SCIENTIST AS SUBJECT:

HOW RESEARCHER BEHAVIORS INFLUENCE PSYCHOLOGICAL KNOWLEDGE

By

NICHOLAS W. FOX

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Lee Jussim

And approved by

New Brunswick, New Jersey

May 2019

ABSTRACT OF THE DISSERTATION

Scientist as Subject:

How Researcher Behaviors Influence Psychological Knowledge

By NICHOLAS W. FOX

Dissertation Director:

Dr. Lee Jussim

Background:

Interacting with the published literature (“knowledge consumption”) and publishing new scientific findings (“knowledge production”) are two key moments in the scientist’s search for truth, and bias in either of these can distort what is known about an area of research. This dissertation details three studies conducted on researchers in psychology that together provide evidence of scientists’ behaviors influencing these key moments of knowledge production and knowledge consumption.

Methods:

Psychologists were recruited to participate in each study (N = 215 and N = 587). Studies used custom web tools and social network methods to collect unique datasets on psychologists’ social networks and how they approach the scientific literature. The analytic approach differed based on each study. For studies on knowledge consumption, Gini coefficients and measures of unpredictability were calculated to better understand the dynamics of the published literature. For studies on knowledge production, the generalized network scale up method was used to estimate the size of the population of current users of questionable research practices, and regression was used to better understand the relationship between attitudes and stigma against certain psychologists.

Results:

The presence of download counts (an operationalization of influential metadata) with scientific literature resulted in larger inequality of downloads, meaning potential readers were more likely to download articles that had been previously downloaded by others. Download count presence also resulted in a higher unpredictability of success. The proportion of psychologists who currently use questionable research practices was estimated as 18.18% by direct estimate and 24.4% by the social network scale up estimate. Finally, these researchers were found to be a stigmatized sub-population of psychologists, which could either help or hinder efforts to reduce this population size.

Conclusions:

There is evidence that psychologists may inadvertently bias the knowledge they generate and consume in several different ways. While this dissertation focused specifically on psychologists, there is potential for this work to be applied in other areas of scientific inquiry. These findings highlight the importance of understanding the scientist as a means of better understanding the science.

ACKNOWLEDGEMENTS

Firstly, I'd like to express my sincere thanks to my advisor Dr. Lee Jussim for his continuous support of my Ph.D research. His patience, guidance, and enthusiasm has helped immensely in the generation and writing of this dissertation. Even more importantly, Lee took a chance in mentoring me as I transferred into the Social Psychology concentration. I doubt I would have written this dissertation if it wasn't for this support, and for it I am extremely thankful.

Besides my advisor, I would like to thank the rest of my dissertation committee: Dr. Edward Selby, Dr. Richard Contrada, and Dr. Katie Corker. Together, they have provided insightful comments, encouragement, and thought-provoking questions that have increased the quality of my research.

It has taken a village of scientists to raise me to this point, and their support over the past years never went unnoticed. Dr. Anne Savitt (Stony Brook University), Dr. Felicia Goodrum (University of Arizona), Dr. Maggie So (University of Arizona), Dr. Janko Nikolich-Zugich (University of Arizona), Dr. Raphael Clynes (Columbia University), and Dr. Alexander Kusnecov (Rutgers University) – thank you all so much.

I'm very grateful for my fellow labmates Akeela Careem, Nate Honeycutt, and Dr. Rachel Rubinstein for all their help – most of which they gave me after I showed up in their office unannounced. Former lab members and successful scientists Dr. Sean Stevens and Dr. Stephanie Anglin have also been immensely helpful and supportive of my work – thank you!

I would not have been able to complete this dissertation without my family; my mom, Angela, for always being a positive and supportive force in my life; my parent-in-laws, Katia and Verne, and my brother-in-law, Nicholas, for their encouragement and support throughout my academic career; and my cat, Sir Neutron Ruffles Sherbet, who provided unwavering positivity and unending love.

Finally, and most of all, I'd like to thank my partner, Mariana. Ever since we met on our first day of college (in statistics class!), you have helped me become a better person, a better scientist, and a more critical thinker. You've supported me so much in the writing of this dissertation, and I am truly grateful for it. My appreciation for you is endless.

DEDICATION

For Mariana.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1:

Overview and Introduction

> *Historians don't want to write a history of historians. They are quite happy to plunge endlessly into limitless historical detail. But they themselves don't want to be counted as part of the limitless historical detail. They don't want to be part of the historical order. It's as if doctors didn't want to fall ill and die.*
>
> *Charles Péguy*, L'Argent, suite

French sociologist Pierre Bourdieu starts his 1988 book, *Homo Academicus*, with the above quote from poet and essayist Charles Péguy, commenting on the preference of the researcher to stand apart from what they research. Bourdieu's theoretical investigation of the academic world at a time of unrest and change (the 1968 University of Paris protests) presents to the reader the academy as an object for study. In doing so, Bourdieu brings the researcher into full display and argues that the authority and objectivity central to an academic's success is not inherent to the individual, but the result of the academic's position in the power structures of academia (Bourdieu, 1988).

In writing his book, Bourdieu aimed to "exoticize the domestic", asking scholars to critically engage with the academic world they inhabit and to question what drives their research questions, methods, and conclusions. Was objective curiosity the sole driver of inquiry, or was research shaped by the influences of academic power and conformity? In order to promote radical change in academic standards and research, Bourdieu asked his peers to honestly reflect on their position of power, the production and consumption of knowledge, and their role in the validation of that knowledge (Forte, 2015).

The beginning of the twenty-first century has been another time of unrest and change for science. For example, academic literature, which has been the foundation for the collection and dissemination of scientific knowledge, is in the process of shifting from the physical to the digital. While this may not sound like a noteworthy change, consider all the additional information that can be delivered along with a digital file (such as how many times it has been viewed or how many people have talked about it on social media platforms), as well as the ease of sharing one with others. For most scientists, academic literature is now easier to access and more daunting than ever, with literally thousands of new articles at their disposal every week (Larivière, Haustein, & Mongeon, 2015; Van Noorden, 2014).

Another example of this unrest is the increase in methodological critique. In psychology specifically, but in the whole of science more broadly, there has been an increased focus on the ways in which scientists do science. In 2005, John Ioannidis published a paper titled “Why Most Published Research Findings Are False”, which detailed how the combination of small sample sizes, high number of studies, and poorly defined *a priori* hypotheses can lead to almost any conclusion being supported by data (Ioannidis, 2005). In a now (in)famous example that illustrates this point, Daryl Bem published findings supporting the existence of precognition in the *Journal of Personality and Social Psychology* in 2011 (Bem, 2011). A published comment quickly pointed out the potential for flexibility in how the data from these studies were analyzed could have led to the conclusion that precognition exists without adequate supporting data (E. Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). At about the same time, several high profile scientific papers in psychology failed to replicate when performed by other research groups, raising doubt in the robustness of the published findings (Pashler, Coburn, & Harris, 2012; Ranehill et al., 2015). By 2019, at least two large-scale projects focused on measuring the replicability of psychology have concluded that much of the published literature is not robust to replication (Klein et al., 2018; Nosek, 2015). This has led to psychology finding itself in a "replication crisis" that continues to this day.

At this moment of change in science, it is important to again ask researchers to reflect on how their motivations and behaviors may knowingly, or unknowingly, contribute to the body of scientific knowledge. This dissertation details three studies conducted on researchers in psychology that together provide evidence of scientists' behaviors influencing the processes of knowledge production and consumption. As the "replication crisis" is centered on psychology, it was considered the most correct field to study in this way. This choice does not assume that other scientific fields are immune to the effects of researcher behaviors, but instead serves as a proof-of-concept that metascientific research can lead to valuable insights into how scientists within a field conduct research.

The following sections provide background information that is useful for better understanding of the context in which modern academic research takes place. While the storybook image of the objective, rational, noble scientist is pervasive (Veldkamp, Hartgerink, van Assen, & Wicherts, 2017), it is important to highlight that scientists are still human, and research is similar to any other career in that success is measured, at least in part, by productivity. Operationalizing productivity in research is an evolving topic, yet historically has been defined as producing high quality peer-reviewed research papers at a sufficient pace (Lee & Bozeman, 2005). Understanding how scientists navigate the world of career research will be important in understanding how their behaviors can influence psychological knowledge.

"Publish or Perish" and Publication Bias

One of the earliest references to the nearly ubiquitous academic phrase "publish or perish" came from the 1942 book *The Academic Man: A Study in the Sociology of a Profession* by sociologist Logan Wilson, who wrote, "the prevailing pragmatism forced upon the academic group is that one must write something and get it into print. Situational imperatives dictate a 'publish or perish' credo within the ranks" (Garfield, 1996; Wilson, 1942). As mentioned previously, publication record is a primary way to measure academic productivity. If a researcher's work never makes it to print, it is hard to justify their position in an academic institution. Unproductive scientists risk losing their jobs if they cannot turn their research into published literature (thus, the "perish" in "publish or perish").

A significant barrier to publishing the results of a scientific study is the preference of scientific journals to publish certain types of results. Journals would rather publish a study with a positive finding (i.e., golden retrievers are bigger than dalmatians) than a study with an ambiguous or negative finding (i.e., the data are inconclusive on whether golden retrievers and dalmatians differ in size). This preference to selectively publish studies with certain results and not others is called "publication bias" (Dwan et al., 2013), and has been considered a serious issue since at least 1963 (Newcombe, 1987).

The way most scientists determine if they have a positive finding (whether two groups of observations are different from one another) is by using statistical tests.

Typically, researchers collect data consisting of many observations, and then calculate if particular groups of observations differ “significantly”. To extend an earlier example, if a researcher was interested in whether two dog breeds differed in size, they would measure many individual animals of each breed, and then use statistics to determine if those two groups of observations were significantly different from one another. This determination relies on several key questions: how different is “significantly” different? Which statistical test should be used? How many individual dogs should be measured in each group? How does one measure the “size” of a dog? To better understand some of these questions, it is important to review the prevailing use of statistics in psychology, null hypothesis significance testing.

Null Hypothesis Significance Testing (NHST)

In psychology, many results are obtained by null hypothesis significance testing (NHST; Bakker, van Dijk, & Wicherts, 2012; Levine, Weber, Hullett, Park, & Lindsey, 2008). It is currently the most widely accepted and used approach to statistical decision making in the field. The modern NHST is a fusion of ideas from statisticians Ronald Fisher (R. A. Fisher, 1925), Jerzy Neyman, and Egon Pearson (Neyman & Pearson, 1933).

In Fisher’s approach (R. A. Fisher, 1925), researchers define a nil-null hypothesis ($H_{0,}$ where nothing happens – golden retrievers and dalmatians are not different in size) and test the probability of observing data under this nil-null hypothesis. Depending on the probability, the researcher then either rejects or fails to reject H_0. Fisher used an arbitrary cut-off probability of 5% to reject the nil-null hypothesis, meaning that under a true nil-null hypothesis, the researcher would only observe data that showed a difference

1 in 20 times. When one rejects the nil-null hypothesis, they are deciding that the nil condition (golden retrievers and dalmatians are not different in size) cannot be true given the observed data. He called this probability value the p value, which has become one of the main statistical justifications scientists use to make a scientific claim.

In the Neyman-Pearson approach (Neyman & Pearson, 1933), the researcher specifies two hypotheses, the null hypothesis (H_0) and the alternative hypothesis (H_1), along with their sampling distributions. The researcher must also specify the alpha level (α, the acceptable false-positive rate) *a priori.* This distinction allows for the measurement of Type I and Type II error, as well as statistical power (Hullett et al., 2008).

Type I error is incorrectly rejecting the null hypothesis when it is true. For example, suppose our true null hypothesis is “golden retrievers and dalmatians are not different sizes”. In this case, a Type I error would be deciding the two breeds are different sizes when they are not actually different. The is also known as a false-positive.

