes in process is mainly about the openness and collaboration. It has more collaborative work during process by connecting, doing research changes from relatively private work to a more open work with sharing data and code.

Meta-analyses

Resulting from researchers and scholars' effort over the decades, the amount of information generated in academic research increases continuously. Unfortunately, variances in findings between thousands of research literature are inevitable. In much of the research literature, the split of finding statistically significant relationships or not is roughly 50–50 (Harlow, Mulaik, & Steiger, 1997). Hence, it brings a difficulty to develop shared and cumulative theories and knowledge. Therefore, academia nowadays urges to seek a toll which can make sense of the bunch of data from independent studies that address similar questions. In order to utilize the vast number of accumulated fings, statistical method suggested by Glass (1976), Meta-analysis, were developed in the late 1970s (Hunter & Schmidt, 2004). Meta-analysis pools and analyzes published results quantitatively under specific criteria, for example, results from identical studies, to generate more reliable estimates (Flather, Farkouh, Pogue, & Yusuf, 1997). Hunter and Schmidt (2004) suggested that meta-analysis can correct the distorting effects of sampling error, measurement error, and other artefacts that produce the illusion of conflicting findings. The cumulativeness of research findings in psychology seems to be weaker than that in physical sciences. However, a meta-analysis showed that there is as much variability across studies in physical sciences as there is in psychology (Hedges, 1987). A well-designed meta-analysis can analyze differences in the results among studies to increase precision in estimating effects (Walker, Hernandez, & Kattan, 2008). Nevertheless, performing a meta-analysis and interpreting its results still face many critical issues, and thus meta-analyses might yield misleading information. Since a meta-analysis based on analyzing data from various research findings. Therefore, performing a meta-analysis might come across a particular bias, such as publication bias, search bias and selection bias. Publication bias referred to trails with favourable results are more likely to be published than those with inconclusive results (Flather et al., 1997). Therefore, we need to pay special attention to the particularly important issue when we conduct a meta-analysis

6.1 Different bias in identification and selection of studies

There are many reasons for the selective publication of studies, maybe the study aimed to favour new treatment, or the results were not aligned with a well-established one (Walker et al., 2008). Selectively publishing experimental result might sound inappropriate in the science world as one of the essential principles of science is to have an open and transparent attitude towards experiment results. However, Turner, Matthew, Linardatos, Tell and Rosenthal (2008) found that publication bias was prevalent in academia. They analyzed the publication situation of antidepressants studies and found that, based on studies registered with the US Food and Drug Administration, 97% of the positive studies were published while only 12% of the negatives ones. As for the search bias, faulty searched of the sample studies will reduce the validity of a meta-analysis. Since the meta-analysis aims at summarizing results from individual studies which shared the same research purpose, the similarity of the experimental design is the critical criteria for choosing the study samples. In order to ensure the validity, most recent meta-analyses include a set of keywords they used which facilitates readers to further interpret the result (Walker et al., 2008). Apart from using keywords for more definitive studies search, researchers have defined some studies scoring criteria chosen by consensus (De Luca et al., 2008). The following are some examples of criteria used to reduce selection bias:

- Objectives
- Populations studied
- Study design (eg, experimental vs observational)
- Sample size
- Treatment (eg, type and dosage)
- Criteria for selection of controls
- Outcomes measured
- Quality of the data
- Analysis and reporting of results
- Accounting and reporting of attrition rates
- Length of follow-up
- When the study was conducted.

6.2 Assessing publication bias in meta-analysis

Since even well-rounded literature searches might not wholly avoid publication bias, some techniques have been developed to investigate the possibility of publication bias. The most straightforward and joint strategy is the Funnel plot developed by Light and Pillemar (1984). It is a scatter plot of a measure of study size against a measure of effect size. If there is no bias, the plot will appear symmetrical funnel-shaped, since effect sizes should be evenly distributed around the underlying exact effect size with more variability in the smaller studies than in the more extensive studies owing to the more considerable influence of sampling error. If gaps in the lower extremities of the funnel are observed, causing the plot to appear asymmetrical, publication bias may be suspected (Light and Pillemar, 1984). Figure 4 has illustrated both circumstances of symmetric and asymmetric results of the funnel plot test conducted by Dentail, Douketis, Gianni, Lim and Crowther (2007).

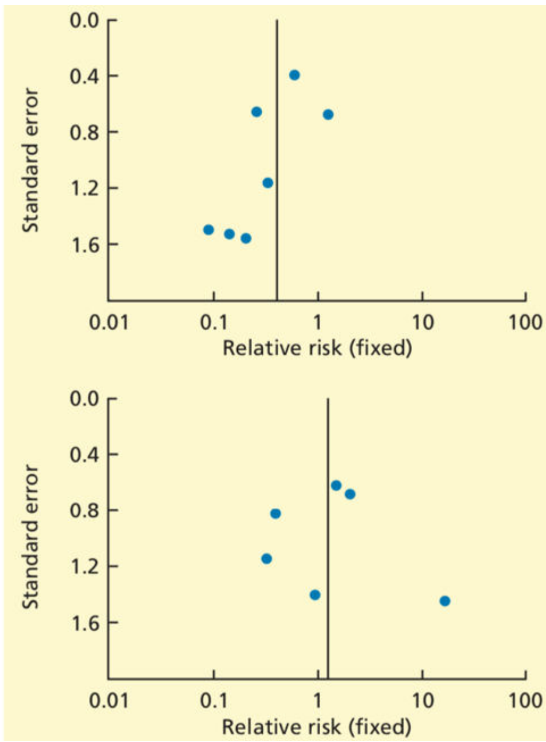


Figure 5. Dentail et al. (2007) conducted a funnel plot test of studies of anticoagulant prophylaxis. The funnel plot on the top showed an asymmetric result which suggested that the presence of selection bias. The funnel on the bottom showed a symmetric result, suggesting absence of selection bias.

An open debate: the p value war

The proposal of Benjamin et al. (2017) of changing the default alpha from 0.05 from 0.005, that a statistical significance is only claimed when $p < 0.005$. The suggestion of changing the threshold from 0.05 to 0.005 was based on two reasons, improved strength of p value and the false positive report probability. Benjamin et al. compare p value 0.05 with the relevant Bayes factor. It was found that the corresponding Bayes factor is only “weak” in Bayesian statistics. A study with $p < 0.005$ was estimated to have a Bayes factor which provides a “strong” evidence supporting the alternative hypothesis. Benjamin et al. also showed that reducing the threshold to 0.005 can greatly reduce the false positive report probability, by mathematical modelling with different prior odds. Benjamin et al. also suggested that the false positive report probability is reduced while the replication rate is higher among studies with $p <$ 0.005 (50%) comparing to those with $0.05 < p < 0.005$ (24%)

Amrhein and Greenland (2018), and McShane, Gal, Gelman, Robert and Tackett (2019) suggested removing and abandoning statistical significance totally while replying to the proposal by Benjamin et al. While it was aware that merely changing the threshold of p value would not improve the problem of using p value (as suggested in the section above), it was suggested that having 0.005 as the default alpha would lead to more serious issue such as increased focus on p value alone and oversimplified statistical reasoning (Amrhein & Greenland, 2018; Lakens et al., 2018). The proposal of having an alternated threshold seems to alleviate the problem of NHST but it does not. The illusion was believed to be exaggerating the existing problem. Amrhein and Greenland added that the above proposal would also lead to more intense p-hacking as the required p value would be harder to obtain. While advocating the removal of p value, McShane et al. emphasised the importance of subordinate factors, including but not limited to factors related to prior evidence, real world costs and benefits and other factors that vary by research domain. Emphasising the existing problem of NHST and p value, McShane et al. suggested authors and editors to consider more about these subordinate factors while having their analysis or evaluation of submitted papers. Journals have already started to ban the idea of NHST due to problems discussed before (Lindsay, 2015; Trafimow & Marks, 2015) and it was found that the quality of the papers without the emphasis on NHST have higher quality in terms of the validity of the conclusions (Fricker, Burke, Han, & Woodall, 2019).

Lakens et al. (2018) addressed the problem concerning replication. It was raised that it was not appropriate to attribute low replication rate to the original alpha level while the replicability is still low among studies with $p < 0.005$. Furthermore, having a lower default alpha would discourage scholars to do replication study as studies with high power required more resources and resources may rather be spent to original studies. Lakens et al. suggested the justification of alpha level, which researchers should transparently explain and justify their design, not only alpha level but also factors such as assumed prior odds or statistical power in registered report before data collection.

All of the ways in handling alpha level were suggested to improve statistical inference in academia, while they might have their own merits and demerits, including their effects and difficulties in application. Nevertheless, they do provide insight for us while thinking about how we should maintain research integrity and what the problem really is.

Change in the way we do science - move to crowdsourcing

Crowdsourcing science, as its name suggests, is the idea that numerous project members collaborate and carry out specific components of a larger project, usually under the direction of a core coordination team (Silberzahn et al., 2018). It is an alternative model of doing science such that the scale and impact of scientific research could be greatly expanded. With the maximization of material and human resources, ambitious projects that would be unattainable by individuals or small teams could be enabled. It is also a great way in assessing the robustness of findings as crowdsourced approach has an edge in determining reliability and generalizability of findings.

Crowdsourcing comes in a multitude of forms as it can vary greatly in terms of the degree of communication between project members and their inclusivity (Silberzahn et al., 2018). For example, citizen-science initiatives that include anyone willing to collect data involve a high degree of independence would be a case of inclusive projects, with low level of communication.

Crowdsourcing is commonly used in data-collection phase of a research project or for conducting replications. It, however, hold a much greater potential to be utilized in the entire scientific endeavor, like generating ideas to designing studies, analyzing the data, writing research reports, providing peer feedback and determining the direction of future analyzes. Data analysis is a process that has a great potential in crowdsourcing, as it is often seen as a mechanical, unimaginative process of revealing results from a research study (Uhlmann et al., 2018). Nonetheless, it is evident that result is significantly dependent on the chosen analytic strategy. A crowdsourced research about the correlation between skin tone and red cards in soccer revealed that, with the same data set and varied analytic approaches, 69% of the research teams found a statistically significant positive effect, while the remaining teams conclude otherwise. Here is their procedure of running a crowdsourced data analysis:

After the building the dataset and recruitment of researchers, the research teams initiated their first round of data analysis. Each team decided its own analytic approach and ran the analysis independently. They then submitted structured summaries of their analytic approach, including information about data transformations, exclusions and covariates. After individual analysis, a round-robin peer evaluation of overall analysis quality was conducted. The analytic approaches were presented in random order, and the analysts were instructed to provide detailed feedback as well as a confidence rating of each analytic approach on a 7-point scale. Each team also gained more insight from analyzing others' analytic approaches and from other's qualitative and quantitative feedback. The second round of data analysis ensued after the peer evaluation, and each team had the opportunity to change their analytic strategies and draw new conclusions. They were encouraged to present their results in the structure of a published article, with method and results sections. An open debate and further analysis was conducted after the second round of data analysis. As the analysts scrutinized each other's results, it was discovered that the differences in results is not merely due to the variations in statistical models, but also the choice of covariates. They concluded that the inclusion of two covariates, 'league' and 'club' may be the root cause of the nonsignificant results obtained by some teams. These teams were allowed to decide whether they would like to revise their model by excluding the covariates.

Crowdsourcing may be the scientific utopia, but is not without problems. Organizing a collective for a globally distributed project may create bureaucracy and induce high transaction cost. It is a much less cost-effective model of research as for the same effort, a much larger number of ideas with initial supporting evidence could have been conducted my individual smaller teams. Crowd projects may also create credit ambiguity and lack incentives for participation which would hamper the recruiting process.

Quiz

Write 10 easy quiz 4 multiple choice questions to ensure reading and understanding of this chapter:

1. Which factor is not the non-fear factor hindering researchers to share their data publicly?
    a. Sharing data is not a common practice in their field
    b. Researchers do not want to lose their control on whom they will distribute the data to
    c. Sharing data is too time-consuming
    d. Researchers want to use their data to earn money
2. Which stakeholder is not mentioned as an initiator in promoting public data sharing?
    a. Governments
    b. Researchers
    c. Funders
    d. Journals
3. What is/are the reason(s) of an asymmetrical funnel plot?
    a. Heterogeneity of studies
    b. Publication bias
    c. Chance
    d. All three above
4. What does *p* value reflect?
    a. Probability that an observed result is due to sampling error or a chance effect
    b. Probability that the null hypothesis is true based on the data
    c. Probability that the alternative hypothesis is true given the data
    d. Probability that a test statistic as extreme or more extreme than the one observed was encountered, while assuming the null hypothesis is true
5. Which of the following is not needed for running GRIM or SPRITE test?
    a. Mean
    b. Sample size
    c. Original dataset
    d. Standard deviation
6. Which of the following concerning Bayesian inference is true?
    a. It can provide evidence supporting the null hypothesis
    b. It is usually used to make a dichotomous decision
    c. It is used to make a two-way distinction decision
    d. It is susceptible to the application of stopping rule by the researchers
7. Which of the following is not the possible drawback of crowdsourcing?
    a. Bureaucracy
    b. High transaction cost
    c. Credit ambiguity
    d. Conflicting ideas

8. Which of the following statements concerning the website AsPredicted is true?

a. Preregistrations done by AsPredicted can be private permanently.
b. AsPredicted requires its users to provide a very detailed plan for their research when doing preregistrations.
c. AsPredicted is a formal way to do pre-registrations.
d. AsPredicted cannot safeguard the privacy of its users.

9. Which of the following preprint servers is not mentioned in this chapter?
   a. Social Science Research Network (SSRN)
   b. bioRxiv
   c. Research Papers in Economics (RePEc)
   d. arXiv
10. Why is SPSS experiencing a decline?
   a. It limits open science
   b. It does not work as an open, shared resource for improvement or amendment
   c. It takes a lot of time for people learning for the interface in SPSS
   d. All of the above

References

AJE Scholar. What are preprints, and how do they benefit authors? Retrieved from https://www2.le.ac.uk/services/research-data/organise-data/version-control

Amrhein, V., & Greenland, S. (2018). Remove, rather than redefine, statistical significance. *Nature Human Behaviour*, *2*(1), 4. https://doi.org/10.1038/s41562-017-0224-0

Bekkers, R. (2016). *Introducing Mega-analysis.* Retrieved 2019, December 15, from https://renebekkers.wordpress.com/2016/07/27/introducing-mega-analysis/

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., … Johnson, V. E. (2017). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10. https://doi.org/10.1038/s41562-017-0189-z

Brown, N. J. L., & Heathers, J. A. J. (2016). The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. http://doi.org/10.7287/peerj.preprints.2064v1

Chartier, C. R., Riegelman, A., & McCarthy, R. J. (2018). StudySwap: A platform for interlab replication, collaboration, and resource exchange. *Advances in Methods and Practices in Psychological Science*, *1*(4), 574-579. https://doi.org/10.1177/2515245918808767

Colling, L. J., & Szucs, D. (2018). Statistical reform and the replication crisis. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-018-0421-4

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Cumming, G., & Fidler, F. (2009). Confidence Intervals: Better Answers to Better Questions. *Zeitschrift Für Psychologie / Journal of Psychology*, *217*(1), 15–26. https://doi.org/10.1027/0044-3409.217.1.15

Cumming, G., & Finch, S. (2001). A Primer on the Understanding, Use, and Calculation of Confidence Intervals that are Based on Central and Noncentral Distributions. *Educational and Psychological Measurement*, *61*(4), 532–574. https://doi.org/10.1177/0013164401614002

De Luca, G., Suryapranata, Stone, Antoniucci, Biondi-Zoccai, Kastrati, . . . Marino. (2008). Coronary stenting versus balloon angioplasty for acute myocardial infarction: A meta-regression analysis of randomized trials. *International Journal of Cardiology*, *126*(1), 37-44. http://doi.org/10.1016/j.ijcard.2007.03.112

Dentali, F., Douketis, J., Gianni, M., Lim, W., & Crowther, M. (2007). Meta-analysis: Anticoagulant Prophylaxis to Prevent Symptomatic Venous Thromboembolism in Hospitalized Medical Patients. *Annals of Internal Medicine*, 278-288. http://doi.org/10.7326/0003-4819-146-4-200702200-00007

Diener, E. & Biswas-Diener, R. (2019). The replication crisis in psychology. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from https://nobaproject.com/modules/the-replication-crisis-in-psychology

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Fidler, F., & Loftus, G. R. (2009). Why Figures with Error Bars Should Replace *p* Values: Some Conceptual Arguments and Empirical Demonstrations. *Zeitschrift Für Psychologie / Journal of Psychology*, *217*(1), 27–37. https://doi.org/10.1027/0044-3409.217.1.27

Flather, M., Farkouh, M., Pogue, J., & Yusuf, S. (1997). Strengths and limitations of meta-analysis: Larger studies may be more reliable. *Controlled Clinical Trials*, *18*(6), 568-579. http://doi.org/10.1016/s0197-2456(97)00024-x

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association: JMLA*, *105*(2), 203-206. https://doi.org/10.5195/JMLA.2017.88

Fricker, R. D., Burke, K., Jr., Han, X., & Woodall, W. H. (2019). Assessing the statistical analyses used in Basic and Applied Social Psychology after their p-value ban. *Statistical Inference in the 21st Century, 73,* 374-384. http://doi.org/10.1080/00031305.2018.1537892

Gardner, M. J. & Altman, D. G. (1986). Confidence intervals rather than p values: Estimation rather than hypothesis testing. *British Medical Journal (Clinical Research Edition), 292*(6522), 746-750. http://doi.org/10.1136/bmj.292.6522.746

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5, 3–8. http://doi.org/10.2307/1174772

Goodman, S. (2008). A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, *45*(3), 135–140. https://doi.org/10.1053/j.seminhematol.2008.04.003

Harlow, L., Mulaik, S., & Steiger, J. (1997). *What if there were no significance tests*? (Multivariate applications). Mahwah, N.J.: Lawrence Erlbaum Associates.