Type II error is incorrectly failing to reject the null hypothesis when the alternative hypothesis is true. For example, suppose our null hypothesis is “golden retrievers and dalmatians are not different sizes”. In this case, a Type II error would be deciding the two breeds are not different sizes when they are truely different. This is also known as a false-negative.

The alpha level (α) is a measure of how much appetite for risk a researcher has in making a Type I error. This false positive rate is traditionally set at 5%, or an alpha level of 0.05 (Lakens et al., 2018). This alpha level is reflected in the traditional cut-off value for a calculated p value of 0.05. This means that if a researcher calculates a p value of less than 0.05, they can decide to reject the null hypothesis (golden retrievers and dalmatians are not different in size) and only be making an incorrect decision once every twenty times (on average). If a researcher decides they are uncomfortable with a 5% false positive rate, they may set the alpha level they use to a lower value, like 0.01 or 0.001, representing a 1% and 0.1% Type I error rate, respectively. However, manipulating the alpha level influences statistical power.

Statistical power is defined as the probability of rejecting the null hypothesis when the specific alternative hypothesis is true. Defined in another way, this is the probability of detecting a difference, given a difference exists. As the alpha level decreases from 0.05 to 0.01, for example, it becomes more difficult to reject the null hypothesis, since making a Type I error is less palatable. Because it is more difficult to reject the null hypothesis, the probability of rejecting it when the alternative hypothesis is true (statistical power) also shrinks. This is important as power is an indication that one can detect a true positive finding given it exists. A study that has low power may inadvertently fail to reject the null hypothesis when the alternative is true, or in other words make a Type II error.

Publication Bias, Revisited

Previous work suggests published research is more likely to have statistically significant ($p < 0.05$) findings than unpublished research (Hopewell, Loudon, Clarke, Oxman, & Dickersin, 2009). This can be due to publishers delaying or refusing to publish studies with non-significant findings, or scientists failing to submit studies with non-significant findings to journals for peer-review, often referred to as the "file drawer effect" (since scientists know about publication bias, they leave these non-significant results in their file drawer, never to be published) (Rosenthal, 1979; Simonsohn, Nelson, & Simmons, 2013).

Knowing about the existence of publication bias has the potential to change how scientists approach their research. For example, it is possible that, in order to find a significant difference in size between golden retrievers and dalmatians, a researcher could measures "size" in several different ways. They could measure weight, height, length, food intake, buoyancy, and any other way they could conceptualize "size". They could then repeatedly statistically test these two groups until they achieve "significant" ($p < 0.05$) results worthy of publishing. This scenario of producing multiple comparisons to find one that stands a chance of being published increases the Type I Error rate, resulting in more false-positive findings being submitted for review and eventual publication in the literature.

It is more likely, however, that unconstrained plans for data analysis lead researchers to have multiple *potential* statistical comparisons, where the details of the data analysis are contingent on the content of the data, without the researcher having any conscious motivation to rig the outcome (Gelman & Loken, 2014). Scientists, like all humans, are biased towards confirmatory evidence (e.g., confirmation bias, or tending to seek information that supports their hypothesis about the world) (Nickerson, 1998). When data analysis plans are unconstrained, and the researchers are incentivized to find evidence (both due to publication bias and confirmation bias), different data outcomes may influence which types of analyses are performed. For example, suppose a research group was interested in gender differences in intelligence. If IQ scores were measured, and the averages between men and women looked different, researchers may perform a t-test comparing the mean IQ of each group. However, suppose the average IQ per group appeared similar. In this case, the researchers might instead conduct an F-test of equality of variances to identify whether IQ varies differently between men and women. Allowing the data to dictate analysis decisions in this way increases Type I Error and false-positive findings in the literature.

Many researchers are aware that multiple comparisons can lead to increased Type I Error and attempt to correct for this by using post-hoc procedures such as the Bonferroni correction. The standard procedure uses a modified significance criterion of α/k, where α is the significance level (typically set at 0.05) and k is the number of significance tests used (Cabin & Mitchell, 2000). For example, suppose for a control group and an experimental group, t-tests are performed between the two groups with five different variables of interest. The standard Bonferroni correction would reduce the significance level from the typical 0.05 to 0.01 instead.

This reduction in the significance level reduces the probability of incorrectly rejecting the null hypothesis, yet it also decreases statistical power. In the previous example, if each group had 30 participants, the power to detect a medium effect (Cohen's d = 0.5) would be 61%. After the standard Bonferroni correction, that power drops to 33%. This reduction in power translates into an increase in the Type II Error rate (β, calculated as 1 – power). In this way, researchers are now more likely to incorrectly accept the null hypothesis when it is actually false, making the finding less likely to be published and potentially contributing to the file drawer effect (Nakagawa, 2004).

Up until the 2000s, these issues of statistical power, multiple comparisons, false positive error rates, and the influence of publication bias went mostly uncriticized as part of the daily life of a career research psychologist. Publication of low-powered studies with potentially cherry-picked analyses were the norm, and careers advanced based on the fruit of scientific labor. Although there have been past crises of confidence in the validity of the published literature (Elms, 1975), the current "replication crisis" is unique as it comes at a time of technological advancement that may allow for the radical change in academic standards and research needed to return career research back to the objective search for truth.

The following three chapters detail studies performed on research psychologists for the purpose of better understanding the role of the researcher in the science they produce. The first study (Chapter 2), investigates how metadata associated with digital versions of academic papers may influence how scientists interact with the published literature. Chapter 3 details a study that sought to estimate the number of psychologists who use "questionable research practices" (QRPs), or research practices that may lead to increased Type I Errors and false positive findings in the literature. Finally, Chapter 4 describes a study on the social relationship between researchers who use QRPs and the general population of psychologists and asks if QRP users are a stigmatized subpopulation of psychologists. Chapter 5 is a general discussion of all research findings in the context of understanding how research behaviors influence psychological knowledge.

Chapter 2:

The Effects of Social Influence on Scientific Literature Choice Among Psychologists

The first academic journals (*Journal des Sçavans* and *Philosophical Transactions* of the Royal Society of London) were both published in 1665. For over 300 years, physical journals were the most efficient way for scientists to distribute their research findings to other scientists, journalists, and lay readers (Larivière et al., 2015). The late 1990s saw the beginning of the digital era of academic publishing, with more publishers producing digital versions of their printed journals, and some new journals being exclusively digital. Currently, science is in the unique in-between era where scientists utilize both print and digital journals to access and disseminate research findings. Importantly, many digital outlets for scientific articles can also present additional data to the would-be reader. This could include supplemental material from the authors, such as more detailed descriptions of methodological or analytical techniques used in the paper. Digital journal articles can also include "metadata", or data about the paper itself, which could include how many times it has been viewed or downloaded, the number of citations the paper has, or how many people are talking about it in blogs or on social media platforms (Piwowar, 2013).

In the process of doing science, it is necessary for researchers to stay current on the work being done by their peers. This is to help understand new findings in their own research, as well as to discover which scientific questions remain unanswered (Subramanyam, 2013). While the rate of scientific papers published has been increasing rapidly, the average number of articles read by scientists per month has remained unchanged since 2005 (Van Noorden, 2014), leading to an increased burden to find and read the most appropriate scientific papers for the amount of time available.

There have been a number of approaches adopted by scientists aimed at finding the most important papers to read for a given line of research. Before the digital era, when academic literature was only in print, researchers could utilize trained science librarians to help guide their reading (Chen, 1974). Now, some use digital services that rely on keywords or trends to alert the reader to new published research in an area of interest (i.e "e-mail me whenever a paper is published using the term 'implicit bias'" or "e-mail me papers others are talking about on Twitter"), or rely on behavioral strategies to search for the most relevant papers (Pain, 2016). All of these strategies may use article metadata as an input in the decision-making process to read a given research article.

A particularly important type of metadata are those that indicate the behavior of other scientists. The number of views, number of downloads, and number of citations an article has all provide a potential reader information on how other scientists have approached the article in the past. Previous research on the relationship between the download counts of biomedical papers and their subsequent citation count five years later found a moderately strong positive correlation (Pearson's $r = 0.5$) (Perneger, 2004), suggesting higher rates of downloading (and presumably reading) a paper leads to higher use in future scientific research. Perneger (2004) ends his article by stating: "Online readers judge the scientific value of an article from the title and the abstract, and if this assessment is favorable, they access the full paper". However, it is possible that these metadata constitute a form of social influence that acts on the potential reader, biasing their decision to engage with an article away from just the presented scientific value.

Solomon Asch's line-length experiments clearly demonstrate the majority effect on individuals and the willingness to submit to existing social pressure (Asch, 1951). Here, individuals had the tendency to yield to the judgement of the group, even when that judgement was objectively false. In a debrief interview, one of Asch's participants stated, "If I'd been the first, I probably would have responded differently", identifying the effect of having to judge the length of a line after seeing the judgement decisions made by a group of peers. Download counts on articles may generate the same biasing effect – a signal of judgement on an academic paper by a group of peers which gives disproportionate weight in favor of some papers over others, potentially (though not necessarily) independent of a paper's quality or content.

This idea has been tested more recently in both real and artificial cultural markets. In studying how cultural items such as songs, books, or movies become popular, sociologists have focused on markers that confer social preference. Muchnik, Aral, & Taylor (2013) experimentally manipulated up-voted and down-voted comments on a popular social news aggregation web site and tracked social herding over a 5-month period. Comments that were up-voted once by the researchers generated accumulated herding effects that increased comments' final ratings by 25% relative to control (unvoted) comments, suggesting a role for social influence in the accumulation of up-votes.

Salganik, Dodds, and Watts (2006) developed an digital music jukebox that let participants listen to and download different songs. They found that, when download counts for each song were displayed to participants, download inequality went up – meaning those songs that accumulated downloads continued to accumulate downloads.

This effect was enhanced when they increased the saliency of social influence by ordering the songs in descending order by download count. When the songs downloaded the most were early in the list of songs, inequality increased further.

Both the studies performed by Muchnik et al. (2013) and Salganik et al. (2006) demonstrate an effect of metadata on decision-making. This decision-making process is thought to have two-steps (Krumme, Cebrian, Pickard, & Pentland, 2012). First, the participant needs to choose to attend to an item, like to read a comment, listen to a song, or to read an academic paper's title and abstract. This first decision is made based on appeal, position, and any other potential social preference markers, like current download count, or number of up-votes. The second decision, in this case to download a song or paper, or to vote on a comment, is based on the perceived quality of the item. Social influence, in the form of metadata, acts on the first step of this two-step process, potentially by increasing appeal, by changing accessibility via list position, or by some other means (Krumme et al., 2012). See Figure 1 for an illustration of this model.

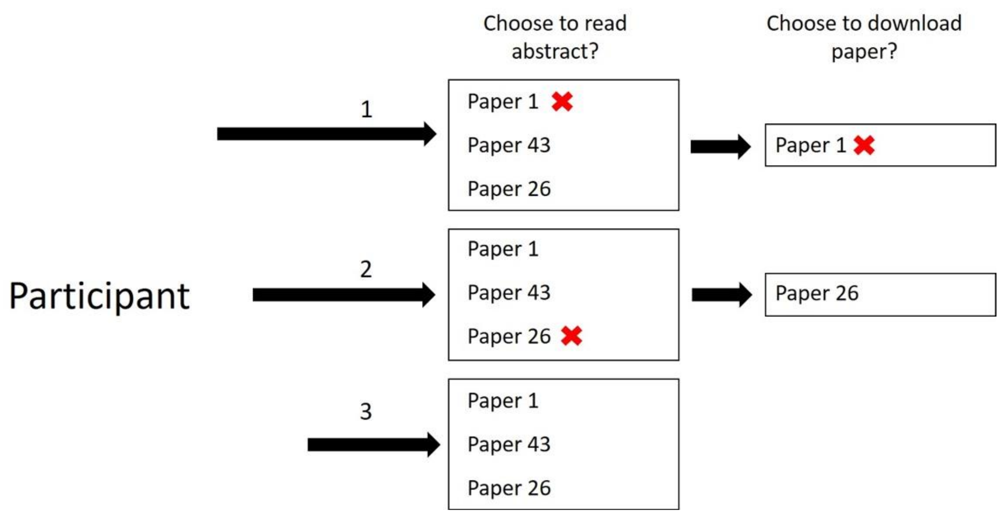


Figure 1. Deciding to download an academic paper is a two-step process. The first decision is to choose to read an abstract, which is affected by the paper title's appeal, location/accessibility, and markers conferring social influence. The second decision is to choose to download the full text paper, which is affected by perceived quality of the abstract. Illustration and explanation adapted from Krumme et al., 2012.

Metadata on journal articles is becoming increasingly common. It is easy to find the citation count for papers in popular academic article search engines such as Google Scholar and PubMed. Download counts have been adopted by article databases such as ScienceDirect and PsyArXiv, and altmetrics are becoming more common as scientists share their work via blogs and social media.

For this reason, it is important to determine how scientists choose which articles they will read. There is significant evidence that metadata that signals peer group judgement can result in herding effects, though this has not been demonstrated in a scientific community. It is possible that scientists, as a highly trained group of professionals, approach the scientific literature in a way that allows the highest quality research to become the most popular research. Alternatively, the behavior of some scientists may influence others to engage with the literature in a process independent of the quality of the published research.

The study reported in this chapter seeks to determine the effect of social influence on the decision to download scientific literature from an artificial academic market. I hypothesize that social influence will increase the inequality of downloads, meaning more downloads will accrue on a smaller number of articles. I also hypothesize that social influence will increase unpredictability of success, meaning across identical versions of the experimental academic market, different papers will become the most downloaded.

Methods

Web tool development and distribution. To collect data for this study, a bespoke web tool was developed in collaboration with Rutgers University Computer Science undergraduate student Steven Mattia. This web tool, called "AbstractFindr", was published online in August, 2018 at http://www.abstractfindr.com.

The main purpose of this tool was to present participants with the title and abstracts of academic papers. When participants clicked on a paper title, they were presented with the abstract of that paper, along with a place to rate the abstract on a one-to-five-star rating system. Once the participant rated the abstract, they were then asked if they would like to download the full article. If they clicked "yes", a pdf file of the full paper was downloaded to their computer, and then they were returned to the list of paper titles. If they clicked "no", they were directly returned to the list of paper titles. Participants had to both rate and make a download decision before they could choose another paper title. Participants were blocked from rereading an abstract after they had made a download decision, either "yes" to download or "no" to not download. See Figure 2 for a depiction of this process.

Forty-eight papers were used as stimuli in this study. This number of papers was reached based on previous social influence literature (Salganik et al., 2006), the desire to

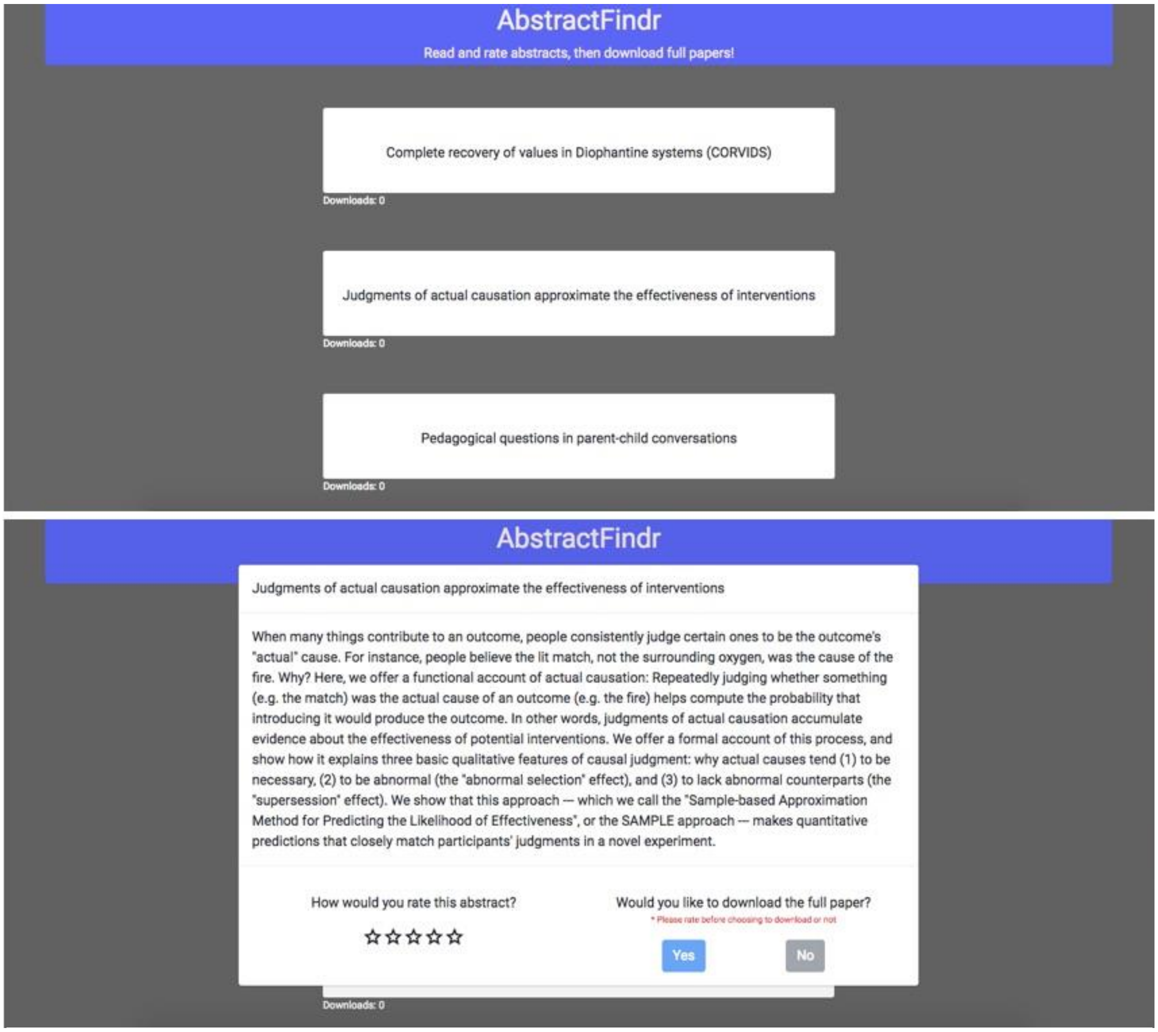


Figure 2. *(top)* The main page of AbstractFindr. Paper titles are displayed in a single column. In the influence conditions, download counts are displayed below the title of each paper, and paper titles are listed in descending order from most to least downloaded. If two papers have the same number of downloads (including zero downloads), titles are displayed in a randomized order. When a paper title is clicked on, a pop-up with the title and abstract text is presented, along with an input for rating the abstract (one-to-five-stars) *(bottom)*. Once rated, a download decision can be made, where a click on "yes" downloads a pdf file of the full paper to the visitor's computer. Participants must rate the abstract and choose to download or not for the pop-up to disappear.

make a list of papers comparably long to what is typically encountered by readers (i.e. on Google Scholar or PubMed), and consideration of the time burden for participants. All papers used as stimuli in this study were preprints publicly available on PsyArXiv, a digital platform for the distribution of non-peer reviewed academic preprints in psychology (www.http://psyarxiv.com). Papers were considered for this study if they met the following criteria:

1. The paper had fewer than 150 downloads on PsyArXiv on the acquisition date (May 9th, 2018).
2. The paper made no indication that it was either published in an academic journal or was submitted for peer review at an academic journal.
3. The paper made no indication that the authors did not want further distribution of the preprint.

Papers were chosen to represent a variety of topics within the field of psychology. See Table 5 for the complete list of papers used.

The population of interest for this study was all individuals who consider themselves researchers in psychology. This included graduate students, post-doctoral researchers, lab managers, professors of all tenure levels, non-tenure track lecturers, and other titles that may be used in other countries. As this population is broad and dispersed geographically, distribution of this web tool was digital.

First, the web tool was packaged inside a Qualtrics survey for digital distribution. This was to provide another point of informed consent for all participants. After providing consent, participants were asked to visit AbstractFindr with a provided link. Invitations to participate in this study were distributed over Twitter, through the "PsychMAP" Facebook group, to members and member laboratories of the Psychological Science Accelerator via their network-wide email list, and via direct email to all American tenured or tenure-track faculty associated with a PhD-granting psychology department. For more detail on this sub-population, see the Methods section of Chapter 3. Participants were also encouraged to share the participation link to their peers, students, and friends within the population of interest.

Procedure. Participants were first asked to accept the End User Agreement, which provided details about ownership of the website, ownership of the academic papers hosted on the website, how collected data will be used, and who has access to collected data. This agreement served as a second level of informed consent. The entire agreement is available in Appendix 1. This page of AbstractFindr also provides instructions on how to use the tool. Screenshots of each front page of AbstractFindr are available in Appendix 2. The main page of AbstractFindr is available in Figure 2.

After accepting the End User Agreement, participants were asked their gender and their sub-field of psychology. They were also asked about how they heard about the study, how they typically access the published literature, and their current academic position. They were then randomized into either the control condition (independent of social influence) or the social influence condition. See Figure 3 for the complete randomization path, which is further detailed in the following section.

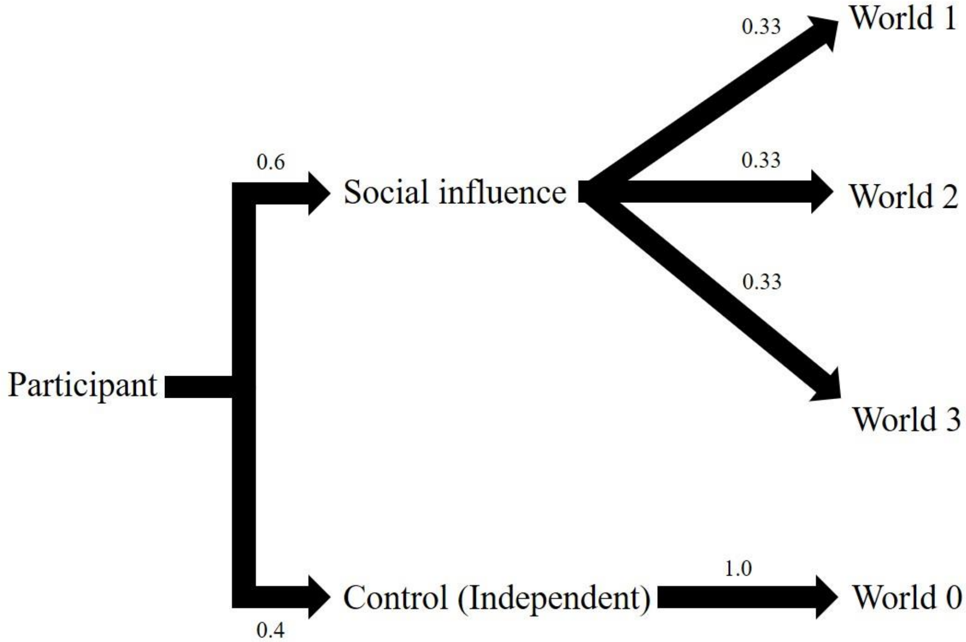


Figure 3. Complete randomization path potential for each participant. Value above each path represents the probability of being randomized into the subsequent category.

Social influence. Participants were presented with academic paper titles in a 48x1 column in the center of their web browser (See Figure 2, top panel). In the experimental condition, a download count for each paper was displayed beneath each title. This download count accurately displayed how many times that paper was downloaded by previous participants. Additionally, in the experimental condition, paper titles were ordered in descending order by download count. In other words, the paper that had the most downloads was presented first in the list, followed by the paper with the second most downloads, and so on. If multiple papers had the same number of downloads (including zero downloads), those titles were presented in a random order per participant within that segment of the list. In the control condition, the download count was not displayed to participants, and all paper titles were ordered randomly in the list per participant.