Heathers, J. A., Anaya, J., Zee, T. V. D., & Brown, N. J. (2018). Recovering data from summary statistics: Sample Parameter Reconstruction via Iterative TEchniques (SPRITE). http://doi.org/10.7287/peerj.preprints.26968v

Hedges, L. V. (1987). How hard is hard science, how soft is soft science: The empirical cumulativeness of research. *American Psychologist*, 42, 443–455. http://doi.org/10.1037/0003-066X.42.5.443

Hornik, K. (2018). R FAQ. Retrieved from https://CRAN.R-project.org/doc/FAQ/R-FAQ.html

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70-85. http://doi.org/10.1177/2515245917751886

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage Publications.

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal Of Computational And Graphical Statistics*, *5*(3), 299-314. http://doi.org/10.1080/10618600.1996.10474713

Inglis, J. R., & Sever, R. (2016). *bioRxiv: a progress report.* ASAPbio.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), 6. http://doi.org/10.1371/journal.pmed.0020124

Kline, N. L. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington DC: American Psychological Association.

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., … Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, *2*(3), 168–171. https://doi.org/10.1038/s41562-018-0311-x

Light, R. J. and Pillemar, D. B. (1984) *Summing up: the Science of Reviewing Research*. Cambridge: Harvard University Press.

Lindeløv, J. (2019). SPSS is dying. It's time to change. [Blog]. Retrieved from https://lindeloev.net/spss-is-dying/

Lindsay, S. (2015). Replication in Psychological Science. *Psychological Science, 26*(12), 1827-1832. http://doi.org/10.1177/0956797615616374

McCarthy, R. J., & Chartier, C. R. (2017). Collections2: Using "crowdsourcing" within psychological research. *Collabra: Psychology*, *3*(1), 1-6. https://doi.org/10.1525/collabra.107

McGrath, R. E. (2011). *Quantitative models in psychology*. Washington, DC, US: American Psychological Association.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician*, *73*(sup1), 235–245. https://doi.org/10.1080/00031305.2018.1527253

Muenchen, R. (2019). The Popularity of Data Science Software [Blog]. Retrieved from http://r4stats.com/articles/popularity/

Normile, C. J., Bloesch, E. K., Davoli, C. C., & Scherr, K. C. (2019). Introducing the new statistics in the classroom. *Scholarship of Teaching and Learning in Psychology*, *5*(2), 162–168. https://doi.org/10.1037/stl0000141

Nosek, B. A., & Lindsay, D. S. (2018, February 28). Preregistration becoming the norm in psychological science. *Observer*. Retrieved from https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science/comment-page-1

Nuijten, M. B., Hartgerink, C. H. J., Van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2015). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*(4), 1205–1226. http://doi.org/10.3758/s13428-015-0664-2

Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37–64). Mahwah, NJ: Lawrence Erlbaum.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... & Carlsson, R. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337-356. http://doi.org/10.1177/2515245917747646

Smith, N. L. (1982). Evaluative Applications of Meta- and Mega-analysis. *American Journal of Evaluation, 3*(4), 43–47. http://doi.org/10.1177/109821408200300412

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology, 37,* 1-2. http://doi.org/10.1080/01973533.2015.1012991

Turner, E., Matthews, A., Linardatos, E., Tell, R., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine*, *358*(3), 252-260. http://doi.org/10.1056/NEJMsa065779

Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., ... & Nosek, B. A. (2018). Scientific utopia III: Crowdsourcing science. *Perspectives on Psychological Science*, *1*, 337–356. http://doi.org/10.1177/2515245917747646

University of Leicester. Version Control. Retrieved from https://www2.le.ac.uk/services/research-data/organise-data/version-control

Van Noorden, R. (2016, May 17). Social-sciences preprint server snapped up by publishing giant Elsevier. *Nature News*. Retrieved from https://www.nature.com/articles/nature.2016.19932#citeas

van Zyl, C. J. J. (2018). Frequentist and Bayesian inference: A conceptual primer. *New Ideas in Psychology*, *51*, 44–49. https://doi.org/10.1016/j.newideapsych.2018.06.004

Veldkamp, C. L. S., Bakker, M., van Assen, M. A. L. M., Crompvoets, E. A. V., Ong, H. H., Soderberg, C. K., Mellor, D., Nosek, B. A., & Wicherts, J. M. (2017). Restriction of opportunistic use of researcher degrees of freedom in pre-registrations on the Open Science Framework. *Preprint retrieved from https://psyarxiv. com/g8cjq.*

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Walker, E., Hernandez, A. V., & Kattan, M. W. (2008). Meta-analysis: Its strengths and limitations. *Cleveland Clinic Journal of Medicine*, *75*(6), 431–439. doi: 10.3949/ccjm.75.6.431

Wansink, B., Just, D. R., Payne, C. R., & Klinger, M. (2012). Attractive names sustain increased vegetable intake in schools. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2079831

Warr, W. A. (2003). Evaluation of an experimental chemistry preprint server. *Journal of Chemical Information and Computer Sciences*, *43*(2), 362-373. http://doi.org/10.1021/ci025627a

Whittington, C., & Pilling, S. (2013). The new statistics: A how-to guide: Guest editorial. *Australian Psychologist*, *48*(3), 171–171. https://doi.org/10.1111/ap.12017

Zyphur, M. J., & Oswald, F. L. (2013). Bayesian probability and statistics in management research: A new horizon. *Journal of Management*, *39*(1), 5–13. https://doi.org/10.1177/0149206312463183

9 - Changes in journals / publishing

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Chi Yuen	Lui	https://www.researchgate.net/profile/Chi_Yuen_Lui	osf.io/8nawr	u3547195@connect.hku.hk	eddielui1999@gmail.com	3035471959
Wing Hang	Lai	https://www.researchgate.net/profile/Wing_Hang_Lai2	osf.io/kh8ej	u3547241@connect.hku.hk	whenrylai@gmail.com	3035472410
Wing Ching	Chan	https://www.researchgate.net/profile/Wing_Chan25	osf.io/tsd4k	u3571437@connect.hku.hk	ashleychan062599@gmail.com	3035714375
Lok Wing Joyce	Man	https://www.researchgate.net/profile/Lok_Wing_Man	osf.io/vyz7c	u3558392@connect.hku.hk	joyceman0701@gmail.com	3035583922
Sze King	Ko	https://www.researchgate.net/profile/Sze_Ko2	osf.io/a7d2z	u3564168@connect.hku.hk	koszeking@gmail.com	3035641683
Yu Tung	Law	https://www.researchgate.net/profile/Yu_Tung_Law	osf.io/nk7qa	u3558079@connect.hku.hk	lawsabine@gmail.com	3035580798
Ziqing	Yao	https://www.researchgate.net/profile/Ziqing_Yao2	osf.io/z7hj9	ziqingy@conncet.hku.hk	ziqingyao0705@gmail.com	3030056118

Managerial summary

Replication is crucial to scientific discovery in the sense that it arrives at a result that is close to the reality. One single finding could barely be convincing, while repeated findings of similar results pave the way for finding the true effect. Conducting replication studies not only could reduce frauds and biases, it also lay the groundwork further research. Recently, some replication crisis has emerged in the field of Psychology, such as the findings related to ego depletion and the Stanford Prison Experiment. Such crisis has resulted in extensive impacts since related psychological theories were circulated in the academic field and were applied in various industries. As such, there is a surging trend of conceptual replications in a bid to verify the facts based on the belief that a true effect underlying different phenomenon exists.

To enhance the scientific rigour of replication study and facilitate the advance in scientific knowledge, several regulations and practices are proposed by scientists. Four dimensions regarding the evolvement of replication will be discussed, namely the use of preprints, implementation of registered report, forcing open data and improving peer-review. These strategies are believed to enhance the transparency and reliability of scientific findings, refining current practices of replication investigation and thus paving the way to finding the true effect.

In depth report

What is changing/ happening: Replication crisis

When we talk about scientific discovery, for example, eating carrots can improve your vision, we expect to see evidence supporting its presence from different findings. If 3 studies state that carrots can improve our vision but another 3 saying that there is no significant result, the effect could hardly be regarded as convincing.   Therefore, to make sure that an experimental result is sound, researchers would replicate the findings in order to confirm the result. In this regard, the study of replication virtually affects every field of science including psychology.

A few years ago, a bunch of a scientist decided to re-ran 100 psychology studies mentioned in top notch journals that were published in 2008. As a result, fewer than half of the published results were replicated (Nosek, Cohoon, Kidwell, & Spies, J. R., 2016), which reinforces the statement of Dr. John Ioannidis, a Stanford researcher, “false findings may be the majority or even the vast majority of published research claims.'' The falsifiability of published findings is worrying and it deserves particular attention

Having a high reproductivity is crucial for scientific research. Not only can we reduce frauds and biases but also lay the foundation for further studies and research. Manipulate data by choosing the data that support the hypothesis is easy yet not ethical. We are not finding the results that we want to find, but those telling the truth and enclosing the true effect. If the results are not significant, there should not have a selective illusionary research result to prove the relationship or correlation between factors in a relectant manner.

The journal Nature published a survey a few years ago and asked researchers if they thought there was a reproducibility crisis in science and 90% of them thought so. And when a medical study having an invalid conclusion, researchers may misguide the treatment based on the faulty conclusion. The patient may get sicker even and tons of money may be wasted to support dead-end research. If the consequence of the replication crisis is so tremendous, why no researcher stop psychology’s replication problem?

There are a lot of answers to this question. But one of the unanimous answers from researchers is that psychology journals are responsible for the replication crisis. As Koole, S. L., & Lakens, D. (2012) has stated, the majority of the researchers understand the importance of replication. However, the editors of the journals generally recommend against the publication of replication (Neuliep & Crandall, 1990, 1993). Using the Journal of Personality and Social Psychology (JPSP) as an illustration, when an independent research team submitted a manuscript describing three failed replication attempts to JPSP, the editor refused to send out the paper for review (see Aldhous, 2011, for more details; see also Galak, LeBoeuf, Nelson, & Simmons, in press, for an update). And this is not the only case of similar publication policy.

Since psychology is originally categorized as human science, it is also studied by the religious and philosophical scholars. They are using hermeneutic-interpretive approach which emphasizes the subjective interpretations of the researchers(Grondin, 1994). Therefore, the result of the research accepts diversified answer and replication serves a very limited function. A replication may consider as impossible since the result is believed to be unique (Koole, S. L., & Lakens, D., 2012). But when modern psychology began to switch from human science to natural science, the old method may be outdated. The old approach does give the field of psychology a lot of insights yet there is a drawback: it goes against the publication of replication. Replication is basically a repetitive version of the original article while publication would like to deliver new discoveries and new results. This phenomenon makes quite a lot of sense since replication didn't sound as attractive as a new discovery and getting much attention. Just imagine, who would like to say 'I found the effect that somebody found yesterday! '. When it comes to the field of psychology, there is a chance that researches are being affected by confirmatory publication bias. Therefore, disconformity replication tends to be more surprising and receives more attention. As a result, researchers tend to generate a disconformity replication which leads to controversy. This phenomenon is putting the field of modern psychology into trouble since it is unproductive.

Here is one of the examples of the replication crisis would be: in 1998, Roy Baumeister investigated the question asking whether self-control is a limited resource that takes energy and motivation to maintain our attention. Therefore, they conducted an experiment to test the hypothesis. The experiment ran like this: there are 2 self-control tasks back-to-back to see if there is a decrease in self-control on the second task. One task is restraining oneself not to eat the cookie and the other is solving an unsolvable puzzle. Then, the researcher compared the group that can eat a cookie (no self-control task) and the one who cannot (with self-control task) to see if there is a significant difference in the amount of time the participants will try to solve the impossible puzzle. They concluded that once we had resisted temptation, it is a lot harder for us to do it again. And they named it ego depletion. And it has hugely affected the research and study later on. It has been applied in many things like dieting strategies and athletic training tactics. However, in a meta-analysis that is led by Martin Hagger (2010), there is no significant difference between the two groups. It was done by 24 difference labs in different languages and countries and showing that there is no ego-depletion effect. Although there is a twist in the experimental method, it suggested that it could be applicable to some people.. The replication argued that ego-depletion may only under very specific circumstances. Some scientists argue that one disconformity replication is not enough. However, it still served as a cautionary tale when it comes to psychological research. This is not the only example of replication crisis.

Another example would be the stanford prison experiment. To test whether brutality is dispositional or situational, a mock prison experiment is being conducted with presetted rules to protect the participants. However, things were out of control when both the research psychologist and the students participating are too involved in their roles. This experiment has been perceived as not moral and ethical but the truth is, a student participant claimed that the experiment was an act to ensure his professor could have a significant result(F.& Jeanne, 2018) It is believed that this is not the only experiment that the participants and data were being manipulated.

In recent years, researchers began to admit that they have not produced results that are as ideal as they hoped. Therefore, Many lab was specially designed for replication study. This is specifically designed to address the criticisms of replication: (i) small sample size, (ii) lack of knowledge on the original experiment and (iii) lack of cross-cultural student. They recruited 60 times more volunteers as the studies from 36 different countries. The 186 researchers from the lab will check every detail of the original experiment beforehand. Although they failed to replicate 14 out of 28 classic psychology experiment, it is the first step for psychology enter the field of psychology as a science.