Participants were randomized into either the control condition (40%) or the experimental condition (60%). Those randomized into the experimental condition were then further randomized into either World 1, World 2, or World 3 – three parallel but independent versions of the experimental condition. Participants were free to navigate the tool as they liked, and could click, rate, and download as many titles as they'd like.

Early participants generated the social influence cues used by later participants and thus did not experience as strong social cues themselves. In other words, the download decisions made by the first participants of each world created the experimental conditions experienced by later participants. The download decisions made within a world (i.e., World 1) were exclusive to that world, and did not influence participants in other worlds. This design may potentially bias differences between control and experimental conditions conservatively, as the experimental condition got stronger as a function of participant count. However, this design allows for the measurement of unpredictability, which would otherwise not be measurable.

Design statement. This study is conceptually a 1x2 design (with 2 levels of social influence: no influence and influence), but was conducted as a 1x4 design, with 2 levels of social influence across 4 worlds (World 0, World 1, World 2, and World 3). Worlds 1, 2, and 3 are identical social influence replicates of each other. World 0 was the control world, independent of social influence.

This study had two outcomes of interest: download inequality, which measured how equal or unequal downloads were distributed across all possible papers, and unpredictability of success, which measured how stable a paper's popularity is across different identical worlds. If social influence affects how researchers decide to download academic papers, both download inequality and the unpredictability of success will be higher in social influence worlds (Worlds 1, 2 and 3) compared to the control world (World 0). I did not hypothesize a difference between experimental worlds, as they were exact replicates of each other. Descriptions of the analyses for download inequality and unpredictability of success follow below.

Bootstrapped Confidence Intervals. All intervals reported were estimated using the percentile bootstrap, which is a non-parametric test. In bootstrapping, the original sample is treated as a miniature representation of the larger population sampled. Observations in the sample are resampled with replacement, until a new sample of the original size is obtained. From this new sample, a new estimate is calculated. These steps are repeated 10,000 times to obtain a distribution for each estimate. Values at the 2.5th percentile and 97.5th percentile represent the lower and upper bound of the 95% CI for each estimate. All CIs were calculated in the statistical program R using the "boot" package (Canty & Ripley, 2017; Davidson & Hinkley, 1997). Unlike parametric tests that generate confidence intervals based on the Gaussian sampling distribution, bootstrapped confidence intervals are generated from the 10,000 draws described previously. Since this distribution is directly drawn from the observed data (which could be dependent or independent observations), there is no assumption of independence or normality.

Inequality of downloads. Download inequality was measured by calculating the Gini coefficient per influence condition per world for the population of 48 papers. The Gini coefficient represents the expected difference in market share between two randomly chosen papers, scaled so that a coefficient of 0 represents complete equality (every paper has the same number of downloads) and a coefficient of 1 represents complete inequality (only one paper was ever downloaded) (Atkinson, 1970; M. J. Salganik, Dodds, & Watts, 2006). Gini coefficients were calculated in the statistical program R using the "ineq" package (Zeileis & Kleiber, 2015).

To determine if social influence affected how papers were downloaded, bootstrapped 95% confidence intervals were generated for each world. For greater detail, see the previous section titled "Bootstrapped Confidence Intervals". Gini coefficients for each world were bootstrapped. Differences between worlds were considered statistically significant if the bootstrapped confidence interval for one world did not contain the estimate for another world.

Unpredictability of success. To determine if social influence changes the predictability that a particular paper becomes popular, or highly downloaded, this study collected data in three different "worlds". The 60% of web tool visitors randomized into the experimental condition were further randomized into either World 1, World 2, or World 3 (0.33 probability per world, so that 20% of total visitors were randomized into each experimental world). Visitors only saw download counts that corresponded to the downloads of earlier visitors in their assigned world. Similarly, the list of papers was only sorted by the downloads of earlier visitors in that same world. For example, the 100^{th} participant assigned to World 1 only saw the download counts corresponding to the previous 99 participants in World 1. They did not see the downloads of the other 100 participants that were randomized into World 2 or World 3. In this way, this study allowed the direct observation of the evolution of paper popularity in three independent, but identical, environments.

Unpredictability of success is defined as the sum total difference in download market share for an individual paper across a pair of experimental worlds (Salganik et al., 2006), for all possible world pairs. If a paper receives the same market share in World 1 and World 2 and World 3, its unpredictability is zero. Unpredictability of each paper (u_i) was calculated across World 1, World 2, and World 3 as,

$$u_i = \frac{|m_{i,1} - m_{i,2}| + |m_{i,1} - m_{i,3}| + |m_{i,2} - m_{i,3}|}{3}$$

where $m_{i,1}$ is the market share for paper i in World 1, $m_{i,2}$ is the market share for paper i in World 2, and $m_{i,3}$ is the market share for paper i in World 3. For the control condition (World 0), data was randomly split into two subgroups, and unpredictability was calculated across these two subgroups.

The unpredictability of a condition (U) (no social influence and social influence) will be calculated as the average unpredictability for that condition,

$$U = \frac{\Sigma u_i}{48}$$

To determine if social influence affected unpredictability of success, bootstrapped 95% confidence intervals were generated for experimental and control unpredictability (U). The difference between experimental and control influence conditions was considered statistically significant if the bootstrapped confidence interval for one condition did not contain the estimate for the other condition.

Results

Participants and web tool usage. The AbstractFindr webtool was visited by 215 unique users between August 17th, 2018 and October 18th, 2018 (a total of 63 days). Participant descriptive data can be seen in Tables 1, 2, 3, and 4. Following the randomization plan in Figure 3, 71 participants were randomized into World 0, 56 into

Table 1. *Reported gender of participants. N = 215.*

Gender	Participants	Proportion
Male	116	54.0%
Female	89	41.4%
Unanswered	10	4.7%
Total	**215**	**100%**

Table 2. *Reported gender of participants by social influence condition. N = 215.*

Condition	Male	Female	Unanswered	Total
Independent	35	30	6	71
Influence	81	59	4	144

Table 3. *Psychological disciplines reported by participants. N = 215.*

Area	Participants	Proportion
social	50	23.3%
cognitive	33	15.3%
clinical	28	13.0%
developmental	17	7.9%
neuroscience	14	6.5%
I/O	12	5.6%
evolutionary	8	3.7%
metascience	8	3.7%
other	8	3.7%
personality	6	2.8%
undefined	5	2.3%
educational	4	1.9%
health	4	1.9%
psycholinguistics	4	1.9%
psychophysics	3	1.4%
applied behavior	2	0.9%
decision making	2	0.9%
forensic	2	0.9%
media	2	0.9%
religious	2	0.9%
history	1	0.5%
Total	**215**	**100%**

Table 4. *Ratings and download descriptive statistics by world. N = 215.*

World	Participants	Ratings	Downloads	Average number of abstracts rated	Average number of papers downloaded
0	71	410	105	5.77	1.48
1	56	196	50	3.50	0.89
2	43	131	52	3.05	1.21
3	45	169	63	3.76	1.40
Total	**215**	**906**	**270**		

World 1, 43 into World 2, and 45 into World 3. A total of 906 abstracts were rated and 270 full text papers were downloaded across all worlds. A vast majority of participants typically access scientific literature digitally, and most participants learned about the study by direct email invitation.

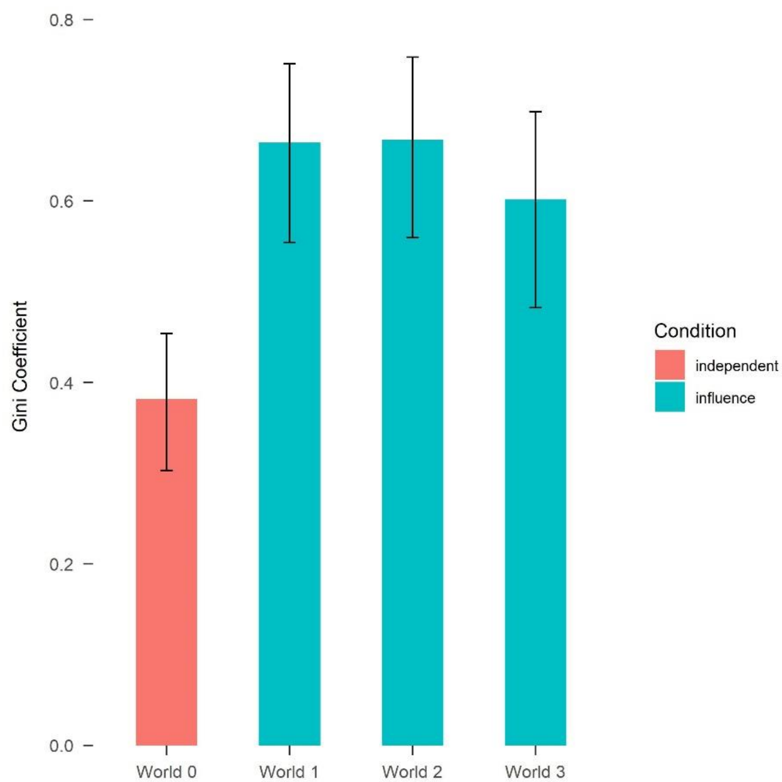


Figure 4. Download inequality, measured by Gini Coefficients, per world. World 0 (*pink*) had no indicators of social influence displayed to participants. Worlds 1, 2, and 3 (*blue*) each displayed accurate download counts beneath each paper title, and paper titles were sorted in descending order by download count. Estimated Coefficient with 95% bootstrapped CIs. From 270 total downloads.

Participants varied in psychological discipline. The greatest proportion of participating psychologists identified as social psychologists. Cognitive and clinical psychologists were also well represented. The full list of psychological disciplines of the participants can be seen in Table 4. A large majority of participants almost always used a digital source when accessing scientific literature (95.3%).

Download inequality. There was higher download inequality in all three influence conditions worlds compared to the independent world (see Figure 4). World 0, the control condition which had no indicators of social influence, had a Gini Coefficient of 0.38, and a 95% bootstrapped confidence interval of [0.30, 0.45]. World 1 had a Gini Coefficient of 0.66 [0.55, 0.75]. World 2 had a Gini Coefficient of 0.66 [0.56, 0.76]. World 3 had a Gini Coefficient of 0.60 [0.48, 0.70]. The total number of downloads per paper per world can be seen in Table 5.

The confidence interval for World 0 does not contain any of the Gini Coefficients for the social influence worlds (1, 2, or 3), meaning the Gini Coefficient for World 0 is significantly smaller than the Gini Coefficients for World 1, World 2, and World 3. The confidence interval for World 1 contains the Gini Coefficient estimates for World 2 and 3, the confidence interval for World 2 contains the Gini Coefficient estimates for World 1 and 3, and the confidence interval for World 3 contains the Gini Coefficient estimate for World 1 and 2, meaning there is insufficient data to determine if the Gini Coefficient estimates calculated from these three worlds are statistically different from each other.

Table 5. *Paper titles and downloads by world.*

Paper Title	World 1 Downloads	World 2 Downloads	World 3 Downloads	World 0 Downloads
Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis	6	2	0	6
The influence of subjective sleep quality on the association between eveningness and depressive symptoms.	6	1	1	1
Do women's preferences for masculine voices shift across the ovulatory cycle?	4	2	3	4
Judgments of actual causation approximate the effectiveness of interventions	4	4	1	3
A Concept Map of Curiosity Literature	3	3	0	4
Pedagogical questions in parent-child conversations	3	0	0	3
Predictors of grit: A multilevel model examination of demographics and school experiences	3	0	4	2
The Effects of Mindfulness-Based Interventions on Cognition and Mental Health in Children and Adolescents: A Meta-Analysis of Randomised Controlled Trials	3	3	0	3
A Brief Report on the Relationship between Attachment Style, Self-Esteem, and Multidimensional Romantic Jealousy	2	3	0	2
What's fair? How children assign reward to members of teams with differing causal structures	2	0	2	1
Does Construal Level Affect the Development of Automaticity? Abstract versus Concrete Mindsets and Choice Repetition	1	0	0	4
Emergence, expectation and causal schemas	1	2	1	3
Eye contact modulates facial mimicry in 4-month-old infants: an EMG and fNIRS study	1	0	0	3
Fast-Forwarding Disgust Conditioning: US Pre-Exposure Facilitates the Acquisition of Oculomotor Avoidance	1	0	0	3
How the Abstract Becomes Concrete: Irrational Numbers are Understood Relative to Natural Numbers and Perfect Squares	1	0	1	1
Modeling Second-Language Learning from a Psychological Perspective	1	1	3	4
Motivation for social bonding promotes high-stakes cooperative strategies	1	5	3	2
Perceptual contributions to racial bias in pain recognition	1	0	8	5
Playground Social Interaction Analysis using Bespoke Wearable Sensors for Tracking and Motion Capture	1	1	3	0
Reciprocity of social influence	1	4	5	6
Social and configural effects on the cognitive dynamics of perspective-taking	1	0	0	3
Testing Expectancy, but not Judgements of Learning, Moderate the Disfluency Effect	1	0	0	1
The influence of first name valence on the likelihood of receiving help: A field experiment	1	1	2	3
Why Passionate Employees Can Have It All: Passion Lowers Time Stress by Enhancing Goal Integration	1	1	0	2
Young Children Police In-Group Members at Personal Cost	1	1	1	3
Behavioral Reconsolidation Interference with Episodic Memory Within-Subjects is Elusive	0	0	1	1
Comparing the Affectiva iMotions Facial Expression Analysis Software with EMG	0	2	1	2
Complete recovery of values in Diophantine systems (CORVIDS)	0	0	2	1
Constructing and model-fitting receiver operator characteristics using continuous data	0	0	2	1
Convex hull as a heuristic	0	0	2	0
Discrimination in the English and Polish Housing Markets	0	0	1	1
Do children privilege phonological cues in noun class learning?	0	1	2	1
How many voices did you hear? Natural variability disrupts identity perception in unfamiliar listeners	0	0	2	1
Impacts of familiarity, conflict, and sex on continuous interpersonal behavior	0	1	0	2
Incidental learning and long-term retention of new word meanings from stories: The effect of number of exposures	0	0	2	1
Near and Far Transfer in Cognitive Training: A Second-Order Meta- Analysis	0	6	3	1
Network models of driver behaviour	0	1	2	1
PERSONA: a Psychometric Monitoring Instrument for High-stress Individuals	0	1	3	2
Psychopathy and ratings of persuasiveness: Examining their relations in weaker and stronger contexts	0	0	0	0
Psychophysics with children: Investigating the effects of attentional lapses on threshold estimates	0	0	0	0
Putting the I" in environmentalist: Explicit (but not implicit) identity predicts pro-environmental action"	0	0	0	5
Reciprocal interactions between audition and touch in flutter frequency perception	0	0	0	1
The emergence of systematicity: how environmental and communicative factors shape a novel communication system	0	0	0	3
The interaction of the Need for Cognitive Closure with implicit and explicit guidance in wiki-based learning	0	0	0	2
The Language of New Terrorism: Differences in Psychological Dimensions of Communication in Dabiq and Inspire	0	1	1	3
The missing link? Testing a schema account of unitization	0	0	0	0
What are reaction time indices of automatic imitation measuring?	0	3	0	1
What’s in a chunk? Chunking and data compressionin verbal short-term memory	0	2	1	3

Abstract rating. Participants were given the ability to rate each abstract they chose to read, on a one-to-five-star scoring system (see Figure 2, *bottom*). Participants needed to rate the abstract before they could choose to either download or not download the full text of the paper they had selected.

The average rating for an abstract in the control condition (World 0) was 3.25 stars. The average rating for an abstract in the influence condition (Worlds 1, 2, and 3) was 3.50 stars (see Table 6). In an exploratory 1x2 ANOVA, the average rating of abstracts in the influence condition was significantly larger than the average rating of abstracts in the control condition, $F(1, 904) = 9.83$, $p = 0.002$. Participants in the social influence condition, on average, rated the quality of abstracts as higher than those in the control condition. As participants could rate multiple papers during their visit, ratings are nested observations within participant. A more appropriate analysis would be a multi-level model that accounts for this nesting structure. Due to the structure of the collected data, this type of nested analysis was not possible, as ratings were collected independently of participant and could not be traced back to the participants that generated them.

At a per-abstract level, 34 out of 48 papers were rated higher in the influence condition compared to the control condition. Only 13 abstracts were rated lower when influence was present, and one abstract received the same rating in both influence and control conditions. Figure 5 depicts the rating for each abstract by influence condition. Points above the line represent abstracts rated higher in the influence condition.

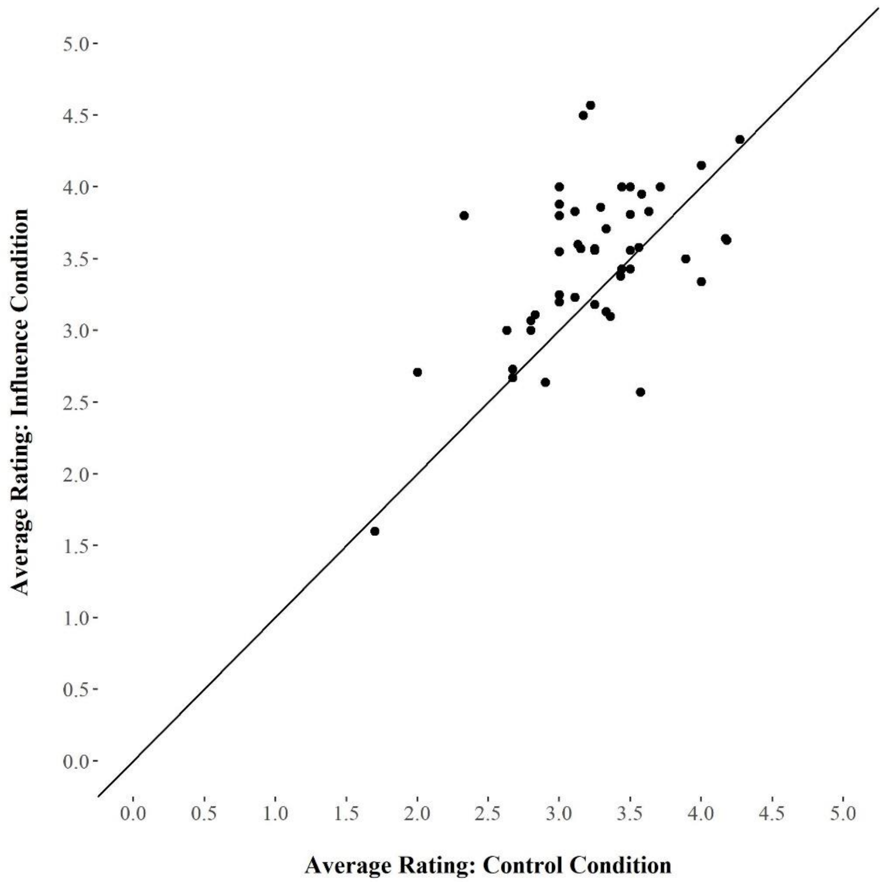


Figure 5. Participant ratings of abstracts by influence condition. Participants in the social influence worlds rated most abstracts higher compared to those in the control condition. Points above the line (y = x) represent abstracts rated higher in the influence condition. From 906 ratings across 48 papers.

Table 6. *Paper titles and ratings by influence condition.*

Paper Title	Average Rating: Independent	Average Rating: Influence	Rating Change
Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis	4.00	4.15	0.15
Reciprocity of social influence	3.15	3.57	0.41
Perceptual contributions to racial bias in pain recognition	3.58	3.95	0.37
Putting the "I" in environmentalist: Explicit (but not implicit) identity predicts pro-environmental action	3.50	4.00	0.50
Modeling Second-Language Learning from a Psychological Perspective	3.89	3.50	-0.39
Do women's preferences for masculine voices shift across the ovulatory cycle?	3.50	3.81	0.31
Does Construal Level Affect the Development of Automaticity? Abstract versus Concrete Mindsets and Choice Repetition	3.33	3.13	-0.21
A Concept Map of Curiosity Literature	2.00	2.71	0.71
The Effects of Mindfulness-Based Interventions on Cognition and Mental Health in Children and Adolescents: A Meta-Analysis of Randomised Controlled Trials	4.27	4.33	0.06
The influence of first name valence on the likelihood of receiving help: A field experiment	3.56	3.58	0.03
Pedagogical questions in parent-child conversations	3.50	3.56	0.06
Judgments of actual causation approximate the effectiveness of interventions	3.43	3.38	-0.05
Social and configural effects on the cognitive dynamics of perspective-taking	3.36	3.10	-0.26
Young Children Police In-Group Members at Personal Cost	3.25	3.57	0.32
What's fair? How children assign reward to members of teams with differing causal structures	3.22	4.57	1.35
The emergence of systematicity: how environmental and communicative factors shape a novel communication system	3.13	3.60	0.48
Eye contact modulates facial mimicry in 4-month-old infants: an EMG and fNIRS study	3.00	4.00	1.00
Fast-Forwarding Disgust Conditioning: US Pre-Exposure Facilitates the Acquisition of Oculomotor Avoidance	3.00	4.00	1.00
The Language of New Terrorism: Differences in Psychological Dimensions of Communication in Dabiq and Inspire	3.00	3.20	0.20
Emergence, expectation and causal schemas	2.67	2.73	0.06
Why Passionate Employees Can Have It All: Passion Lowers Time Stress by Enhancing Goal Integration	3.44	3.43	-0.02
Comparing the Affectiva iMotions Facial Expression Analysis Software with EMG	3.17	4.50	1.33
Predictors of grit: A multilevel model examination of demographics and school experiences	3.11	3.23	0.12
The interaction of the Need for Cognitive Closure with implicit and explicit guidance in wiki-based learning	3.00	3.25	0.25
PERSONA: a Psychometric Monitoring Instrument for High-stress Individuals	2.90	2.64	-0.26
Motivation for social bonding promotes high-stakes cooperative strategies	2.83	3.11	0.27
A Brief Report on the Relationship between Attachment Style, Self-Esteem, and Multidimensional Romantic Jealousy	2.80	3.07	0.27
Impacts of familiarity, conflict, and sex on continuous interpersonal behavior	1.70	1.60	-0.10
What’s in a chunk? Chunking and data compression in verbal short-term memory	4.18	3.63	-0.56
Near and Far Transfer in Cognitive Training: A Second-Order Meta- Analysis	4.17	3.64	-0.53
The influence of subjective sleep quality on the association between eveningness and depressive symptoms.	4.00	3.34	-0.66
How many voices did you hear? Natural variability disrupts identity perception in unfamiliar listeners	3.63	3.83	0.21
How the Abstract Becomes Concrete: Irrational Numbers are Understood Relative to Natural Numbers and Perfect Squares	3.50	3.43	-0.07
Discrimination in the English and Polish Housing Markets	3.44	4.00	0.56
Constructing and model-fitting receiver operator characteristics using continuous data	3.33	3.71	0.38
Incidental learning and long-term retention of new word meanings from stories: The effect of number of exposures	3.29	3.86	0.57
Do children privilege phonological cues in noun class learning?	3.25	3.56	0.31
What are reaction time indices of automatic imitation measuring?	3.11	3.83	0.72
Network models of driver behaviour	3.00	3.88	0.88
Testing Expectancy, but not Judgements of Learning, Moderate the Disfluency Effect	3.00	3.80	0.80
Reciprocal interactions between audition and touch in flutter frequency perception	2.80	3.00	0.20
Behavioral Reconsolidation Interference with Episodic Memory Within-Subjects is Elusive	2.63	3.00	0.38
Complete recovery of values in Diophantine systems (CORVIDS)	2.33	3.80	1.47
Psychophysics with children: Investigating the effects of attentional lapses on threshold estimates	3.71	4.00	0.29
Psychopathy and ratings of persuasiveness: Examining their relations in weaker and stronger contexts	3.57	2.57	-1.00
Playground Social Interaction Analysis using Bespoke Wearable Sensors for Tracking and Motion Capture	3.25	3.18	-0.07
Convex hull as a heuristic	3.00	3.55	0.55
The missing link? Testing a schema account of unitization	2.67	2.67	0.00
Average	**3.25**	**3.50**	**0.26**

Unpredictability of success. There was higher unpredictability of success in the social influence condition compared to the control condition (see Figure 6). The unpredictability of success for the control condition (World 0) was 0.008 [0.006, 0.011]. The unpredictability of success for the influence condition (Worlds 1, 2, and 3) was 0.05 [0.037, 0.064].