But anyways, the idea of replication emerged due to the assumption that nature behaves lawfully. (Dilworth, 1996) To enhance the reproducibility of the journals, it is necessary to state the primary information concerning both material and immaterial information clearly to decide what to keep and change in the replication aiming at verifying a fact or a knowledge. Examples are the characteristics of the participants with specific research history, the physical setting of the experiments, the selection and the allocation of the participants to name but a few. (Schmidt 2009)

Furthermore, a point to note is that it is not possible to replicate an experiment in exact. In the first place, it is impossible to run the experiment again by the same subjects, at the same time spot and holding all the variables the same. Also, an exact replication do not have confirmatory power (Collins, 1985). Imagine that even if we want to see whether the experiment that complies with the hypothesis is valid back in ten years, it is impossible to have the exact same subjects, tools and participants in the same state physically and mentally. On top of that, having the exact same components and research materials cannot restate that the hypothesis is valid since we cannot inference. Therefore, conceptual replication is being adopted commonly in replication study so as to use methods and measures that are similar to the original study for inferencing the concepts and theories.

On the other hand, there are also people claiming that fail to replicate is not an alarm to the field of psychology as it is part of the new discoveries under specific circumstances and conditions. It is claimed that as long as the experiment is well designed and executed, the same procedures can lead to opposite results. (Owens, 2018) This could be related to sampling size, yet it is suggested that the hypothesis in the original study before replication might not be correct as the relationship between two factors should be strongly correlated with even in different contexts. For example, in the famous fear learning experiment (Barrett, 2015) while a rat is being put on an electrical grid in a box, giving an electrical shock after a tone would cause that rat to freeze and also its heart rate and blood pressure would rise. But does this means that whenever the rat is shocked or listened to the tone, same reactions would occur? Probably no. Because when the rat is being restrained during the tone, its heart rate decreases. Conceptual replication, which means rather than aiming for replicating the experiment fully to test whether the phenomenon is approaching reality, testing the concept using different subjects and similar tests can then help inferring the significant elements in the context that contributes to an effect, but not merely concluding that every situation has a different outcome.

Consequence of the replication crisis

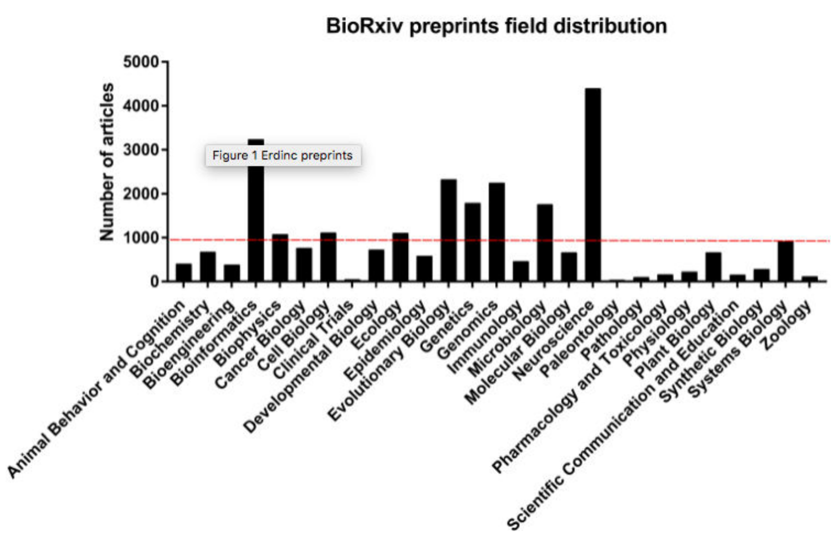
The replication crisis was brought under the limelight by the scientists. The journal *Nature* survey reported 52% of scientists believe the replication crisis is significant (Baker, 2016). In practice, 70% of researchers failed to reduplicate another scientist's findings (Baker, 2016). Therefore, recent journal focus on this replication crisis have stricter rules and regulations to ensure the quality of the published research. Online Psychology platform such as the Collabra started to accept replication reports which are scientific, methodological, and ethical rigor. There are a few guidelines proposed by the editors and reviewers of the Collabra:Psychology :

- Not trying to predict the paper's impact to the field
- No topic bias
- Check for rigorously and transparently conducted
- Statistically sound
- Significant powered
- Worthy of inclusion in scholar record

More specific and influential replication.

Preprint revolution

Preprint is an article which has not yet gone through the process of peer review. There are usually online public servers allowing writers to upload their article and receive comments from peer reviewers, before sending the paper to the publishers (Sezgin, 2018). For authors, the peer review platform provides advice to amend their work. For reviewers, the platform serves as a useful learning platform for them to interact and learn from the authors. Such a process of exchanging academic knowledge serves as a stepping stone to the improvement in rigor of scientific findings. Nowadays, multiple journals that come from numerous fields utilize the preprint servers to improve improve the quality of scientific study, especially in neuroscience and bioinformatics journals, as shown by the figure below:



Source: Medical Sciences Division, University of Oxford

Implementation of Registered Reports

There is a change in the criteria for the acceptance of a publication (Nosek, Ebersole, DeHaven, & Mellor, 2018). In order to achieve scientific breakthroughs, publication should be granted based on the relevance of the question investigated and the appropriateness of the chosen methods and study design. If the answers to the important questions are obtained in an appropriate manner, even negative results, the publication is still important. Thus, publication should not depend on the results (Spiller, 2018). Pre-registered studies and registered reports are therefore highly encouraged in the field. (Meta-psychology, n.d.)

Stage 1: review of introduction, method, proposed analyses, and pilot data

Direct replications are not only important to address earlier work. They are also necessary if researchers want to further explore a finding stated in exploratory research, for example in a pilot study. In particular, the approach should be make explicit the procedure that is likely to reproduce the finding and pre-register that procedure, followed by running the experiment (Zwaan, Etz, Lucas, & Donnellan, 2018). With this approach, one would not assume that the initial procedure was an appropriate test and makes the whole scientific investigation more transparent. In practice, some key steps need to be followed in this stage, which are as follows:

> (i) First, authors submit a manuscript outlining the relevance of the question aimed to be investigated, their proposed methods, a power analysis with a sample size estimation, a detailed data analysis plan, and preferably pilot data.
>
> (ii) Furthermore, the plan is registered in a database open to the public. Then, the manuscript will undergo stage one peer review. If the manuscript fulfils the above mentioned criteria, the journal offers in-principle acceptance. This means that as long as the authors follow the protocol they submitted, the manuscript will be published when the study is completed, regardless of the results

It is sometimes argued that conducting replication studies may not be desirable or even possible due to practical concerns. For instance, large-scale observational and clinical-epidemiological studies (Coyne, 2016). Researchers who work in areas where replication is difficult should be particularly alert to such concerns and make concerted efforts to avoid the problems that result. Large scale developmental studies that follow participants for 30 and 40 years are one example as is research with difficult-to-study populations such as infants, prisoners, or individuals with clinical disorders. Researchers in such areas would benefit from pre-registering their hypotheses, designs, and analysis plans, to protect themselves from concerns about researcher degrees of freedom and the use of questionable research practices (Zwaan, Etz, Lucas, & Donnellan, 2018).

Stage 2: peer review of intro, methods, results, and discussion

In the next step, the study is conducted and the manuscript is finalized hy filling the results and discussion sections. On the other hand, the introduction and methodology should remain unchanged. If there are deviations from the protocol, the author(s) should accurately and reasonably explained why do they need to change the original plan.

After that, the completed manuscript goes through stage 2 review, meaning the reviewers assess if the protocol is addressed and followed and whether power is sufficiently reached. Last but not least, the reviewers would assess the relation between results and discussion. If available, the manuscript is published without regard to the results (Zwaan, Etz, Lucas, & Donnellan, 2018).

In short, the practice of the registered report is indeed a comprehensive and structural process of developing scientific findings. Without pre-registered process and peer reviews, authors might seek to obtain positive results, which might result in researcher biases and many weakly founded studies. By implementing stages of peer reviews and pre-registration, the rigor, reliability and persuasiveness of scientific findings can be enhanced by concerted efforts between authors and readers.

Forcing open data/code - JDM (Journal of Database Management) / Cognition

Data should be clear and shared so as to allow further replication in the future. When using the original data set, it is recommended to use a trustworthy digital repository to preserve data. The requirement for these trusted digital repositories is that they have to be discoverable, accessible, usable, and preserved for the long term (Collabra: Psychology, n.d.) Usually repositories provided by institutions like Universities will be a good target. Accession numbers or DOIs are also required to be provided (if any), so as to allow access to these data for replication. Program codes and all materials allowing full replication of the experiment is required for authors reusing data from public repositories. There are even guidelines to authors using repositories which are difficult to access due to the protection of privacy. They should explain the restriction in detail, also to follow up the problem by giving alternative ways to access for a set of data.

Authors are in these journals are required to provide data set as well as information listed below (Collabra: Psychology) :

- all variables, treatment conditions, and observations described in the manuscript.
- a full account of the procedures used to collect, preprocess, clean, or generate the data.
- provide research materials and description of procedures necessary to conduct an independent replication of the research.

Open-access / no-fee journals

Researchers on the topic of replication are suffering from the lack of publishcation resources functions. Therefore, many of the replications are waiting to be published.(Suls & Martin, 2009) Under this trend, the idea of open-access journals is first proposed by Hartshorne and Schachner (2012). However, the idea may have a main drawback which is the lack of readers. Most of the readers are fond of the traditional journals which will not specialise on replication reports, even when there are some open-access journals subjecting to replication issues, the number of readers is still worrying. Therefore, there are suggestions that traditional journal can make room publishing these replications. Traditional journals may need to concern on the maximum pages per issue in the past, but in this electronic era, the online e-version journals have no such restriction anymore. (Nosek & Bar-Anan, 2012)

The idea of online publishing appear to be beneficial to no-fee journals. Journal like the Meta-Psychology provide a free platform for authors and readers to read scholarly articles. The journal provides a permanent archive which will be convenient to preserve papers for a long time. Although it is a no-fee journal, the authors can still own their copyright under the CC-BY license. (Meta-Psychology, n.d.)

Improving peer-review: Open peer-review / signing reviews / forcing data/code sharing.

Using Collabra: Psychology, an online journal, as an example to investigate the improving of peer-review. They are operating with two types of peer reviews: single-blind or non-blind peer review process.

The idea of single-blind peer review process is that only the name of the author is available to the reviewers, while the reviewers remain anonymous. And for non-blind peer review process, reviewers can sign their reviews. (Collabra: Psychology). Researchers can also blind the analysis, or set aside a certain proportion of the data for a confirmatory test. As a result, discussion sections from papers that describe these results can be appropriately calibrated to the strength of the evidence. (Zwaan, Etz, Lucas, & Donnellan, 2018)

Open peer-review is also created when requested. It is a system that the author of the submitted paper can see the comments from the reviewers in a single document. There will not be any secret comments given by the reviewers to the editor. All reviewers can have their comments cited and everything is done is a transparent and open way.

The significance of replications

Replication values in a scientific research

A fundamental process of scientific research is the systematic procedure of conducting replications. This is a crucial practice within the evaluation course, and serves multiple purposes (Schmidt, 2009). First is that replication allows verification of scientific findings and solidify our understanding of nature. Second, replication also serves as a norm in methodology which contributes to establishing scientific facts. In addition to the above benefits, replications also provide relevance and improvement to the original study from a functional approach (Schmidt, 2009). Type I errors in research can not be eliminated completely, thus replications assist by controlling the sample error in studies, improving the quality of study and robusting the hypothesis in hand. Many studies also states that there is no one particular variable that is responsible for the change, therefore replications can aim to replicate the primary focus of study as identical as possible while altering other variables, strengthening the correlation between the primary focus and the effect of interest, which in turn also minimises the chance of fraud in studies. Furthermore, replications can enhance the sample size and thus are generally more well-powered than the original findings.

However, the concept of replications and its findings generated many controversies among researchers. More specifically the phrasing of replication studies, the statistics results of replications, message and interpretations conveyed by replications and the relative significance of different methods of replications (Zwaan, Etz, Lucas & Donnellan, 2018).

Concerns of replication

The difficulty to replicate original findings can be explained by various reasons, the first may be the “publication bias”, which refers to the subjective preference to publish findings that holds support to a certain hypothesis while disregarding other studies that have shown null findings and is of insignificance. When this becomes a habit among publishers, the frequency of false positives in published research will increase as an aftermath. Another possible factor may be due to the contribution of unrestricted “researcher degrees of freedom”. When researchers expects to find significant results, they are allowed to manipulate variables in favor to show significant results, yet the methodology and analysis are not always clear and open to the public (Zwaan, Etz, Lucas & Donnellan, 2018).

When conducting a replication, it also raises many concerns regarding the quality of replication. One common argument for the failure to replicate findings is that the change of context from the original to the replication may render different results due to historical, geological or even unidentified factors replications may fail. A good explanation by Barsalou (2016) states that there are too many uncertainties within an individual which would alter their personal experience to the stimulus, hence it may be difficult or even impossible for researchers to conduct a replication accurately. This dilemma also handicap evaluations of whether the original article is a false-positive or whether replication failed because of context changes.

Another concern is related to the basic nature of replications. Some claim that replication are irrelevant since replications provide inaccurate or false information. The challenge to replications come in both ways as stated by Crandall and Sherman (2016), if the replication fails it may imply that the particular method used is problematic, if the replication succeed it may also imply that the theory itself has high replicability yet there is still poor test of theory, meaning that the replication is only supporting an invalid statement. Therefore it may seem that the implementation of a replication is redundant and holds low value to examine the original results.

Replications also comes with limitations and cost (Zwaan, Etz, Lucas & Donnellan, 2018). Subjects such as psychology as a domain where the focus is on subjective feelings but measured objectively, replication seems to be incomprehensible under practical concerns. Meanwhile some studies are associated with exclusive events such as natural disasters, doing a replication of that would be impossible. The outcomes of a replication comes with costs, it might damage the reputation of original authors or even cause perceptions that replications are tools for accusations.

Proposed methodology for replications and implementations

In prospect of replications, there are few improvements that could be made in replications so that the results derived could hold more significant values in scientific research. One of the current controversies that replications have now is that they are not mainstream and are irrelevant to research. In response, Coles, Tiokhin, Scheel, Isager and Lakens (2018) suggest that future replications should be based on a decision-theory framework, where researchers evaluates the expected utility (cost and benefit) of a replication. Factors such as reliability of the literature, relevance among community or significance of theory drives the motive whether or not replication is necessary. In return, researchers can gain productive discussions and optimize efficiency in research.

The analysis of replications should also shift from comparing with target article to a continuous evaluation and multilevel modeling (meta-analysis). Since original article lies on conventional methods of “false-positives”replications may turn out to be useless even when there are findings. Instead, continuous measures are preferred over threshold-based criterion (e.g. $p < .05$) (Gelman, 2018). Therefore direct replications should be discarded and replaced by a meta-analytic approach to better evaluate replications.

The current incentive structure in the domain of psychology is that researchers are better off doing original research as they get rewarded mostly through increased publications and reputation. Yet the ideology of replication practise promotes that researchers collaborate and examine each other’s work, thus the quality of research done could not be benefited by replications. In order to make replications mainstream, Koole and Lakens (2012) proposed strategies to encourage replications. First would be co-publications, where the creation of new publication outlets for replication could be promoted. The following are examples of new journals in support of open-science and peer evaluations:

Collabra : Psychology

Collabra : Psychology is an Open Access journal that is provided to be publicly evaluated by the psychology community. The research in The Collabra : Psychology is peer-evaluated. It is written in a scientific and methodological way. The research will be checked in a very thorough and transparent way by a large team of peer-editors in the Collabra : Psychology team. The journal contains 7 sections standing for the broad field of psychology.

Meta Psychology

Meta Psychology is a free Open Access journal that is open for professional peers to review in the psychology community. Novel methods and tutorials are very welcome by Meta Psychology. This new journal also welcomes the cross-disciplinary work that is beneficial to the psychological field. Even more, it links up the discussion of psychology in social media such as Facebook and Twitter and traditional journals. This Journal will offer a way for the editors to upload their academic work after a thorough review by peers. The manuscripts are peer-reviewed openly and everyone can also be the co-editors after adding rigorous comments on the manuscripts.

AMPPS

Advances in Methods and Practices in Psychological Science are provided to conduct the whole scope of areas within psychological science. Different part in psychological science is encouraged to combine methodological and analytical questions and write up the advanced methodology. The new methodology is encouraged, for instance, direct replications which is not commonly publish and multi-lab collaborations. The editor of the journal added amendments in reporting practices to have a more thorough and transparent report. The editor also encouraged the open data, open materials, and pre registration by rewarding badges.

International Review of Social Psychology

The Journal valued the research of social psychology that using high quality of scientific method. It is encouraged to the scholar to discuss about advances in practices, research design, and statistical methods. The high-transparency and high-completion of research is encouraged by the editor through rewarding badges. The Journal is reported in a brand-new way on the articles and research. It newly launched Registered Replication Reports to appraise the advantage of evidence for substantial effect. Social psychological analysis of collective movements is encouraged to be published. The editors welcome the research analysis on the factors of breaking out social movements and the impact on representations and intergroup relations.

Yet there are limitations to this approach, where long-established journals consisting of the majority of readers may not necessary turn focus to specific replication journals, thus it can only bring minimal contribution to the popularity of replications. An alternative strategy would be to let existing journals incorporate replications into their publications, despite the fact that there are limits to publishing replications within existing journals, nevertheless it can foster the practice of replications.

Apart from co-publications, co-citations may also bring awareness to replications. By requiring studies to cite multiple replications along with the original study as a regulation for establishing a phenomenon, it can guarantee research done in a more meta-analytic approach (Cumming, 2008). This also incentivize researchers to conduct replications, increase appealing to journals and reinforce popular original studies.

One single finding holds little to no value, but repeated findings of similar results says something about the nature. Replication in scientific research is without doubt a pivotal focus, and it aids the advancement of our knowledge and studies. Since implementation of a contemporary practice will surely be flawed, there are ways to refine and polish current ways of doing replications. The trend of replications in studies is only on the rise, soon it will redefine our perspectives on analysis of studies, and we hope that the current replication crisis could be resolved.

Quiz

Write 10 easy quiz 4 multiple choice questions to ensure reading and understanding of this chapter:

1) Which of the following is NOT the common criticism of replication?

A. Small sample size
B. Lack of cross-culture study
C. <u>Conformational bias</u>
D. Lack of knowledge on the original experiment

2) How can we categorize psychology?

A. Human science
B. Natural science
C. <u>It can be human or natural science, just depends on the approach of the researcher</u>
D. It is not science at all

3) What is the assumption of replication?

A. <u>Nature behaves lawfully</u>
B. Every person is rational
C. There are no sampling error
D. Every individual is the same

4) Why traditional journals found it very difficult to make rooms for replication report to publish in the past?

A. They have never heard of replications
B. Replication reports are always poorly done
C. No one is interested in the replication reports
D. <u>They need to concern on the maximum pages per issue</u>

5) Which field is producing the greatest amount of preprints ?

A. Cell Biology
B. <u>Neuroscience</u>
C. Bioinformatics
D. Zoology

6) In-principle acceptance offered by the journal secures the publication of manuscripts regardless of which of the following components?

A. Research background
B. Proposed methodology
C. Relevance of the question aimed to be investigated
D. Results

7) What is the peer review process called when reviewers can sign their reviews ?

A. Open-eye peer review process
B. Single-blind peer review process
C. <u>Non-blind peer review process</u>
D. Non-closed peer review process

8) Which of the following is NOT a function of doing replications?
A. Verification of scientific findings
B. Solidify knowledge
C. Analysis of original study
D. <u>Disproving results</u>

9) Which of the following is NOT a concern regarding the quality of replications?
A. Context inconsistency
B. Limitations and cost
C. <u>Popularity of replications</u>
D. Findings from replication

10) What is the current incentive structure of doing research?
A. <u>Researchers are favored in doing individual work</u>
B. Researchers are favoured in doing collaborative work
C. Original studies are less preferred
D. Replication studies are more preferred

References:

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature. DOI: https://doi.org/10.1038/533452a.

Barsalou, L. W. (2016). Situated conceptualization offers a theoretical account of social priming. *Current Opinion in Psychology, 12*, 6–11.

https://doi.org/10.1016/j.copsyc.2016.04.009

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? Journal of Personality and Social Psychology, 74(5), 1252–1265. https://doi.org/10.1037/0022-3514.74.5.1252

Coles, N. A., Tiokhin, L., Scheel, A. M., Isager, P. M., & Lakens, D. (2018). The costs and benefits of replication studies. *BEHAVIORAL AND BRAIN SCIENCES, 41.*https://doi.org/10.1017/S0140525X18000596

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology, 66*, 93–99. https://doi.org/10.1016/j.jesp.2015.10.002

Cumming, G. (2008). Replication and p intervals: P values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science, 3*, 286–300.https://doi.org/10.1111/j.1745-6924.2008.00079.x

F. Neil Brady, Jeanne M. Logsdon, (1988). Zimbardo's "Stanford Prison Experiment" and the relevance of social psychology for teaching business ethics. https://link.springer.com/article/10.1007/BF00382981#citeas

Gelman, A. (2018). Don't characterize replications as successes or failures. *BEHAVIORAL AND BRAIN SCIENCES, 41.* https://doi.org/10.1017/S0140525X18000638

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. Psychological Bulletin, 136(4), 495–525. doi: https://doi.org/10.1037/a0019486

Hartshorne, J. K., Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. Frontiers in Computational Neuroscience, 6. Advance online publication. doi:10.3389/fncom.2012.00008

Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science, 7*(6), 608-614.https://doi.org/10.1177/1745691612462586

Lisa F. B. (2015). Psychology is not in crisis, New York Times. Retrieved from https://www3.nd.edu/~ghaeffel/ScienceWorks.pdf

Meta Psychology. (2019). About the Journal. Retrieved from https://open.lnu.se/index.php/metapsychology/about.

Nosek, B. A., Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. Psychological Inquiry, 23, 217–243. https://doi.org/10.1080/1047840X.2012.692215

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences, 115(11), 2600–2606. Doi: 10.1073/pnas.1708274114.

Owens, B. (2018, November 19). Replication failures in psychology not due to differences in study populations. Retrieved from https://www.nature.com/articles/d41586-018-07474-y.

Patrick E. S, Joseph L. R. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis,Annual Review of Psychology 2018 69:1, 487-510 https://doi.org/10.1146/annurev-psych-122216-011845

Sander L. K., Daniël L. (2012).Rewarding Replications: A Sure and Simple Way to Improve Psychological Science. doi: https://doi.org/10.1177/1745691612462586

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*, 90. https://doi.org/10.1037/a0015108

Sezgin, E. (2018). Preprints: revolution and resistance. Medical Science Division, University of Oxford. REF: https://www.imm.ox.ac.uk/about/blog/preprints-revolution-and-resistance.

Spiller, T. & Olff, M. (2018). Reducing uncertainty in research: Introducing registered reports. *European Journal of Psychotraumatology*, *9:1.* Doi: 10.1080/20008198.2018.1554417.

Stefan S. (2009).Shall We Really Do It Again? The Powerful Concept of Replication Is Neglected in the Social Sciences, Review of General Psychology 13(2):90-100, doi: 10.1037/a0015108

Suls, J., Martin, R. (2009). The air we breathe: A critical look at practices and alternatives in the peer-review process. Perspectives on Psychological Science, 4(1), 40–50. doi:10.1111/j.1745-6924.2009.01105.x

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replications mainstream. *BEHAVIORAL AND BRAIN SCIENCES, 41*. https://doi.org/10.1017/S0140525X17001972

Nosek, B. A., Cohoon, J., Kidwell, M., & Spies, J. R. (2016). Estimating the Reproducibility of Psychological Science. doi: 10.31219/osf.io/447b3

10 - Changes to education / academia

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Radhika	Bansal	https://www.researchgate.net/profile/Radhika_Bansal4	https://osf.io/wzrp7/	radhikab@connect.hku.hk	radhikaban2407@gmail.com	3035453828
Joshua	Bernal	https://www.researchgate.net/profile/Joshua_Bernal2	https://osf.io/x5jwv/	jdkb9701@connect.hku.hk	josh.dkb97@gmail.com	3035399852
Hiu Nam Trista	Hung	https://www.researchgate.net/profile/Trista_Hung	https://osf.io/9pqsz/	u3547693@connect.hku.hk	hung.hiunam@gmail.com	3035476935
Shubham	Kumar	https://www.researchgate.net/profile/Shubham_Kumar150	https://osf.io/wt2p9	u3553652@connect.hku.hk	troysk97@gmail.com	3035536527
Tsz Foon Justin	Leung	https://www.researchgate.net/profile/Justin_Leung13	https://osf.io/udcsq/	u3547502@connect.hku.hk	justinleungfoon@gmail.com	3035475022
Haeon	Park	https://www.researchgate.net/profile/Haeon_Park	https://osf.io/58s6d/	u3512036@connect.hku.hk	haeonpark@gmail.com	3035120366

Managerial summary

In this chapter, we discuss about the replication crisis and how it is affecting researchers and academicians in the field of Psychological Science and teaching environments. As more researchers recognize the problems with reproducing replication studies, e.g. an insufficient statistical power and non-transparent experimental method, many are striving to bring changes. Examples of changes in academia include the establishment of the Psychological Science Accelerator (PSA) and Collaborative Replications and Education Project (CERP). They are based on the idea of crowdsourcing in which everyone can share their projects and peer-review them. Prominent university institutions such as Michigan State University and the University of Hong Kong are also taking the replication crisis into the classrooms. Undergraduates are encouraged to perform independent replication studies. Furthermore, the Open Science Framework (OSF) lays out several principles for Open Science. They provide step-by-step procedures for obtaining reproducible experimental scientific findings.

This chapter consists of five sections: Changes in academia, Changes to the syllabi, teaching replications, HKU mass replication and Open Science Manifesto. It is hoped that after reading this chapter, readers will understand further the various efforts in tackling the replication crisis in academia.

In depth report

Changes in academia

The previous chapters have emphasised that results based on insufficient findings and non-representative samples pose a threat to the accuracy and generalizability of results. In light of this development, psychological practices have evolved towards flexible data analysis and appropriate statistical power. As researchers are becoming aware of the aforementioned limitations of previously conducted studies, the popularity of crowdsourced research is increasing. Crowdsourced research is a large-scale collaboration of research projects that are conducted across multiple labs. The Psychological Science Accelerator (PSA) is a distributed network of laboratories that formalised crowdsourced research. PSA's mission is to contribute to the collection of generalisable and reliable findings. PSA is ongoing, efficient, diverse and inclusive. Its efficiency comes from reusing structures and platforms for different projects. It is diverse in both its subjects and researchers as well as inclusive of all relevant input from anyone (Moshontz et al., 2018).

Benefits of crowdsourced research and PSA

1. Crowdsourced research projects have a large sample size and thus can achieve high statistical power. The number of subjects available for a study can be a confound. Crowdsourced research help researchers reach results that they might not have been able to achieve independently by accumulating data from many labs. Precision is maintained by controlling variables to reach better effect-size estimates.
2. Crowdsourced research has transparency of documentation and research process. This allows for secondary publication based on these findings and data.
3. Crowdsourced research endorses inclusion and diversity. Researchers from any part of the world can collaborate with labs and find participants from all over the world. This includes countries that are under-represented in current scientific literature.
4. In addition to the aforementioned, the PSA promotes decentralisation of authority and projects are chosen by committees. All PSA projects must be pre-registered to ensure further transparency. Lastly, all projects are reviewed by multiple researchers with an expertise in data analysis to endorse statistical rigor.

These features of crowdsourced research can further the reliability and generalizability of findings in psychological science. In implementing crowdsourced research, large amount of resources are required to establish a wide network and maintain collaboration between various labs. Nevertheless, the PSA evaluates proposed research ideas, assigns them to different labs, gives ethics approval and oversees data collection and analysis. All these help lower the barriers to crowdsourced research.

The internal working of PSA:

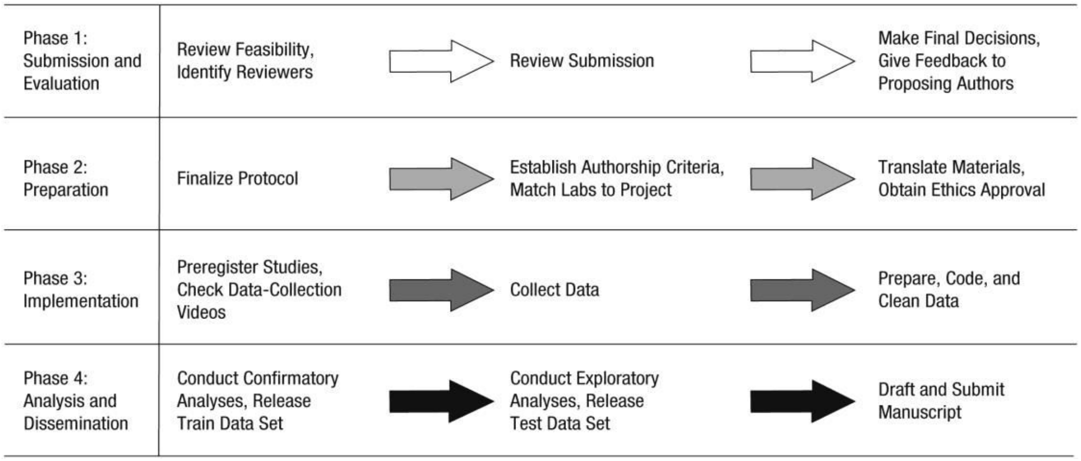


Figure 14. Adapted from the psychological accelerator: advancing psychology through a distributed collaborative network (Moshontz et al., 2018)

The democratic and diverse research of PSA ensures that the research conducted within its network reduces biases and reflects the errors of an individual. NO single individual in the network has ultimate control over the research process (including which projects to choose, materials, analysis etc). For each of the tasks, there are committees that lead the decisions. This eventually leads to high levels of scientific rigor.

Changes to the syllabi

Collaborative Replications and Education Project (CREP) is a volunteer crowdsourced replication projects managed by the faculty and students of academic institutions focusing on psychology. Students have to complete replication projects from the start till the end as a part of their coursework. Students partake in open-science practices and make use of transparent methodologies (open data and pre-registered hypothesis etc). Students also engage in peer-review and revise-and-resubmit processes until their work reach the acceptable standards of the scientific community. The process also ensures data fidelity and quality both before and after data collection and analysis. Students taking part in the replications can contribute to authorship (Wagge et al., 2019).

Independent of the CREP, some institutions are also conducting replications at an undergraduate and postgraduate level. These replications follow the protocols and are conducted using open-science methodologies. These disclose the pre-registration, data, materials and ethics approval beforehand (Hawkins et al., 2018).

Teaching replications

Current status of replication crisis

The replication crisis is affecting universities and academias in terms of how students should approach experimental methodology in psychology. Many researchers and psychologists acknowledge the lack of certainty in ensuring replication reliability of published scientific literature. According to Nature (2016), 1,576 scientists were asked about their reproducibility of other group's experiments. Strikingly, 70% of the scientists agreed that they have failed to reproduce the experiments and there was indeed a replication crisis as an ongoing methodological crisis. Even though scholars have acknowledged such crisis and reckoned the importance of the published literatures' reliability, no one can afford replicating every important factor of the experiments. Although a combination of appropriate sampling procedures and statistical testing can help estimate reliability, there are still certain loopholes not being handled. These include the number of other experiments conducted by the same researchers (Rosenthal, 1979) and the analytic decisions not accounted for in statistical calculations (Simmons, Nelson, & Simonsohn, 2011).