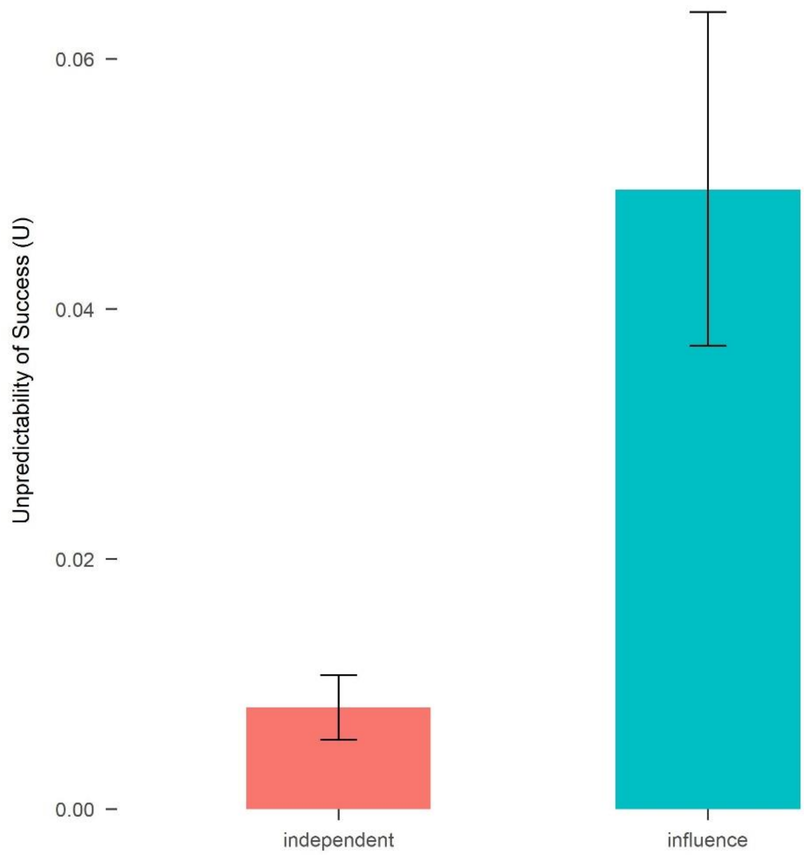


Figure 6. Unpredictability of success by social influence condition. The independent condition (World 0, *pink*) had no indicators of social influence displayed to participants. The influence condition (Worlds 1, 2, and 3, *blue*) displayed accurate download counts beneath each paper title, and paper titles were sorted in descending order by download count. Estimated unpredictability of success with 95% bootstrapped CIs. From 270 downloads.

The confidence interval for the independent condition does not contain the calculated unpredictability of success for the social influence condition, meaning the unpredictability of success for the independent condition is significantly smaller than that of the influence condition.

Discussion

In this study, there were two outcome measures that spoke to the behavior of scientists when they transact with the published literature, and one outcome measure that examined the effect of that behavior. The first measure on scientist behavior is the Gini coefficient, which is a measure of variability among values of a frequency distribution, first described by Corrado Gini in 1912 (Sen, 1973). In this study, a Gini coefficient of 1 represents maximum inequality, where one academic paper receives all the downloads from participants. Conversely, a Gini coefficient of 0 represents an equal distribution of downloads across all papers.

As seen in Figure 4, World 0, the control condition where no social influence was present, had a significantly lower Gini coefficient compared to the three worlds where social influence was present. There was insufficient evidence to detect a difference in Gini coefficient between the three influence condition worlds, suggesting similar levels of inequality across all three replicate worlds. If this finding were the result of random noise, Worlds 1, 2, and 3 would have randomly varied, and not have generated the extremely similar coefficients observed in Figure 4.

These data support the stated hypothesis. In the presence of social influence markers (an accurate download count per paper and a popularity-sorted list of papers), a smaller pool of papers accrued a larger share of downloads compared to the control condition, reflected in a larger Gini coefficient. This was true for all three replicate worlds. This result indicates a narrower approach to the published literature in the presence of social influence indicators, like download count.

The second outcome measure on scientist behavior was the subjective rating score. After selecting a paper title, participants were presented with the abstract to that paper, and then asked to rate the abstract on a one-to-five-star scale (see Figure 2, *bottom*). This rating was left intentionally without specification, as different readers may value different qualities in an abstract (such as writing clarity, novelty of methods or findings, general interest, or specific interest to one's own research). This type of rating scale was chosen as it is a relatively common way to assess general quality and is used extensively in many domains (i.e., rating service establishments on Yelp!, or hotels on TripAdvisor) (Wang, Lu, & Zhai, 2010).

As seen in Figure 5 and Table 6, participants on average rated abstracts significantly higher in the social influence condition compared to the control condition where no social influence markers were present. All 48 abstracts were rated in both conditions, and of them, 34 abstracts were more highly rated when social influence was present. The average rating increase for an abstract from the independent to the influence condition was 5.16%.

It is important to note that participants were not presented with any indication of how previous participants rated abstracts at any point in their use of the webtool. The only social influence presented was at the paper title level, where paper titles were ordered based on how many times the full text of that paper was downloaded, and the current count of full text downloads.

It is possible that participants used download count as a proxy measure of quality, and that influential metadata is reflecting the wisdom of the readers that came before. In this scenario, one would predict an abstract's rating would be correlated with the number of downloads it has received in the social influence condition. (a paper of higher quality would receive more downloads). There is insufficient evidence to conclude that this is the case, as the correlation between an abstract's rating in the influence condition and the download count in the influence condition is $r = 0.11$, $p = 0.44$.

The results from both outcomes suggest that social influence effects both steps in the two-step decision-making process put forth by Krumme et al. (2012), not just the first step. First, social influence markers are presented at the presentation of paper titles. This may influence which titles are clicked on for subsequent reading of the abstract. Although no social influence markers were present at the abstract level, readers rated abstracts higher in quality, potentially shifting the decision to download the full text of the paper.

The third outcome measure of this study was unpredictability of success, which measures an outcome of scientists' behaviors when transacting with the literature. This unpredictability was observed by the multi-world design of this study (Salganik et al., 2006). Worlds 1, 2, and 3 were designed to be identical in all ways. The only difference was how downloads were accumulated in each world based on the download behavior of the participants. The number of downloads per paper were set to zero at the onset of the study in every world. As participants were randomized into a world, their download behavior was recorded to that world only. For example, if Participant A downloaded Paper 15 in World 1, Paper 15 would show 1 download in World 1, but zero downloads in World 2 and World 3. In this way, these three separate worlds could generate three unique distributions of downloads across the corpus of 48 papers.

As seen in Figure 6, unpredictability of success was significantly higher in the social influence condition compared to the control condition. What this means is that there was observed instability in the rank position of a paper across worlds. For example, the paper "Doing Good vs. Doing Bad in Prosocial Choice: A Refined Test and Extension of the Morality Preference Hypothesis" tied with another paper in receiving the most downloads in World 1 (6 downloads), received 2 downloads in World 2, and 0 downloads in World 3 (see Table 5). Although the structure of the worlds was the same, the interaction between the participants and the literature developed an environment where this paper could be very popular in one world, and not at all popular in another. This instability is captured by the measure of unpredictability of success. If this finding were the result of random noise, unpredictability of success in the control condition would have been unlikely to have been significantly lower than the unpredictability of success in the social influence condition.

Strengths, limitations, and future directions. This study is unlike a real academic cultural market in several ways. First, much of the information typically associated with an academic article was stripped away so that participants were only using a very specific set of information in this study. The two biggest pieces of information not available to participants were the paper's authors and the name of a publishing journal.

All articles used in this study were pre-prints available on PsyArXiv.org, and none of the chosen articles indicated that they had been published in an academic journal.

Scientists have long used journals as indicators of topic interest and quality (Chen, 1974), and this influence may be even more important now given the limitations of researchers to adequately sample the growing corpus of published literature (Van Noorden, 2014). To limit the influence of journal name on the behavior of participants, only pre-prints were used as full-text articles. An important next step will be to investigate how journal prestige potentially changes the perceived quality and importance of published academic literature.

Authorship was removed from the papers used in this study as an attempt to prevent participants from engaging their scientific social network while participating in this study. Co-authorship rates have risen in the past 20 years (Henriksen, 2016), and Chapter 3 of this dissertation describes the average academic social network size of psychologists as over 180 individuals. While it is possible that participants could possibly identify the originating lab based on the content of the title or abstract, by removing the authors' names, direct recognition of individuals was eliminated. Future studies focusing on the relationship between authorship and reader's choice of academic literature will help to elucidate how social circles within the research community shape which findings become widely read.

There were some variables that we could not measure due to the architecture of AbstractFindr. Specifically, time spent per abstract and attrition rate are two variables that would have informed how participants interact with the site. Given that these variables went unmeasured, it is hard to determine the mechanism of action for the behavior observed. Future studies will benefit from a more robust webtool that can measure variables such as these.

One limitation was that popularity was confounded with article positioning on screen. The articles presented at the top of the screen were also shown as most popular. Thus, what seemed like social influence (more popular articles being downloaded more) may be a stimulus-positioning effect (articles presented at the top of the screen are more likely to be downloaded). It is unlikely this explains the observed data, as there was no evidence that participants only viewed the most popular paper (on average, participants rated between 3 and 5 abstracts per visit, see Table 4). Nevertheless, future studies in this domain would benefit from an additional condition where paper position was randomized while download count was displayed.

Figure 7. An example of digital metadata being used to influence literature search. An email from SAGE publishing (which publishes the journals of the Association for Psychological Science, or APS) promoting the most-read academic papers of 2018.

The real academic cultural market also includes other potential markers of social influence, such as mentions on Twitter or Facebook, emails from publishing journals (see Figure 7), and word-of-mouth from academic peers. The environment in which researchers judge and read academic papers is rich with social cues. Although this study reports effects of the presence of social cues as operationalized by download count, this may not have been the optimal cue. If researchers rely more heavily on recommendations from academic peers, or from Twitter, they may be more relevant social cues to participants. In the future, studies on the effect of social influence on researchers should investigate how different social cues are used.

The original proposal for this study called for the recruitment of 1,450 participants to detect a true effect size of d = 0.3 with a total balanced error rate of 5%. This level of participation was not achieved in this study. This could have been due to lack of any incentive for participating, only being available in English, or most likely, the relatively short timeframe of the study (63 days). An immediate drawback of this is the possibility of selection bias reducing the generalizability of the reported findings. Another potential issue is increased Type 2 error. Even though an effect was detected, it is possible that the observed variance in the gini coefficient and unpredictability of success is overly broad due to the relatively low number of participants.

There is evidence that the presence of social influence can modify the way participants subjectively rate the value of stimuli via modulated engagement of the nucleus accumbens and orbitofrontal cortex (Zaki, Schirmer, & Mitchell, 2011). Zaki et al., (2011) found this effect 30 minutes after presenting their social influence condition, suggesting the neuronal modulation could last at least this long. Considering the possible effect of social influence on the subjective quality rating of an abstract (devoid social influence) reported in this study, it is not unreasonable to suspect neuronal modulation as a possible mechanism of action. Continued work on the neuroscience of social influence may shed more light on this potential pathway.