Thus, special issues on replication, refined journal policies on transparency and study design, and new ways to detect questionable research practices have all emerged for quantifying and alleviating the concern over the reliability of psychological research (Simmons, Nelson & Simonsohn, 2011, 2013; Spellman, 2015; Wagenmakers, 2015). Moreover, researchers and psychologists are suggesting that students should replicate cutting-edge studies, in an effort to testify reliability of experiments, and provide evidence for their credibility in serving as a basis for further research (Frank & Saxe, 2012). In terms of resolving the replication crisis, such an approach would provide the scientific community with a larger captive workforce, combining the efforts from both students to teachers. For students, the younger generation would be able to learn about the contemporary replication crisis plus the standard scientific process through hands-on experience, since as a replicator doing real science rather than merely a reader, they would be more motivated to focus on details in methods or analysis. Involving students into such huge projects, the benefits to them would be profound. They can develop their own experimental methodology and obtain deeper insights into scientific analysis along the way. To look at the approach in a bigger picture, it can promote a culture shift that new work in the experimental-psychology community builds more on previous findings.

Bringing replication crisis into classrooms

As mentioned, forward-leaning researchers and academicians has accepted that there is a lack of reliability in today's scientific literature. Hence, prestigious universities and research institutions suggested that we should involve students in replications, from which students would also gain the merits of developing their own scientific methodology and gaining the ability to evaluate the replicability of findings. However, before analysing the practicality of the proposal, we should also ponder on whether students can fully understand the nature of replication crisis.

Professor William J. Chopik and Professor Ryan H. Bremmer developed a 1-hour lecture discussing the replication crisis in Psychological Science (Chopik et al. 2018). It aims to gather the issues surrounding study rigor and the replication crisis as well as recommendations for increasing reproducibility. The professors also introduced other scholars' assessment on the replication crisis as well as statistical discussions on reproducibility and data analysis that could be easily comprehended by undergraduates. Their lecture slides and pre/post survey questionnaires can be found on Open Science Framework website. From their findings, they showed that the lecture was indeed effective in conveying the most important issues about the replication crisis — after the lecture, students 1) realized that media attention was not an accurate indicator of the reliability of studies; 2) showed high levels of agreement with current suggestions about transparency and reproducibility, such as determining a sample size before running a study, making data publicly available, and reporting studies that "don't work out."; 3) understood that flexible statistical decision-making can lead to questionable and significant findings; and 4) correctly stated that the studies presented in the press releases may have trouble replicating (Chopik et al. 2018). Although students showed a lower level of trust towards previous psychological studies, it is still a success as it indicated that students reckoned the replication crisis, which would in turn motivate them to focus on the study design elements that increase reproducibility, hence allow them to identify replicable/ irreplicable studies. Furthermore, they shared their lectures and data onto OSF for everyone's free access. This serves as an effective pedagogical tool for teachers with only a cursory knowledge of issues related to rigor and reproducibility; as for students, they can learn about the existing replication crisis and approaches for enhancing reproducibility.

In-class replications

Some may have concerns over the practicality of performing in-class replications, which then leads to the part considering how to make it work. It requires the efforts combining multiple facilitating factors, and the first and foremost one would be choosing the right experiment to replicate. It is found to be helpful if students are provided with a menu of options, ranging from pencil-and-paper survey studies to more complex experimental paradigms (Frank & Saxe, 2017). With the help from tutors in creating the materials for more complicated experimental design, students can choose the project based on their own strength and interest. Besides, adequate data collection is essential in making replications possible. Students should pay close attention to the inferred effect size from original studies, rather than just p-values. They should also make sure they can collect enough data, e.g. making the experiment web-based and collecting data via M-Turk. Finally, frequent check-ins and feedback from tutors are critical to keep the replication projects on track. It allows chances to review key details and hence, students would be able to make timely refinement.

Some may argue that we should allow students to establish their own original study, given that one of the aims of such approach is to empower students to develop decent scientific methodology. However, due to the time constraint, this is arduous for inexperienced undergraduates to handle. Together with the lack of expertise, the experiments are often poorly designed, and unlikely to connect to current issues in psychological science (Frank & Saxe, 2017). Neither the students nor the community would be yearning for the result. On the other hand, direct replication can provide a jump start for students. As the original researchers have already contributed the idea and experimental design, students can focus on skillswise learning, which can greatly benefit them in making their own original studies in the future.

Apart from criticism over the ideology of the proposal, some may also worry if the relications would be costly. However, this is rarely the case — by making use of the free resources and university-licensed tools for data analysis and collection etc., it is possible to make low-cost, yet powerful replications. Of course, the proposal is not absolutely perfect, and it still induces a variety of concerns, e.g. how close to the original studies the replication should be, and whether students will be biased to choose simple, more replicable studies rather than more complex behavioural research. However, we should not be stumbled, and classrooms can be the place where we start to make a change.

The Gold Standard

Janz (2016) suggests that scientists should take the reproducibility and replication as the Gold Standard for scientific research. The Gold Standard can be embedded them into students' coursework. This can help establish the culture for replication and reproducibility.

The Gold Standard requires a differentiation between reproducibility, duplication and replication studies. "Reproducibility" refers to the quality which information of the studies is provided for readers to understand, evaluate and build upon a prior work (King, 1995). The American Political Science Association (APSA) emphasize that researchers must provide (i) data access, (ii) details of how they collected the data, and (iii) details of the analysis that led to their conclusions (Lupia & Elman 2014); in practice, we should provide supplementary sheet including the data file, software codes, original sources of the data, and explanation of how variables are transformed (Dafoe, 2014). "Duplication" refers to the process in which we verify previous research findings by aiming to generate exactly the same results with exactly the same data set and methods (Janz, 2016). On the contrary, "replication" is more than sheer duplication. It involves reanalysing the published work based on the original data, then replicate the study with newly collected data, new variables or new model specification (Casey, 2014).

In line with the aforementioned opinion, Janz (2016) suggested designing a new software for helping researchers and students to conduct research much easier, and increase transparency. In the overall, many researchers are devising ideas on ways to implement the teachings related to the replication crisis in universities, for the sake of creating a transparent and continued culture of replication and reproducibility.

HKU mass replication

The University of Hong Kong (HKU) plays an important role in the promotion of replication studies. To illustrate, the 6th World Conference on Research Integrity was hosted by The University of Hong Kong on June 2 to June 5 2019. This Conference provides participants from over 47 countries with opportunities to discuss and promote integrity in research, and in particular advancing discussion in key issues on integrity, innovation, and impact, e.g. open data, open access, research transparency, reproducibility, research misconduct management, and education on responsible research conduct (6th World Conference on Research Integrity, 2019). In addition, since joining the psychology department of HKU as an assistant professor in December 2017, Dr. Gilad Feldman and his team of teaching assistants have run 11 undergraduate and masters courses in HKU that focus on replication studies; and in HKU alone, Dr. Feldman has held five open-science workshops since 2017 (Feldman, 2019). He has also massmobilized undergraduate students to devote their efforts into numerous pre-registered replications. In 2018, Dr. Feldman and his students had completed 45 replications of impactful articles about judgement and decision-making (Feldman, 2019). This results in one of the largest replication efforts in social psychology. For the year of 2019-20, the HKU teaching development grant has provided funding for advancing the 20+ planned replications, which is valued at $250,000 HKD/~$32,000 USD (Feldman, 2019).

For each of the replication projects, students had fully pre-registered the research method design, collected data and done the coding, and written up replication reports in APA-style. Many of the replications also included interesting and insightful extensions. In order to store previous replications completed, Dr. Feldman has created a cloud drive named "HKU—Replications". This folder contains various invaluable resources, including but not limited to the importance of replications, means of pre-registering replication studies, procedures of seeking ethics approval, design and administration of Qualtrics surveys, objectives and directions for designing extensions, as well as HKU replication projects done previously under the guidance of Dr. Feldman. More importantly, all the replication studies are in support of Open Science. Pre-registered reports, data/code, and all finalized materials are publicly accessible through the cloud drive. This would facilitate future replication projects.

The following shows eleven examples of replication studies completed by HKU undergraduate students in 2018:

1. Baron & Hershey (1988): The role of outcome bias on the evaluation of decisions (outcome bias exists when people take outcomes into consideration in an irrelevant manner during the evaluation of a decision's actual quality)

2. Epstein, Lipson, Holstein & Huh (1992): The two information-processing systems—experiential and rational system—to which the cognitive-experiential self-theory refers

3. Fischhoff (1975): Hindsight bias (the tendency for people to estimate a higher likelihood of an event to happen when knowledge of an outcome is provided

4. Hsee & Weber (1997): Self-others differences in risk preferences (people tend to predict others as being more risk-seeking than themselves when facing risky options, no matter if the choices involved would bring about negative or positive outcomes)

5. Hsee (1998): The "less-is-better" effect (a phenomenon in which an option with lower normative value is judged more favourably than an option with higher normative value)

6. Kruger, Wirtz, Boven & Altermatt (2004): The effect of effort heuristic (our general tendency to use effort information, e.g. the amount of time spent to perform a task, as a mental shortcut to assess the quality of work) on the level of favourability of evaluation people give for artistic works

7. Kruger, Wirtz & Miller (2005): Action-inaction (whether most people tend to feel more regret when they fail because of taking action instead of inaction)

8. Shafir (1993): The effect of question framing on the respective proportion of people selecting the enriched or impoverished option

9. Slovic & Fischhoff (1977): Effect of hindsight bias (tendency to overestimate the predictability of an event's outcome after it is known) on perceiving predictability of experimental outcomes

10. Tversky & Shafir (1992): People's tendency in circumstances of uncertainty to violate the Sure-Thing Principle (if prospect x is preferred to prospect y no matter a person knows if event A has happened or not, then x will surely preferred to y even if it is unknown whether event A has occurred) by being reluctant to think through the implications of each outcome

11. Zeelenberg, Beattie, van der Pligt & de Vries (1996): People's tendency to be regret-averse and to make regret-minimizing choices

Aside from repeating the classical study, quite many replication projects done at HKU have also extended the study by adding one simple extension so as to supply additional insights that go beyond the original article. One of the following three types of extensions have been adopted:

1. Additional dependent variables (DV): The added DV will either be about evaluations/attributions/judgments regarding the scenarios/vignettes presented or present participants with a choice related to the presented scenario.

2. Additional well-known and validated individual difference scale at the beginning of the survey as predictors of the effect (independent variables, iv).

3. Additional condition(s) that make slight changes to the scenario presented: The added conditions are anticipated to pose no harm to participants going beyond the replication materials.

Regarding five of the replication projects above, the following extensions are implemented:

1. Baron & Hershey (1988): Perceived level of responsibility of the decision-maker in, respectively, a successful and failed outcome

2. Epstein, Lipson, Holstein & Huh (1992): Two dv scales for measuring the illusion of personal authorship (i.e. cognitive bias that we have caused certain events even though we have not)

3. Fischhoff (1975): Effect of outcome knowledge on the level of feeling surprised

4. Kruger, Wirtz & Miller (2005): Whether one perceives the bias that comes from counterfactual (i.e. thinking of possible choices based on past experience and then creating results that are equally advantageous and harmful to the person) and first instinct (i.e. the

answer that a person first thinks is correct) fallacy as a norm, whether one would be satisfied if he takes action and a positive outcome turns up, whether one perceives it a norm if most others feel satisfied if they take action and a positive outcome turns up

5. Shafir (1993): Respective rejection rate of the enriched option in situations where negative attributes are present and not present in the impoverished option

Open Science Manifesto

The need for embracing Open Science stems from the threats that have been identified from past literature. Over the recent times, researchers have been accused or even recognized as being tempted by issues like failing to control for biases while coming up with a hypothesis for their study, designing a study with low degree of statistical power, or even *P*-hacking(Munafò et al., 2017). As special emphasis is placed on *P*-hacking. *P*-hacking is a form of inflation bias and it denotes misreporting of the true effects in the published paper as the researchers run numerous statistical tests and selectively choose and report those which are significant. As success in academia is measured by the number of papers published and the publishers' prestige, researchers are incentivized to only report findings which are statistically significant (Head, Holman, Lanfear, Kahn, & Jennions, 2015). By doing so, academics are not only deceiving the world for their personal gains and popularity, but are also prosisting for phenomenon or factors which are insignificant or are remotely applicable universally. These corrupt practices should be eradicated from the academia and the only solution to it is acceptance of Open Science Framework for pursuing scientific endeavours.

Drawing Inspiration from Karl Popper

A prominent figure and advocate of Open-Science was Karl Popper. His ideology of science can be simply laid out as the fundamental principles of Open Science and its practice as demonstrated by Open Science Repository (The Editors of the Open Science Repository, 2012). While several principles put forward by Karl Popper are taken as the doctrine for Open Science, only a few are highlighted here.

1. Method of Science is by checking and evaluating theories and proposing a better explanation to the phenomenon concerned.
2. Previously upheld theories should be disregarded once deemed false, whereas new ideas and proposals of follow-up theories should be considered as a mark of scientific progress.
3. Scientific knowledge and its properties such as theories and arguments are all hypothetical in nature and susceptible to feasibility.
4. Science is not bounded by any authority.
5. Science entails our understanding of errors and their solutions and finding errors in proposed or accepted theories is one of the challenges. We should help discover errors in the commonly accepted or recently proposed theories.
6. Scientific knowledge should not be confined by financial barriers; scientific papers and associated communication should be legally available to everyone.

7. Science is an infinite process of discovery and discovered laws and attempts should be made to explain such discoveries using new theories.
8. Scientific theories are vulnerable to errors despite being numerously tested for errors.

The selected points from the aforementioned Open Science Manifesto have two highlights: a) scientific knowledge should be advanced by constant checks for errors and testing of new ideas and approaches b) science should not be financially or legally bounded and everyone should have equal access to the scientific materials. These points set out the premise on which different bodies and organizations have created their own guiding principles for Open Science.

OSF Manifesto

Variants of Open Science Manifestos are available in the literature as different research bodies opt for the one most suitable for their respective field. For this book, we put forward "A Manifesto for Reproducible Science" (Munafò et al., 2017) as a guide to put open science into practice to due it's popularity with 518 citations being reported at the time of writing and publication in *Nature Human Behavior*. The Manifesto is laid in an attempt to curb all the perceived threats to reproducible science as briefly discussed earlier in this section and to detail the necessary steps for practicing reproducible science (research design, data handling, collaboration etc.).

The following is an adopted Manifesto by Munafò et al. (2017).

METHODS - Measures to be Implemented in Research

1. Protection against Cognitive Biases

   Academics, just like any other human beings, are susceptible to cognitive biases. These biases can consciously or unconsciously lead to measures or actions being taken which any researcher would deem as malpractice. A feasible precaution is "blinding" of those who are involved in the study process. Blinding refers to concealment of experiment conditions or the testing hypotheses from data collectors, participants or the data analyst. This makes up a good protection against any sort of cognitive biases as those involved are unaware of what to expect from the findings or even what is being tested.

2. Improvements in Methodological Trainings

   The world wants evidence and the evidence lies in data. Every research comprises of data analysis which provides the justification for the existence of the effect being studied. Therefore, if one wants to advocate for open science, he/she should start by fully comprehending a certain extent of knowledge of all the statistical tests and tools available to refrain himself/herself from common statistical misconception such as the interpretation of *p*-values. Realizing that not all academics are continuing education, the authors advised advocates of open science to design modules or easily accessible resources which researchers can refer to at their convenience.

3. Implementation of Independent Methodological Support

Similar to the nature of independent or external standing committees in public service sector, setting up of independent committees to provide assistance, guidance and monitoring of methodological designs and the study procedures is recommended as it ensures that no sponsor or any individual or entity associated with the study could influence or distort the findings obtained for self-benefit. An example of this is the CHDI Foundation which puts effort in the identification of steps and measures essential to the methods rigor of studies focusing on Huntigon's disease.

4. Support for Collaboration and Teamwork in Science

   Collaboration amongst researchers is beneficial to progression of our scientific knowledge as it's commonly noted that papers present findings with low statistical powers. Publication of such papers regardless of the effort and time devoted to the particular study becomes pointless if the findings obtained cannot represent the phenomenon in the majority of the population. This also creates room for false and inconclusive data to be pursued. This is often the case when researchers are confined by limitation of resources. Collaborative with other researchers add to the likelihood of obtaining findings with high statistical power and assures that any significant findings would be generalizable across the population if multidisciplinary perspectives are taken into account. One possible way to do so would be by promoting collaboration amongst students across institutes and open distribution of relevant resources

REPORTING & DISSEMINATION - Measures to be Implemented in Communicating Research

5. Promotion of Pre-Registration of Studies

   This measure was suggested for tackling the publication bias (fewer number of studies being published than the number of studies being actually conducted). It's likely that a study with positive and favorable results would be published while those which report negative or neutral findings are not. This leads researchers to highlight and put heavy emphasis on the factors reported as significant, which might heighten the chance of publication but would in fact be regarded as malpractice. Pre-registering studies comprises of registration of the basic study design with additional specifications of the procedures, statistical analysis plan and expected outcomes prior to data collection and analysis. By doing so, it becomes possible for others to evaluate the study design and try to act on it. It also counters the threats of *p*-hacking or any data tampering as all the integral details have been fully disclosed prior to data collection.

6. Improvements in Reporting Quality

   This principle puts weight on adhering to numerous guidelines and manuals that are easily accessible so as to ensure that the pre-registered information is presented clearly. Reproduction of the registered study would also be facilitated. Guidelines-stating bodies include *Transparency and Openness Promotion (TOP)*, *Consolidated Standards for Reporting Trials (CONSORT)*, *Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA)* *etc.* The need for such guidelines stems from the possibilities of unclear reporting. Unclear findings and ambiguous report would only add to the difficulty in carrying out the study since the purpose, exact measures or even the analysis tools and measures used would be unknown. This would be in complete contradiction with the very purpose that Open Science exists to serve. Adherence to these Reporting Guidelines facilitates detection of observed behaviors and their impact.

REPRODUCIBILITY - Measures to be Implemented in Support of Verification of the Report

7. Promotion of Transparency and Open Science

This principle is a direct application of Karl Popper's view of scientific knowledge as discussed previously. The authors posit a stance that scientific knowledge and its advancement is not exclusive to any individual or entity. Rather, it should be transparent and public. Another support comes from the fact that reviews from community, evaluation and extension designs are important for the advancement of scientific knowledge. Also, by practicing transparency and open science, the researchers can gain greater approval and credibility for their work as their work is easily verifiable. One can simply put transparency and open science in practice by making the contents and procedures opted in the course of the study accessible and viewable by others without any discrepancies.

EVALUATION - Measures to be Implemented in Evaluation of Research

8. Diversification of Peer Review

With the integration of technology in distributing and publishing papers, the journal editor's role of monitoring and reviewing the papers submitted for publication is gradually diminishing as journal is no longer the only means to get papers published. Therefore, little to no peer review is prevalent in some cases, resulting in no proper evaluation of the papers. Possible solutions have emerged with some portals such as *PubMed* allowing members of the public to freely comment on the work without restrictions. The authors of the manifesto highlighted two forms of reviews which can alleviate this problem: pre-publication review and post-publication review. Pre-publication reviews allow researchers to gain quick feedback on their proposed work; post-publication enables instant critique of paper in contrast to the slow bureaucratic-like procedures followed previously.

INCENTIVES - Rewards for the Novelty of the Idea

Success in academia is measured in terms of publication which compels researchers to distort their findings to comply with the commonly held belief or observation of only positive and statistical findings being reported. This creates the need to reconsider where should the academia place value when it comes to evaluating the researchers' performance and rewarding them appropriately. Therefore, it is suggested by the authors that academia should also recognize the novelty of the ideas or phenomenon being worked upon instead of solely focusing on the statistical findings. One such example is the badging system adopted by numerous journals which recognize the papers published in compliance with open science and transparency.

As mentioned earlier, the above stated Manifesto one of many established by different organizations and research bodies such as *The Hong Kong Manifesto for Assessing Researchers: Fostering Research Integrity* designed and presented at the 6th World Conference on Research Integrity (Moher, Bouter, Kleinert, Glasziou, & Sham, 2019) and *The Manchester Manifesto* adopted by the Institute for Science, Ethics and Innovation at The University of Manchester (The University of Manchester, n.d.). Regardless of the publication or the research body, they all resonate the very principles that composed the major component of this section.

We began this section by referring to Albert Einstein and we close this section by referring to Sir Isaac Newton.

" If I have seen further it is by standing on the shoulders of Giants." ~ Sir Isaac Newton

Conclusion

To conclude, this chapter has discussed how the replication crisis in the field of psychological science gives rise to changes in academia. The acknowledgement of insufficient findings and under-representative samples has given rise to the popularity of crowdsourced research. Crowdsourced research projects enables the use of a large sample size for achieving high statistical power. Researchers and subjects worldwide can participate in a diversity of projects. Both the Psychological Science Accelerator (PSA) and the Collaborative Replications and Education Project (CREP) provide the hardware and software needed for conducting crowdsourced research.

The replication crisis has been increasingly incorporated into the teaching syllabus of many tertiary education institutions. Students are encouraged to gain hands-on experience of carrying out replication projects. Not only can this help them understand about the modern replication crisis and the importance of reproducibility, but it can help them develop their own scientific methodology and the ability to evaluate the replicability of findings. The University of Hong Kong is one of the universities that embraces the education on replication crisis. Dr Gilad Feldman from the Psychology Department is one of the key figures who mobilize both post-graduates and undergraduates to conduct replication projects on previous scholarly studies about judgemental and decision-making (JDM).

Finally, this chapter discusses about the importance of embracing Open Science—an important solution to the corrupt practice of fabricating research findings and creating statistically significant results. The Open Science Manifesto consists of eight principles for researchers to make reference of when designing, reporting and evaluating their studies.

Quiz

(Unless otherwise specified, please select only one answer for each question.)

1) Which of the following is a feature of crowdsourced research?

    a) Diversity
    b) Transparency
    c) Generalizability
    d) All of the above

2) What is the name of the system that is a distributed network of laboratories formalized by crowdsource researchers?

    a) HKU mass replication
    b) Open Science Manifesto
    c) Psychological Science Accelerator (PSA)
    d) Open Science Framework (OSF)

3) Which of the following are disclosed before conducting a replication study using open science?
    a) Pre-registration
    b) Data
    c) Ethics approval
    d) All of the above

4) What are the two components of the gold standard? Please select two.

    a) Replication
    b) Sample size
    c) Data
    d) Reproducibility

5) The 6th World Conference on Research Integrity was held in HKU. Which of the following was not a key discussion issue?
    a) Data cleaning
    b) Exploratory data analysis
    c) Computer programming
    d) All of the above

6) Which aspect of Psychology does Dr. Feldman and his HKU students focus on when doing replication projects?
    a) Biological psychology
    b) Moral development
    c) Judgement and decision-making
    d) Psychotherapies

7) Which of the following has been adopted as a type of extension in HKU replication projects?
    a) Reaction time
    b) Additional individual difference scale
    c) Gender difference
    d) Age difference

8) Which of the following is not a key component of the Open Science Manifesto?
    a) Reproducibility
    b) Evaluation
    c) Reporting & Dissemination
    d) Emphasis on Significant Findings only

9) A number of manifestos are available online. What is the major difference amongst them?
    a) Not all address the same research domain or research area
    b) Not all adhere to transparency
    c) Not all emphasize on clarity
    d) Not all emphasize on peer-reviewing

10) There are many variants of Open Science Manifesto. In this chapter, we have outlined several feasible steps for practicing reproducible science. Which of the steps is not mentioned as our suggestion for Open Science Manifesto? (Refer to the section OSF Manifesto)

    a) Methods
    b) Reporting & Dissemination
    c) Subjective Data Analysis
    d) Incentives

References

6th World Conference on Research Integrity. (July 3, 2019). WCRI2019. Retrieved November 10, 2019, from

https://wcri2019.org/

Carsey, T. M. (2014). Making DA-RT a Reality. PS, Political Science & Politics, 47(1), 72–77.

doi: 10.1017/s1049096513001753

Cohen, J. (1994). The earth is round (p < .05). American Psychologist, 49, 997.

doi: 10.1017/s1049096513001753

Feldman, G. (2019). Pre-registered replications: Background. Retrieved November 10, 2019, from

http://mgto.org/pre-registered-replications/

Feldman, G. (2019). Pre-registered replications: Other Registered Report initiatives. Retrieved November 10,

2019, from http://mgto.org/pre-registered-replications/

Feldman, G. (2019). Workshops. Retrieved November 10, 2019, from

http://mgto.org/pre-registered-replications/

Frank, M. C., & Saxe, R. (2012). Teaching replication. Perspectives on Psychological Science, 7(6), 600-604.

doi: 10.1177/1745691612460686

Hawkins, R. X., Smith, E. N., Au, C., Arias, J. M., Catapano, R., Hermann, E., ... & Salehi, S.

(2018). Improving the replicability of psychological science through pedagogy. *Advances*

in Methods and Practices in Psychological Science, 1(1), 7-18. doi: 10.1177/2515245917740427

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The Extent and Consequences

of P-Hacking in Science. PLOS Biology, 13(3). doi: 10.1371/journal.pbio.1002106

Moher, D., Bouter, L., Kleinert, S., Glasziou, P., & Sham, M. H. (2019). The Hong Kong Manifesto for Assessing Researchers: Fostering Research Integrity. Retrieved November 10, 2019, from https://wcri2019.org/uploads/files/2019_new/Hong_Kong_Manifesto_0527.pdf

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., ... &

Castille, C. M. (2018). The Psychological Science Accelerator: Advancing psychology

through a distributed collaborative network. *Advances in Methods and Practices in*

*Psychological Scienc*e, 1(4), 501-515. doi: 10.1177/2515245918797607

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. D.,

… Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human*

Behaviour, *1*. doi: 10.1038/s41562-016-0021

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

Sharpe, D. (2004). Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research.
Canadian Psychology/Psychologie Canadienne, 45(4), 317–319. doi: 10.1037/h0087004

The Editors of the Open Science Repository (2012). Scientific Manifesto Based on the Philosophy of Science of Karl Popper. *Open Science Repository Philosophy, 1*, 09, OSR0000071.

The University of Manchester. (n.d.). *The Manchester Manifesto*. Manchester. Retrieved November 10, 2019,

from http://www.isei.manchester.ac.uk/TheManchesterManifesto.pdf

Wagge, J. R., Brandt, M. J., Lazarevic, L. B., Legate, N., Christopherson, C., Wiggins, B., &

Grahe, J. E. (2019). Publishing research with undergraduate students via replication

work: The collaborative replications and education project. Frontiers in psychology, 10.
doi: 10.3389/fpsyg.2019.00247

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations.

American Psychologist, 54, 594–604. doi: 10.1037//0003-066x.54.8.594

11 - Where are we headed? What is the future?

Team names and contribution

First name	Last name	Researchgate profile	OSF profile	Institutional email	Personal email	Student ID
Qing Chun Tina	Zhu	https://www.researchgate.net/profile/Tina_Zhu9	osf.io/7wzdk	tinaqcz@connect.hku.hk	tinaqcz0504@gmail.com	3035341805
Valborg	Johannesen	https://www.researchgate.net/profile/Valborg_Johannesen	osf.io/cjs6z	u3565178@hku.hk	vajo94@hotmail.com	3035651781
Kar Man Carmen	Lei	https://www.researchgate.net/profile/Carmen_Lei	osf.io/xryvh	carmen22@connect.hku.hk	carmenleikm22@gmail.com	3035553367
Wing Yan Gloria	Chan	https://www.researchgate.net/profile/Wing_Yan_Gloria_Chan	osf.io/vk5t8/	cwysmile@connect.hku.hk	cwysmile@gmail.com	3035573018
Martina	Simovicova	https://www.researchgate.net/profile/Martina_Simovicova	osf.io/wpv9z	u3565326@connect.hku.hk	mata.simovicova@gmail.com	3035653260
Oliver	Read	https://www.researchgate.net/profile/Oliver_Read	osf.io/kwvht	or2g17@connect.hku.hk	or2g17@soton.ac.uk	3035650828

Managerial summary

The present chapter is about examines the issue of replicability (also known as replicability crisis), data reproducibility and generalizability of research findings within the field of social psychology. Firstly, we are addressing the distinction between replicability and reproducibility of the research. We then choose to focus on the replication crisis, identifying the problem from the approach of researchers, research system, and research topics. Recommendations are given on resolving the current threats for reproducible science, on the design of replicable research, and education for novice researchers that can be implemented for future studies. Out of the given recommendations, designing statistically high powered research with “non-questionable” research questions, avoiding publication bias by having the standardized research, and having an open collaborative research environment are the key points that shall be implemented in order to improve the present situation. In addition, other than increase the dependability of the current research by improving current research system or teaching on service researchers, we justify that current psychology students could work with replication of contemporary, cutting edge study for training of their research skills and critical thinking, in addition, replications would help on the original work that they replicated on.

In depth report

Over the past few years, psychology researchers have become more aware and concerned as to whether findings from psychological studies are replicable or reproducible. Such concern is still on the rise due to the constant discovery of relatively weak rate of replicability of published literature (Asendorpf et al., 2013). This ongoing difficulty or impossibility to replicate or reproduce scientific studies is known as the replication crisis. This chapter will focus on the future direction of tackling the replicability crisis.

Now that you have come to the end of the coursebook, we would like you to take a minute to think about what constitutes a good research paper and how we could sort out the criteria for only allowing quality research to be published. According to Asendorpf et al. (2013), standard of quality is defined with reference to internal and external validity of the research as well as research that is based on the reputation of the journal in which it was published. As with the latter, it is to note that there are twofold to judging the quality of a research based on the reputation of the journal. On one hand, a better journal would compare different research papers before favouring the one with the highest scientific validity to publish, hence correlating such elimination to the positive quality of the research. However, on the other hand, such competition during the publication process may promote the need of "perfect data", which thereby encourages questionable practices. For example, the exclusion of data that are inconsistent with the hypothesis or only reporting selective studies that worked and excluding those that did not or weren't significant. Furthermore, questionable practices include deciding whether to collect more data after looking to see whether the results will be significant.

Thus, we are left to think about what we could do to improve not only the internal and external validity of future research, but also ensuring that all published research will not undermine the knowledge and reputation of the social psychology field.

Firstly, there are two major components which drive the replicability crisis; namely the importance of the *replicability* of research findings and the *reproducibility* of findings obtained from the same data set. Scholars have defined Replicability "as the ability of random samples (like individuals or groups, situations that are natural or experimental, and operationalizations (experimental manipulations, methods, and measures)) to capture the important facets of the research design" (Asendorpf et al., 2013). Replication is achieved if there is minimal differences in the facets of the design, between the finding in the original study A and the findings in replication studies B. In order to achieve this replication, reproducibility of a research finding from the same data set is a necessary requirement (Asendorpf et al., 2013). Asendorpf et al (2013) states how data reproducibility places more emphasis on Researcher B obtaining exactly the same results as originally reported by Researcher A when following the same methodology. To ensure that the reproducibility is valid, Researcher B must have the raw data, the code book (variable names, labels, values and codes for missing data) and acknowledge the analysis previously done by Researcher A. Hence, the key point is that if data reproducibility is not done correctly, biased or contains errors, replicability would be limited and flawed as well.