Conclusion. This study is the first the report on the relationship between digital academic literature and reading behaviors among academics. It found that psychologists use social influence cues to guide their reading of the academic literature and that the presence of social influence cues make it more difficult to predict which academic papers will become most read. It also found the presence of social influence cues increased the subjective quality rating of abstracts. Together, this study sheds light on the processes that govern how psychologists interface with the published literature, a critical first step in generating new scientific knowledge.

Chapter 3:

How Many Psychologists Use Questionable Research Practices? Estimating the Population Size of Current QRP Users

It is the researcher's job to generate theories, test hypotheses, collect and interpret data, interpret results, and to publish findings. This is all done to learn more about the world and how it works. In the course of doing science, the researcher has many decisions to make: What past research most informs my current research? How many subjects will I use? How will I operationalize my variables of interest? What is the population of interest that I am studying? Should certain observations be excluded from the final analysis? Which statistical tests will I use?

Each decision point is a "researcher degree of freedom" (Simmons, Nelson, & Simonsohn, 2011), a decision in science with the potential of introducing error. Since there is a high level of ambiguity in research, these degrees of freedom can resolve in different ways. For example, in reviewing how researchers dealt with outlying observations, Simmons et al., (2011) found different research groups made independent decisions on the best course of action. When these researchers chose to remove outlying responses that were deemed "too fast", some defined this as two standard deviations below the mean response speed, some defined it as observations below 200 milliseconds, and others removed the fastest 2.5% of observations. None of these definitions are inherently incorrect interpretations of "too fast", which is part of the problem: without clear standards in place, flexibility in decision making can change the overall interpretation of a study's results.

There are many “researcher degrees of freedom” available to researchers that exploit the grey areas of acceptable practice (Wicherts et al., 2016) and may negatively infleunce the published literature by introducing false-positive or false-negative findings. Ten of these behaviors have been studied previously and have been collectively called “questionable research practices”, or QRPs. These ten behaviors occur during data collection, analysis, and reporting that have the potential to increase false-positive findings in the literature (L. K. John, Loewenstein, & Prelec, 2012). These 10 QRPs do not include behaviors that increase false-negative findings.

In previous literature on QRP use, data fabrication is included as a questionable practice. Fabrication, along with falsification and plagiarism, have been previously labeled “FFP” and are not considered in this dissertation as they are not questionable but instead academically dishonest (Steneck, 2006). Each of the remaining nine QRPs (with examples) can be found in Table 7. While there are many examples of other behaviors that could be considered questionable, these nine have been defined in previous literature as “Questionable Research Practices” and have been investigated previously (Agnoli, Wicherts, Veldkamp, Albiero, & Cubelli, 2017; Fiedler & Schwarz, 2016; L. K. John et al., 2012).

While there are some instances when QRP use may be justified, when they are used, they contribute to the false-positive rate observed in the published literature (Banks, Rogelberg, Woznyj, Landis, & Rupp, 2016; Fanelli, 2009). Not only does QRP use increase the number of false-positive findings (i.e. taking a non-significant result and pushing it over a designated threshold into being “significant”), but using multiple QRPs within a study can inflate the reported effect size (Button et al., 2013). Thus, QRP use can lead to field-wide interpretations that are not warranted by the data (Hopewell et al., 2009).

Prevalence of questionable research practices. Consider one of the most basic questions about the current replication crisis: How many people are contributing to it? John et al. (2012) reported that 63% of psychologists had published work without reporting all dependent measures (at least once in their academic career). As articulated by Simmons et al. (2011), this is problematic because increasing the number of dependent variables is correlated with an increase in the probability of finding a statistically significant result. Without reporting all dependent measures, readers are left with a false impression of the rarity or truthfulness of the reported findings.

The estimate reported by John et al. (2012) was contested by Fiedler & Schwarz (2016). In their conceptual replication that used differently worded questions, used a different conceptualization of "prevalence", and tested a German (as opposed to American) cohort of psychologists, Fiedler & Schwarz (2016) found less than 10% prevalence of the same questionable research practice (omitting dependent variables).

Even more recently, Agnoli et al. (2017) attempted to replicate the original John et al. (2012) study in an Italian cohort of psychologists, and found moderately high levels of QRP use (for example, 47.9% of respondents had omitted dependent variables). Consequently, there is no current consensus on the prevalence of QRP use in psychology, nor any indication of how these behaviors, reportedly used at least once in the span of a career, may be related to the current replication crisis in the field.

Given the inconsistencies in assessing the prevalence of QRP use, the present study seeks to expand the existing literature in several ways. First, this study will investigate current QRP users, operationalized as a person who has used at least one QRP "in the past 12 months". This orients QRP use into the timeframe of the current replication crisis. Second, it will address the larger issue of "prevalence", by defining behaviors performed within a specified time period. Previous work estimating QRP prevalence has done so over career-long timespans, or via estimating frequency of QRP use, both providing limited insight on the current issues in the field.

A third unique contribution of the present study is that it will assess prevalence of QRP use with three starkly different methodologies. One is a direct estimate, which is firmly based on previous research (Agnoli et al., 2017; Fiedler & Schwarz, 2016; Leslie K John, Loewenstein, & Prelec, 2012; Sijtsma, 2016): Researchers will be asked to report their own QRP use.

Although this assessment is straightforward, it is prone to confirmation and response biases (R. J. Fisher, 1993). For this reason, two other estimating methods will be used. One is the unmatched count technique, an indirect estimate aimed to reduce social desirability bias by removing the requirement for participants to identify as QRP users to the researchers (Arentoft et al., 2016). The second estimator generates an indirect estimate of QRP use by using social network information from the general population of psychologists (Jing, Qu, Yu, Wang, & Cui, 2014; Salganik, Mello, Abdo, & Bastos, 2011; Zhang et al., 2010; Zheng, Salganik, & Gelman, 2006). By asking psychologists about the behaviors of other psychologists they know, rather than their own behaviors, this estimator reduces the risk of socially desirable responses compared to more traditional direct estimate. Additionally, this social network method minimizes selection bias by accessing members of the population of interest through the social ties of participants. In this way, the current study generates three estimates of the prevalence of QRP users: two from investigating behaviors performed by study participants (via direct and unmatched count estimates) and one by asking about individuals the participants know (via the social network estimate).

Methods

Population of interest and target population. The population of interest for this work was all tenured or tenure-track researchers associated with a PhD-granting psychology department in the United States. QRP users (the target population) are therefore a subpopulation of this population, with a size greater than zero and maximally the size of the population of interest.

A complete list of names and contact information for the population of interest was provided via private correspondence with Dr. Leslie John (Leslie K John et al., 2012). The list provided was current as of 2010, so name and contact data was updated in May, 2017. When updating this data, an individual was coded as “absent” if their name was no longer associated with an institution, “new” if they were newly associated with an institution (i.e., their association to their current institution did not exist in the original list), “present” if their association and contact information did not change, and “updated” if their association did not change but their contact information changed. After updating the provided list, the total size of the population of interest was 7,101 researchers.

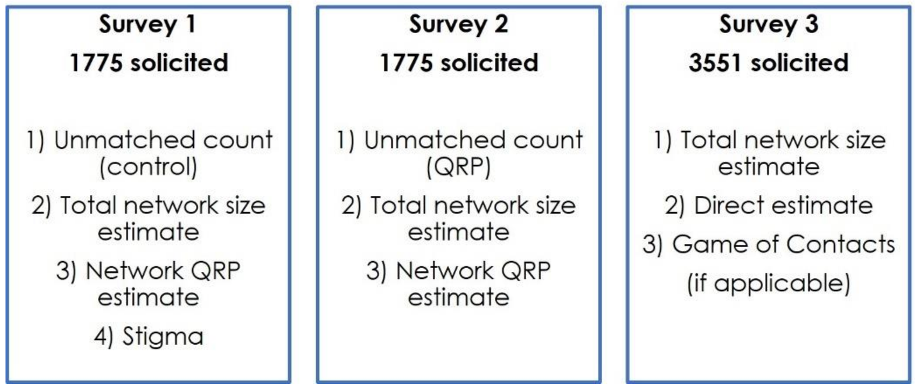


Figure 8. Estimators and question blocks within each survey.

Survey distribution. Members of the population of interest were invited via email to participate in a brief survey on personal social network size and attitudes towards researchers. All invitations were sent and all surveys were administered using the Qualtrics web tool (Qualtrics, 2005).

All members of the population of interest (N = 7,101) were solicited via email to participate. Emails were sent in a total of 10 waves, with each wave consisting of 200-400 invitations. This was to reduce the possibility of a technical error corrupting a large number of potential participants. All initial emails were sent to potential participants on a Thursday, and a single follow-up “reminder” email was sent on the following Monday. Participants who had finished the survey were sent a “thank you” email on the Thursday following the initial solicitation. All invitations were sent between September 2017 and December 2017.

Three surveys were distributed simultaneously. This was to facilitate the different types of direct and indirect estimates that will be described in the following sections. Surveys 1 and 2 were each distributed to 1,775 researchers. Survey 3was distributed to 3,551 researchers. All surveys included relevant instructions and definitions (i.e., defining behaviors identified as QRPs). See http://osf.io/d9bg5 for the survey materials distributed. The survey distribution design can also be seen in Figure 8.

In these surveys, "QRP use" was defined as having used at least one of the nine QRPs in Table 7 in the past 12 months. Similarly, a "QRP user" was defined as a person who has used at least one of the nine items in Table 7 in the past 12 months. Therefore, a QRP user is only defined by performing at least one of nine specific behaviors within a defined timespan. Participants are presented these definitions at the start of the survey and are shown that these definitions will always be available by hovering over text using their computer mouse.

The time frame of "in the past 12 months" is a reasonable time frame for investigating current or near-current behaviors. While some research has shown a 12 month timeframe may underestimate recall for some events (Connelly & Brown, 1995; Landen & Hendricks, 1995), this timeframe is used frequently to measure current behavior in major national data collection surveys such as the National Health Interview

Table 7. *Nine questionable research practices with examples.*

	Questionable Research Practice	Example
1	Failing to report all of a study's dependent measures	After measuring baseline anxiety with three different scales, only reporting one
2	Collecting more data after looking to see if the results were significant	Collecting 25 subjects' worth of data, achieving p = 0.07, then recruiting 25 more subjects.
3	Failing to report all of a study's conditions	Collecting data on three experimental conditions (no/low/high anxiety), but only reporting two (no/high anxiety)
4	Stopping data collection earlier than planned because one found the result one was looking for	After collecting half the final number of subjects, achieving p < 0.05, then stopping recritment
5	Rounding off p values to achieve significance	Achieving p = 0.054, then reporting it as p = 0.05 or p < 0.05
6	Selectively reporting studies that "worked"	Running eight studies, but only reporting the 5 that had significant findings
7	Deciding whether to exclude observations after seeing the effect of doing so on the results	Removing potential outliers based on their effect on the data and not on an a priori guideline
8	Reporting unexpected findings as being predicted from the start	Predicting a relationship between A and B, but finding a relationship being A and C, reporting the A-C relationship was predicted. Also know as Hypothesizing After the Results are Known (HARKing)
9	Reporting results are unaffected by demographics when actually unsure or not tested	Reporting no gender differences when gender was not collected or relationship not tested

Survey (NHIS) (United States Census Bureau, 2018) and the National Survey on Drug Use and Health (NSDUH) (Ahrnsbrak, Bose, Hedden, Lipari, & Park-Lee, 2017). Additionally, recall errors were further mitigated in Estimate 3, detailed below.

Survey responses. Of the 7,101 email solicitations sent, 214 emails bounced (3.01%). 613 full responses were collected (8.63% full response rate), and 296 partial responses were collected. There was no compensation offered for participation. Only full responses were used in the generation of population size estimates. Additionally, 26 participant responses were removed for either being marked complete erroneously by the Qualtrics webtool, or due to breaking estimate-specific criteria. For example, if a respondent claimed to know 200 individuals who have used a QRP in the past 12 months, yet the estimate of the size of their total social network was only 150 individuals, that respondent would be excluded from analysis.