In this chapter, we will delve deeper into the concepts of reproducibility and replicability as well as exploring the solutions and improvements that future social psychology researchers, reviewers, editors, journal management, teachers and granting institutions etc. should take into consideration in an attempt to decrease the replication crisis.

Reproducibility Crisis

Reproducibility in social psychology, is the ability to reproduce the data and results of published research. However, there are threats to reproducible science; namely the lack of replications, hypothesizing after obtaining the results, poor study design, low statistical power, analytical flexibility of data and test hypothesis, publication bias and lack of data sharing (Munafo et al., 2017). During data analysis and interpretation, researchers often fail to recognise that they may be subjected to confirmation and hindsight biases which encourage the favoritism of an outcome that they desire, and rejection of an outcome even if the results show otherwise. If uncontrolled and unaware, it would dramatically increase false discovery rate and in fact encourage self deception.

Munafo et al. (2017) proposed four evidence-based set of actions that can be implemented by researchers, institutions, journals and funders as solutions and improvements to reduce the threats of reproducible science; they are methods, reporting and dissemination, reproducibility, evaluation and incentives.

Methods

In this section, we would focus on what measures could be implemented when conducting research (like the study design, methods, statistics). Blinding of participants and data collectors to the experimental conditions that participants have been categorised into and to the research hypothesis can prevent cognitive biases of researchers, participants and data analyses (Munafo et al., 2017). This could be done during data preparation, where variable labels can be masked, hence the results cannot be interpretable in terms of the research hypothesis (Munafo et al., 2017). Furthermore, having a pre-registration of the study design (whereby researchers would publish their research plan and it will undergo a peer review in advance of observing the research outcomes) is a good form of blinding as the data do not exist yet and the outcomes are still unknown.

Secondly, the methodology of research could be improved in terms of statistical training. Munafo et al. (2017) note that research design and statistical analysis are mutually dependent and interpretation of *P* values, limitations of null-hypothesis significance testing, accuracy of reported effect size and importance of statistical power are often incorrectly understood, which ultimately leads to flaws within the data of the research published - limiting its reproducibility in the future. A solution to this would be to develop easy-to-understand *educational resources* specifically for statistical interpretation of research data and to also customised it for particular research applications. Munafo et al. (2017) provided some current examples like the “Experimental Design Assistant” that supports animal experimental designs, while “P-hacker tackles ‘P-hacking’ (a concept when researchers selectively report or misreports the true effect sizes in their published studies - like trying out several statistical analyses then selectively reporting those that produce significant results) and creates statistically significant findings by using analytic flexibility.

Finally, it is inevitable that sometimes statistical power of a research study may be low (due to sample size, effect size, significance level, and the power of the statistic used) and this increases the likelihood of obtaining both false-positive and false-negative results. Hence, in order to prevent such low-powered research to persist (perhaps due to the poor understanding of low power or lack of resources to improve power), *team science* could be implemented. Instead of basing the study on limited resources conducted by single investigators or just one laboratory or experimental study, distributed collaboration globally across many study sites will not only encourage high-powered designs, but will also increase the study’s generalisability across different settings and populations - touching upon different cultures and theories.

Reporting and Dissemination

In this section, we will focus on what could be improved in regards to reporting standard and study pre-registration. Publication bias and analytical flexibility (in terms of outcome switching) are common problems in research practices. Publication bias is the publication of studies that obtain positive results rather than studies that have obtained negative results. This means that many more studies are being conducted but are selectively published favouring studies with significant findings, while null findings are ignored (Munafo et al., 2017). Outcome switching refers to the possibility of changing the outcome of interest depending on the observed results (Munafo et al., 2017). However, *pre-registration* can protect against these problems.

Not only does pre-registration mean that the study’s results will be made public, it also requires researchers to clearly state the study design, primary outcome and analysis plan before conducting the research (Munafo et al., 2017). Pre-registration of a study will decrease publication bias as it allows all research (be it a positive outcome or a negative outcome) to be published. Furthermore, pre-registration of a study would allow the discovery of more null findings in RRs and pre registrations than in the rest of the literature. To test this, Allen and Mehler (2019) assessed the percentage of hypotheses that were not supported and compared it with percentages previously reported within the wider literature. Of the hypotheses they surveyed, 60.5% were not supported by the experimental data (null findings), which contrasts to the estimated 5% to 20% of null findings in the traditional literature.

Reproducibility

In this section, we will focus on what can be implemented to support verification of research, mainly sharing the data and methods. Nowadays, science lacks ‘openness’ where most published articles are not readily available and accessible unless people have personal or institutional subscriptions. One solution could be to encourage open science and transparency - the process of making content and producing evidence and claims transparent and accessible to others (Munafo et al., 2017). For example, researchers could adopt such an open practice where their publications could be accessible, open to peer-review and verification of their findings through the Open Science Framework (OSF) platform. Indeed, there have been concerns of copyright issues of the research article and this has sparked an ongoing debate between publishers and scholars. In short, in the publisher's perspective, they would try to limit public access to the research article in order to safeguard their interest. However, in a scholar's perspective, 'Scholarly communication' is seen as a crucial part as scholars are often judged by the public on their academic output and list of publications. Moreover, open access journals has aided this process by providing a means for scholars to publish their research regardless of perceived importance, as is the case with traditional journals.

Evaluation and incentives

With the progress in technological advancements, there is an increasing ease of dissemination of research information to the research community. It could be argued that it facilitates easy sharing and discovery of research, but on the other hand, it means that publishers do not get the final say in the research's impact in the community. Nevertheless, the role of publishers in evaluating still remains crucial in the research enterprise. If researchers adopt the *pre- and post- publication peer review* mechanism, it would dramatically improve and expand the evaluation process. According to Munafo et al. (2017), researchers sharing their preprints will enable a rapid feedback on their work from diverse communities rather than the conventional closed peer review process. Employing post-publication services would allow reviewers from diverse qualified institutions to make critical comments. Indeed, this pre and post processes of peer-reviewing will allow the community to justify collectively (given that the research is open) the suitability, generalisability, and reproducibility of a research study.

Finally, researchers, institutions, journals, funders, authors, and reviewers all contribute to the cultural norms that create and sustain dysfunctional incentives (Munafo et al., 2017). Changing the incentives could alter the problem and the reward structure too. For example, providing incentives favouring transparent and reproducible science shall encourage more researchers to conduct their research in line with the open science guidelines. Journals could also adopt badges to acknowledge practices, to promote registered reports and blinding, guidelines to promote openness and transparency, and having open-science practices in institutions (Munafo et al., 2017). As time passes, collective efforts of using incentives can shift science to more transparent, reproducible and credible.

Replication Crisis

Replication crisis can be explained in three ways focusing on researchers, research system and the research topic respectively (De Boeck & Jeon, 2018). Therefore, some suggestions to improve these three problems were introduced.

Proximal problem explanations focuses on problems in research practices. Selection of intriguing and novel research questions, problematic practices and weak inferences are factors that contribute to replication crisis in terms of research practices. Novel research questions often lack replication. Also, novelty is not a proper indicator of the quality of research. P hacking and file-drawing are some examples of problematic selective practices that result in replication failures.Countering selectivity can improve the inference method for proximal problems (De Boeck & Jeon, 2018).

Distal problem explanations refers to the context of research and focus on the research system. It is common for journals to evaluate research on its results instead of focusing on study design. Such practice intensifies publication bias as it encourages research to produce significant results to be published. Reducing the bias for significant results and new discoveries in research and modification of related policies in funding agencies can be possible ways to deal with distal problems.

Another explanation focuses on the topic researchers investigate. The problem can be caused by two factors including false discoveries due to high probability of the null hypothesis and variable effects. Variable effects can be a result from the variations in terms of the context and methodologies used in different researches. Heterogeneity of effects found in research is another significant cause of replication failures. Performing a false discovery rate analysis, which is similar to a power analysis and meta study, can be solutions to problems regarding behavioral reality.

There are eight standards and three levels of the guidelines of The Transparency and Openness Promotion (TOP) (Nosek et al., 2015). These standards include standards on citation and replication in open science practices. It emphasizes design, research materials, data sharing and analytic methods on transparency. Last but not least, standards on pre-registration and pre-registration analytic plans are regulated .

Recommendations for Research Design (for replication research)

The following recommendations could be implemented while designing a research in order to improve the dependability of the study: 1) improvements to make for statistical primer, for example, avoiding an underpowered study; 2) use “appropriate” hypothesis, because questionable research would cause questionable conclusions; 3) choose the appropriate reporting method, be aware of publication bias; 4) promote collaborative work, enhance the openness about data and methods for replication; 5) solutions to improve credibility and efficiency, understand the motives of operation for current science; 6) incentives to encourage scientific integrity, evaluate the current disadvantages in research practices on an open and flexible basis, and give suggestions for alternative system.

1. Statistical primer

Unlike other scientific research areas, psychology is based on the research on human being, typically for social and personality psychology. Researchers cannot control many aspects that vary in each individual. A finding based on one sample of 100 participants may be very different from the result of another sample of 100 participants, since all participants are genetically diverse (Funder et al., 2014). These variations across samples raise the importance and needs for statistics in research. With the appropriate usage of different statistical primers, the measurement of one sample would be able to tell the finding as if it is done to the entire population. Such statistical inference would have significant implications on the operation, analysis, and reporting of research (Funder et al., 2014). Some statistical primers that should be considered in psychology research include: effect size, statistical power, sample size, p value, and sources of error (i.e. type I error).

Decreasing the sources of error would increase replicability of a research study (Asendorpf et al., 2013). Firstly, there are two types of errors – type I and type II; type I errors , occur when the true null hypothesis is rejected, thus, the findings that are said to be significant may actually just occur by chance; type II errors , occur when the false null hypothesis is accepted, therefore, research may fail to report the significant effect (McLeod, 2019). Power can be used to better understand the errors, which refer to the probability of detecting an effect when there is one, on another word rejecting the null hypothesis if it is false. P value depends on sample size, effect size (Asendorpf et al., 2013; Funder et al., 2014). Because of the negative correlation between the two errors, when planning the study a choice should be made as which error to minimize, yet, increasing statistical power can minimize both types (Asendorpf et al., 2014). Thus, if decreasing the sources of error would help with replicability, having sufficient power could increase the quality of the study. In addition to the two errors related to null hypothesis, inaccuracy of parameter estimation is the third error to consider when designing a study. Smaller confidence intervals (CI) in the initial study are flavoured for better implement replication of the study (Asendorpf et al., 2014). Increasing the sample size would increase the statistical power while decreasing the CI, thus, the results obtained from a larger sample would be more replicable than those from a smaller sample (Asendorpf et al., 2014). Effect size can affect the strength of the experimental manipulation or the reliability and precision of measurement (Funder et al., 2014). Higher reliability of the measurements would increase replicability (Asendorpf et al., 2014). In addition, presenting multiple studies is a way to increase replicability. If multiple underpowered studies get the same results then people may consider it as replicable as well; however the truth is the opposite; thus, for future improvements, researchers should avoid multiple underpowered studies (Asendorpf et al., 2013).

2. Avoid “questionable” research

For the previous recommendations on statistical primer, an assumption is made of having “appropriate” hypotheses, and the findings are being “fully” presented. However, many practices may be at the origin of false positives in scientific literature (Świątkowski & Dompnier, 2017). Questionable research practices include a) fail to have statistical correction when conducting multi testing on a data set; b) data peeking, keep on running test with participants until significant result; c) dropping “insignificant” measures, findings, observations, etc. after knowing the outcomes; d) only reporting the “good” experiment results after multiple experiments (Funder et al., 2014). Questionable research would lead to questionable results, thus, sometimes even well-designed replication studies would fail to support the original findings (Funder et al., 2014). In order to help with future replication of the study, the research and analytic process should be fully reported and described.

3. Reporting Method

Gerber and Malhotra (2008) offer three potential ways to tackle publication bias, regarding journal decisions, submission practices of scholars, and research practices.

Firstly, they hypothesized that reviewers, especially those working for top journals, evaluate research studies and selectively choose those with statistical significance (Gerber & Malhotra, 2008). To tackle this issue, reviewers should be motivated to focus on evaluating research designs, ideas, operationalization, method of analysis, choice of sample instead of judging studies by the estimated significance of individual study.

Secondly, regarding submission practices of scholars, critical values for statistical significance is often found to be inaccurate in published studies. For instance, the probability of having a Type error is often underreported in many published studies. Obviously, it increases unauthentic confidence levels and has a negative effect on those follow-up studies. In addition, problematic journal practices as mentioned above play a role in this case. They mainly intensify the distortion of statistical calculation in research. Data mining and adjustments on post hoc sample sizes are some cases in point, moreover, the standard of $p < 0.05$ is not helpful to encourage accurate estimates in research. A possible solution can be producing a collection of many individual studies that have insignificant statistical results as long as the measurements are accurate and without bias. In this way, it can encourage researchers to focus on the study design and worry less if they may not be able to produce very precise measurements.

Thirdly, regarding research practices of scholars, the research result of z-scores just over 2 is very common (Gerber & Malhotra, 2008). It can be an indicator for publication bias. For instance, it may reflect the research practices of pushing results above critical value. A possible solution is to encourage scholars to emphasize the accuracy of evaluations, especially on confidence intervals and the practical implications of the results.

4. Collaborative work

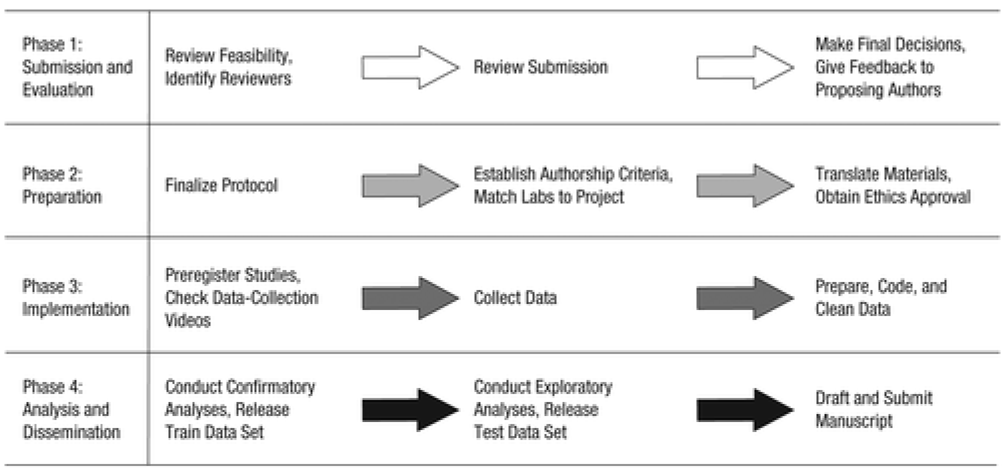
Collaborative work which is aligned with data sharing policy can be a possible solution to the current issue of replication crisis Psychological Science Accelerator (PSA) is an example of a distributed collaborative network to enable and support crowdsourced research projects (Moshontz et al., 2018). Its background, structure, principles, procedures, benefits, and challenges will be briefly discussed as follows.

The PSA is introduced to address the issues in the field of psychology including failures of replication of many studies. As there was a reform to improve the generalizability, crowdsourced research was introduced to tackle the issues. For instance, it can increase sample size and thus has high statistical power, as well as contributing to higher generalizability of research, aligning with open science practices. Another benefit is that it promotes interactions and diversity among researchers. Hence, it offers more opportunities to researchers facing different difficulties in research. However, it is difficult to recruit and maintain large collaborative network. For this reason, PSA was created to increase diversity and effectiveness of collaboration network.

In terms of structure and principles of PSA, it consists of five principles including diversity and inclusion, decentralized authority, transparency, rigor, and openness to criticism. Regarding the principle of decentralized authority, it refers to the collective effort of members to direct PSA without a central leader in it. Regarding the principle of rigor, it encourages large samples, reviews and analysis with high quality.

In terms of procedures of PSA, there are four phases in total, including submission and evaluation, preparation, implementation, and analysis and dissemination. The following Figure 15 illustrates the procedures of PSA in different phases.

Figure 15 Four Phases of Psychological Science Accelerator



(Moshontz et al., 2018, p.505)

In terms of the benefits of PSA, it is more cost-effective and efficient compared to other crowdfunded research as it significantly reduces obstacles for entry to crowdfunded research. Also, it adds more diversity in sample in research. Furthermore, it aligns with open science practice with a high degree of transparency, and reduces the bias of individuals due to the characteristics of its procedures. To sum up, PSA offers training opportunities for young researchers, promote collaborations and benefit research development in the future.

In terms of challenges of PSA, it uses resources from many different sources from institutions which increases its opportunity cost. PSA also faces challenges in terms of difference in cultures and languages. Protecting the rights of subjects is another issue that needs to be addressed. Another limitation is shown as PSA may not support studies that require more funding. More importantly, it is very challenging for so many stakeholders involved in PSA to collaborate and cooperate together.

5. Solutions to improve credibility and efficiency

Solutions to improve credibility and efficiency can be mainly divided in three aspects: interventions to improve research practices, understanding the motives of various stakeholders who operate in scientific research, and modifications in the reward system for science/research (Ioannidis, 2014).

In terms of interventions to improve research practices, it is crucial to assess the effectiveness and feasibility of the intervention before putting it into practice. For instance, it is impossible to reject every study that involves some insignificant results. It is argued that there are many pre-existing problems in the research structure of various fields of science, such as bureaucracy. Large-scale collaborative research and encouraging replication are some of the examples.

In order to understand the motives of various stakeholders who operate in scientific research, it is noteworthy that researchers have their personal interest in terms of benefits of publishable study. Apart from influence from private investors, it is clear that the increase of English research studies produced in non-English speaking countries complicate the situation as more stakeholders are involved.

In terms of modifications in the reward system for research, changing the rewards for publications and promotion in academic fields can be some possible solutions. It is important to tackle the issue of low credibility of research by addressing the motives of researchers and other stakeholders. Values of study ideas instead of publications should be rewarded. Moreover, studies that provide practical values in real life and those with high quality peer review should be valued and encouraged.

6. Incentives to encourage scientific Integrity

To provide incentives to encourage scientific integrity, disadvantages of current system on research practice will be addressed first, followed by suggestions encouraging ethical research practice.

There are several disadvantages of current system on research practice. One salient example is the provision of unreasonable incentives that intensify bias in research studies, such as putting emphasis on the number of publications of researchers (Edwards & Roy, 2016). In addition, the number of publications become an indicator of performance of scholars and benefit those who manipulate their statistical figures. Similarly, other factors including intense competition for funding in the academic field intensify the misconduct and unethical practices in research which is harmful to future scientific environment.

In order to alter the system to encourage ethical research practices, it is important for the problems that exist in the academic field to be properly addressed by different stakeholders. Also, another suggestion is to commission scholars with expertise to review the funding in their academic field (Edwards & Roy, 2016). This helps to prioritize the interest of the public instead of the interest of private investors when the scholars conduct research. It also minimize corrupted practice in academic research and encourage the promotion of ethical researchers. Moreover, it is crucial to understand and address the external factors that may intensify the unethical practices in research. For example, educating students who might be future researchers to tackle these issues is a practical and feasible way to improve the academic environment in the future. Likewise, reinforcing values that train ethical researchers can be another strategy that can be introduced in education. Universities should modify their policies so that it encourages ethical practices and punish misconduct, rather than having incentives that may in turn intensify the problematic research practices.

Recommendation for Education of Replication Research Practice

Recommendations for research practices can increase the quality of the published research, however, it is important to educate those who in the publication process about the value of replication research (Funder et al., 2014). These recommendations include encouragement the culture of "getting it right", teaching about the transparency of data, improving methodology instructions for scientist working (including editors, reviewers, etc.) in the research field, as well as educating students knowing the importance of reporting standards. Thus, knowing how to teach reproducible and transparent research is important for building a better future for scientific research, since all those students force can help to make real scientific contributions.

1. Encourage a culture of "getting it right"

Educational process for better research practices should begin with the promotion of building a "getting it right" culture rather than "finding significant results" (Funder et al., 2014). The "getting it right" culture can be implemented in both the classroom (school courses/textbooks) or even in editorial/reviewer guidelines, then, both experienced and novice researchers should be reminded to carefully evaluate the research design, choice of sample, statistical analysis, and data collection in establishing a valuable research (Funder et al., 2014). Soundness of the research should be valued over publishability, and this standard of good practice should be taught by teachers in advance to support junior researchers and students to seek the goal of "getting it right" (Asendorpf et al., 2013, Funder et al., 2014).

2. Teaching the transparency of data reporting and the whole research transparency

Researchers like to "tell a good story" instead of "telling the whole story", tending to hide the imperfect results, and trying to adjust the hypotheses when needed to fit the results (Funder et al., 2014). Though omitting the unnecessary and unexpected data may help with the flow of the report, having the "whole story" available would help for further replication of the research (Funder et al., 2014). Other then promoting openness of data reporting, providing a comprehensive review (key prior studies), decisions for sample size (i.e. why choosing these many samples for the study), pre-registering research are important for open and transparent science (Asendorpf et al., 2013). Multiverse analysis can be a way to increase the transparency of research, where a whole data set can be processed using different reasonable choice of data process (Steegen et al., 2016).

3. Improve methodology instructions

Nowadays, statistics courses are rather important in undergraduate and graduate studies. No matter what the research topic is about, correlated to the recommendations for a better research, students or novel researchers should be aware of the consequences of using questionable research practices, the meaning and usefulness of effect sizes, CIs, and statistical power (Asendorpf et al., 2013). Studies with sufficient power, yet the results are nonsignificant should also be appreciated, as long as the research is conducted rigorously (Asendorpf et al., 2013). Besides the basic knowledge on statistics, students should also notice about the importance and challenges of replication research, understand how under the same design, with highly similar samples, outcomes maybe diverse (Asendorpf et al., 2013; Funder et al., 2014).

Teaching replication in classroom

Teaching replication methods for students in laboratory / psychological researching classes could be a win-win plan for both the scientific community and students' own knowledge gaining. Frank and Saxe (2012) argued that due to the time restriction of courses, in many cases, students are not able to perform a full original and scientifically meaningful experiment; thus, in order to help students learn the skills on research and scientific methods, they should be actively involved in replicating a recent cutting-edge experiments, and promote critical thinking about the problem (Asendorpf et al., 2013).

1. Students working on recent, cutting-edge experiments

Students in the laboratory class would be given an opportunity to make some real scientific contributions with appropriate supervision, where lessons may object on the scientific process, importance of reporting standards, and talk about the value of openness. Hence, replications of new research can be implemented as a part of experimental methods in students coursework (Frank & Saxe, 2012). Classic laboratory class for psychology is usually held in two ways: 1) verify an original study results by using the acquired skills; or 2) replicate an original experiment (Frank & Saxe, 2012). Version 1 can be dry and boring, whereas version 2 does not ensure the original experiment to be successfully replicated. Thus, in a pedagogical point of view, the laboratory class should be motivating and meaningful, which, both of the classic psychology laboratory classes fail to perform. Frank and Saxe (2012) suggested that students should learn the experimental methods in the same logic as it motivates working scientists. Working on the replication of a recent published experiments can have the following benefits: 1) original authors have done all the hard work on experiment questioning and designing; 2) recent article means the question refers to the interest of current working scientists; 3) the reliability of the results remain unknown (Frank & Saxe, 2012). While learning the skills of scientific methods, students' interests are promoted by working on replicating a recent research.

There are significantly more students than working scientists. Thus, for the scientific community, teaching replication in class would provide a large captive workforce, and the students' replications can be valid as long as the supervision of instructor is of a good standard (Frank & Saxe, 2012).

2. Critical thinking

Critical thinking is something discussed profoundly in the education field in the 21st century. It is thought to be one of the top of transferable skills that students need for life beyond the class (Lee, 2015). In general, critical thinking can be concluded as thinking that are clear, rational, logical, and independent (Lee, 2015). Professors should teach young researchers how to critically read and assess others' studies, to read scientific papers critically, see the advantages of research design, data interpretation, analysis, etc. (Asendorpf et al., 2013). Students should also be able to critically evaluate evidence, be aware of the interpretation of effect sizes, CIs, taught about the importance of meta-analysis at a more advanced level; shall be sensitive of statistical tools – i.e. optional stopping, data fishing, deletion of cases or outliers for arbitrary reasons, and other tricks used to reach significance (Asendorpf et al., 2013). Overall, critical thinking skills plays a significant role for current psychology researchers, and should be passed on to novice researchers.

Conclusion

Overall, though in the sections above we have effectively identified the problems that contribute to the replicability crisis, we have also recommended ways to reduce the reproducibility crisis. However, offering such solutions does not guarantee implementation of changes. Unless we make collective changes in the field of psychology, amending cultural norms and incentives that will spur behavioural changes when conducting and publishing research, the process of change will be slow and difficult. Furthermore, Munafo et al. (2017) remark cynically that some solutions may be ineffective or even harmful to the efficiency and reliability of science, though conceptually, they appear sensible. Proposed solutions may give rise to challenges such as the uncertainty about which studies deserve to be replicated and what would be the most efficient replication strategies. The current state of psychological science is not ideal, but transparent and open research practices contribute to the quality and reliability of science. However, if efforts are continuously recognized and it is done collectively in the field of psychology, self-examination and self-correction would increase, causing the drive for the evaluation and improvement of the scientific process itself.

Quiz

Write 10 easy quiz 4 multiple choice questions to ensure reading and understanding of this chapter:

1. What are the components which drive the replicability crisis?
    a. Replicability only
    b. Reproducibility only
    c. Replicability and reproducibility
    d. Generalizability

2. Which one is NOT a threat to reproducible science?
    a. Hypothesizing after obtaining results
    b. Publication bias
    c. Data sharing
    d. Low statistical power

3. In "reporting and dissemination", (one of the solutions to reduce the threat of reproducible science) what can researchers do to avoid the likelihood of reported results being of only positive outcomes while negative evidence get ignored?
    a. Share their data only to the publisher
    b. Convince the other party that their results are justified
    c. Conducting a pre-registration
    d. Not disclosing their study design only to the public.

4. According to The Transparency and Openness Promotion (TOP), how many standard and levels of guidelines are there on replication in open science practices?
    a. 5 standards and 2 levels
    b. 8 standards and 1 level
    c. 8 standards and 3 level
    d. 3 standards and 4 level

5. *p*-value can be used to understand errors (like the probability of rejecting the null hypothesis if its false). What does the *p*-value NOT depend on?
    a. Sample size
    b. Effect size
    c. Type II error rate
    d. Type I error rate

6. For which domains is replication crisis relevant?
    a. Research practice
    b. Research system
    c. Research topic

d. All of the above

7. Which one is NOT an aspect that contributes to publication bias?
    a. Journal decisions
    b. Submission practices of scholars
    c. Research practices of scholars
    d. Publishers decisions

8. Which of the following is one of the principles of The Psychological Science Accelerator ?
    a. Diversity and inclusion transparency
    b. Centralized authority
    c. Privacy
    d. None of the above

9. What are the possible ways to improve credibility and efficiency?
    a. Improve research practices
    b. Understanding the motives of various stakeholders who operate in scientific research
    c. Modifications in the reward system for science/research
    d. All of the above

10. What are some recommendations for Education of Replication Research Practice?
    a. Encourage culture of "getting it right"
    b. Encourage "getting significant results"
    c. Value publishability of research
    d. All of the above

References

Asendorpf, J.B., Conner, M., De Fruyt, F., De ouwer, J., Denissen, J.J.A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*, 108-119.

De Boeck, P., & Jeon, M. (2018). Perceived crisis and reforms: Issues, explanations, and remedies. *Psychological bulletin*, *144*(7), 757.

Edwards, M. A., & Roy, S. (2016). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*. DOI: 10.1089/ees.2016.0223 http://online.liebertpub.com/doi/pdfplus/10.1089/ees.2016.0223

Frank, M., & Saxe, R. (2012). Teaching Replication. Perspectives on Psychological Science, 7(6), 600-604.

Funder, D.C., Levine, J.M., Mackie, D.M., Morf, C.C., Sansone, C., Vazire, S., & West, S.G. (2014). Improving the dependability of research in personality and social psychology: Recommendations for research and educational practice. *Personality and Social Psychology Review*, *18*, 3-12.

Gerber, A., & Malhotra, N. (2008). “Do statistical reporting standards affect what is published? Publication bias in two leading political science journals”. *Quarterly Journal of Political Science*, 3, 313-326.

Ioannidis, J. P. (2014). How to make more published research true. *PLoS medicine*, 11(10), e1001747.

Lee, W. C. (2015). The Importance of Teaching Critical Thinking. Retrieved from: https://www.wabisabilearning.com/blog/the-importance-of-teaching-critical-thinking

McLeod, S. (2019). What are Type I and Type II Errors? Retrieved from: https://www.simplypsychology.org/type_I_and_type_II_errors.html

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., ... & Castille, C. M. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501-515.

Munafo, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. du, … & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour* , 1. doi: https://doi.org/10.1038/s41562-016-0021

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422-1425.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*, 702-712. doi:10.1177/1745691616658637

Świątkowski, W. &Dompnier, B. (2017). Replicability Crisis in Social Psychology: Looking at the Past to Find New Pathways for the Future. *International Review of Social Psychology* , *30*(1), 111–124, DOI: https://doi.org/10.5334/irsp.66

Additional Resources

Links

1. Open Science Teaching and Training Resources
2. Framework for Open and Reproducible Research Training (FORRT) Resource Submission (Responses)
3. Replication crisis on Wikipedia
4. Reproducibility and replicability reading list
5. Course Syllabi for Open and Reproducible Methods
6. Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J., ... & Schulte-Mecklenbeck, M. (2018). 7 Easy Steps to Open Science: An Annotated Reading List.
7. https://github.com/ReScience/ten-years/ (doing a replication study on your own data?)

Impactful papers:

1. Ioannidis, J. P. (2005). Why most published research findings are false. PLoS medicine, 2(8), e124.
2. Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22, 1359-1366.
3. DeCoster, J., Sparks, E. A., Sparks, J. C., Sparks, G. G., & Sparks, C. W. (2015). Opportunistic biases: Their origins, effects, and an integrated solution. American Psychologist, 70, 499.
4. Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: causes, consequences, and remedies. Psychological Methods, 9, 147.
5. Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories publication bias and psychological science's aversion to the null. Perspectives on Psychological Science, 7, 555–561.
6. Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. Psychological Inquiry, 23, 217-243.
7. Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. Science, 348, 1422-1425.
8. Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... & Laitin, D. (2014). Promoting transparency in social science research. Science, 343, 30-31.
9. LeBel, E. P., Campbell, L., & Loving, T. J. (2017). Benefits of open and high-powered research outweigh costs. Journal of Personality and Social Psychology, 113, 230.
10. Lishner, D. A. (2015). A concise set of core recommendations to improve the dependability of psychological research. Review of General Psychology, 19, 52.
11. Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1, 0021.

Review articles:

1. Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. Annual review of psychology, 69, 511-534.
2. Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. Annual review of clinical psychology, 15, 579-604.
3. Freese, J., & Peterson, D. (2017). Replication in social science. Annual Review of Sociology, 43, 147-165.
4. Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. Behavioral and Brain Sciences, 41.
5. The Replication Crisis: Flaws in Mainstream Science