299 (48.78%) participants identified as female, 279 (45.51%) identified as male, and 19 (3.10%) chose not to identify their gender. 131 (21.37%) participants identified as Assistant Professor, 141 (23.00%) identified as Associate Professor, and 208 (33.93%) identified as Full Professor. 113 participants identified as tenured or tenure-track, but chose not to disclose their tenure level.

Design statement. This study utilizes a population-based survey design to generate estimates of the size of the population of QRP users, as defined in the following sections. As such, estimates are valid only for the population of tenured or tenure-track faculty associated with PhD-granting American psychology departments. Since the entire population of interest was solicited to participate, there is no sample from which to infer about a larger population.

Estimate 1: direct estimate. The first of four estimates is a direct estimate of QRP usage. This serves both as a control estimate and as an estimate to compare to others than already exist in the field (Agnoli et al., 2017; Fiedler & Schwarz, 2016; Leslie K John et al., 2012). For this estimate, participants were asked if they used at least one of the nine QRPs (seen in Table 7) in the past 12 months. The proportion of researchers using QRPs was then calculated as:

$$\rho = \frac{c}{n}$$

Where ρ is the proportion estimate, c is the number of participants indicating they had used at least one QRP in the past 12 months, and n is the total number of participant responses. One question in Survey 3 was used to generate this estimate.

Estimate 2: unmatched count. The first indirect estimate utilized the unmatched count technique (UCT) to estimate the size of the QRP using population. In this estimate, two groups of participants are each given a list of innocuous items that could apply to them (i.e., I own a dishwasher, I exercise regularly, I enjoy modern art – examples from (Gervais & Najle, 2017)). The list of items is the same for both groups except for one additional item that one group receives and the other does not. This extra item asks about a sensitive identity (i.e., I own a dishwasher, I exercise regularly, I enjoy modern art, *I smoke crack cocaine*). For the current project, the sensitive identity is using at least one QRP in the past 12 months. See Table 8 for the full list of items and which group received which list of items. Participants were asked to count the number of items in the list that applied to them and to report that number into a text box. At no point did a participant have to identify themselves with any particular list item, only the total number of applicable items.

The proportion of participants that identify with the sensitive identity is calculated as follows:

$$\rho = \frac{\Sigma x_y^s}{n^s} - \frac{\Sigma x_y^i}{n^i}$$

where ρ is the proportion estimate, x_y^s is the number of reported items for participant y in the sensitive list s (list 2), n^s is the total number of participant responses in group s, x_y^i is the number of reported items for participant y in the innocuous list i (list 1), and n^i is the total number of participants in group i. Essentially, the mean difference between the two list groups is the proportion of participants who identified with the extra list item associated with the sensitive identity. Bootstrapped 95% confidence intervals were calculated to determine stability of the estimate.

Table 8. *Items used in the unmatched count technique. List 1 is the innocuous list, and List 2 is the sensitive list.*

	Item	**List**
1	I am a vegetarian.	1 & 2
2	I own a dog.	1 & 2
3	I work on a computer nearly every day.	1 & 2
4	I have a dishwasher in my kitchen.	1 & 2
5	I can drive a motorcycle.	1 & 2
6	My job allows me to work from home at least once a week.	1 & 2
7	I jog at least four times a week.	1 & 2
8	I enjoy modern art.	1 & 2
9	I have attended a professional soccer match.	1 & 2
10	I have used at least one QRP in the past 12 months.	2 only

One strength of the unmatched count estimate is that it generates an indirect proportion estimate without having to sample the target population (in this case, QRP users) specifically. It is a self-report of behavior that may reduce response bias as it is clear to participants that identifying group information cannot be gathered by their responses. However, estimate accuracy depends on participants being aware of their own group membership. Participants who are group members but are not aware of it (either due to self-deception or ignorance) may truthfully respond to list items but not count the sensitive item as one they associate with. Additionally, since the unmatched count is still a self-report, individuals may still choose to conceal their identity even when anonymity is methodologically guaranteed. Ultimately, this bias may lead to the unmatched count generating an underestimate of QRP user population size.

The first question block of Survey 1 contained the innocuous list condition for this estimate. The first question block of Survey 2 contained the sensitive list condition for this estimate (see Figure 8). No individual received both question blocks, and a total of 3,550 individuals were solicited to participate in this estimate. Data from these question blocks of Survey 1 and 2 were used to generated the final unmatched count estimate.

Estimate 3: the generalized network scale-up method. The second indirect estimate utilizes the social networks of participants to generate and estimate of group size and is called the generalized network scale-up method (GNSUM). The GNSUM is composed of three parts: the network scale-up method (NSUM) and two adjustment factors. Each part is described in detail below.

The network scale-up method. The network scale-up method estimates the size of a target group by utilizing the fact that it exists within a larger group, called the frame population (Bernard et al., 2010; McCarty, Killworth, & Bernard, 2001; Russell Bernard, Johnsen, Killworth, & Robinson, 1991) In this project, QRP users are the target groups that exists within the larger population of interest, tenured or tenure-track faculty associated with American PhD granting psychology departments.

In using the network scale-up method, each participant is asked about the total number of people they know to generate an estimate of the total size of their social network. They are also asked how many people they know in the target group (in this case, QRP users). Across many respondents, one can estimate the size of the target group as:

$$\rho = \frac{\Sigma y_i}{\Sigma d_i}$$

where ρ is the proportion estimate, y_i is the number of people known in the target group y by participant i, and d_i is the estimated total social network size d of participant i. Essentially, it is the number of QRP users known out of all the people known in the frame population. This approach of asking others to consider their peers to estimate y_i is not unreasonable; Fanelli (2009) estimated over 70% of researchers have witnessed QRP use in colleagues.

For this estimate, a participant knows a person if they fulfill the following criteria: 1) the participant knows the person and the person knows the participant (this is also known as reciprocal knowing), 2) the participant could get in touch with the person by any means, and 3) the participant has been in touch with the person in the past two years.

This definition was presented to participants before they were asked questions about the number of people they knew (McCormick, Salganik, & Zheng, 2010). Additionally, the population of interest was defined for each participant. Participants were told that questions would ask about how many research psychologists they know and were specifically told to focus on "research psychologists you know in any sub-field who are tenured or tenure-track faculty members associated with a PhD granting psychology department in the United States". Participants were then told that "in the next set of questions, these people will be referred to as 'research psychologists'". See http://osf.io/d9bg5 for the complete survey text.

Participants' total social network size (d) was calculated by asking about the number of research psychologists they know with certain first names and summing the responses per participant. Twenty-four first names from the population of interest were

Table 9. *Names used to estimate social network size within population of interest.*

Common	**Uncommon**	**Rare**
David	Alan	Dean
James	Donald	Seth
Jennifer	Sandra	Janet
Susan	Cynthia	Wendy
John	Keith	Guy
Mark	Gary	Jansen
Elizabeth	Emily	Susanne
Lisa	Andrea	Chantel

used: 12 male, 12 female, with a total of 8 common names, 8 uncommon names, and 8 rare names (4 per gender). Commonality of names were determined based on their frequency within the total population of interest. See Table 9 for the names used.

For each, participants were asked "how many research psychologists do you know named _______?". These names were randomized per participant, and all participants responded to all 24 names. These 24 names account for 964 individuals in the population of interest, or 13.57%. Responses from participants are summed and the percent population known is calculated, then multiplied by the population size to generate the estimate of the size of their social network. Due to the differential visibility of research psychologists with these names, a randomly chosen psychologist is more likely to know a David than a Jansen. Additionally, those participants with larger networks are more likely to know individuals with rarer names. Finally, although two participants may know the same psychologist named David, their responses in estimating their own social network members are independent from one another.

A major strength of the network scale-up method is that it does not require a sample of target group members (QRP users). This means a truly random sample can be drawn from the population of interest to estimate the size of a sub-population of individuals. Another strength is that it is a relatively short, straight-forward set of 30 questions that can be added to a battery of questionnaires.

However, the accuracy of the network scale-up estimate relies on two broad assumptions: 1) there is perfect information transfer between all members of a social network, and 2) the members of the population of interest and members of the target population have social networks of equal size. Put another way, the network scale-up estimate assumes an individual knows everything about everyone in their social network. To relax these assumptions, I also calculated the generalized network scale-up estimate, detailed below.

The generalized network scale-up method. The generalized network scale-up method elaborates on the network scale-up estimate by adding two adjustment factors to the equation, tau (τ) and delta (δ), to generate the final equation:

$$\rho = \frac{\Sigma y^i}{\Sigma d^i} * \frac{1}{\tau} * \frac{1}{\delta}$$

where ρ is the proportion estimate, $\frac{\Sigma y^i}{\Sigma d^i}$ is the network scale-up estimate, τ is the information transmission rate, and δ is the popularity ratio.

The information transmission rate, tau (τ). Most people do not know everything about the people in their social network. For example, you may not know that one of your friends (John Smith) was born in July. If I were to ask you how many people you know who were born in July, you count would be less by at least one, because you did not count John Smith, even though he was born in July. The information transmission rate, τ, adjusts the network scale-up estimate to account for the fact that information is not completely transparent. In this example, you did not know the birth month information for John Smith, so you did not count him when asked about the group membership of the people you know. This is especially important for identities that can be concealed, such as QRP use.

The information transmission rate is calculated using the game of contacts method (Salganik et al., 2012). The final output of this method is a value between 0 and 1, representing the transmission rate of certain information. A transmission rate of 1 represents total information transparency, and a value of 0 represents complete concealment. Values between 0 and 1 represent partial concealment. For example, a transmission rate of 0.5 would indicate 50% of social network members are aware of a participant's identity and 50% are unaware.

This method has participants answer a set of questions about information they know about other individuals in their social network, and what those individuals know about the participant. The questions and answers are given on a 2x2 grid, representing the four possible ways information can flow between a given two-person relationship (i.e., both know information about each other, one person knows and the other doesn't (either the participant or the individual), and neither person knows information about each other). An example question from the game of contacts can be seen in Figure 9.

Figure 9. An example question illustrating the game of contacts method. In this example, the participant indicated that they knew three research psychologists named David. They were they asked to think of each David, and to place the card corresponding to that David in the appropriate square.

The transmission rate was then calculated as follows:

$$\tau = \frac{\Sigma w^i}{\Sigma x^i}$$

where w^i is the number of individuals that know the participant is a member of the target population, and x^i is the total number of individuals identified by the participant.

Essentially, this is the sum of names placed in the top two boxes (see Figure 9) divided by the sum of individuals known by the participant.

The popularity ratio, delta (δ). Some people have larger social networks than others. Because of this, not everyone has the same level of network exposure. Individuals will smaller social networks will be known by fewer people. We can estimate the relative difference in social network size between the population of interest and the target population (QRP users) by comparing the responses to how many people participants in each group know when collecting data used to calculate d for the network scale-up estimate (described previously). If we find that members of the target population know, on average, one person named Mark, and members of the population of interest know, on average, two people named Mark, then assuming the target population is not especially likely or unlikely to know people named Mark, we can estimate that their network is 50% smaller than the network of the population of interest (Salganik et al., 2011). This calculation was made for all 24 names and was calculated as,

$$\delta = \frac{dE}{dT}$$

where dE is the average network size for the target population (QRP users), and dT is the average network size for the population of interest (tenure or tenure-track faculty associated with PhD granting psychology departments in the United States).

Data to calculate the transmission rate and the popularity ratio requires sampling from the target population. Survey 3 contained one question that directly asked participants if they had used at least one QRP in the past 12 months. This question primarily serves to directly estimate the number of QRP users (see Estimate 1).

Participants that respond that they have used at least one QRP in the past 12 months were then asked to complete the game of contacts questions (see Figure 9). Data for the popularity ratio was calculated using the direct estimate question response to categorize participants as either members of the target population or members of the population of interest.

There are three types of error that are associated with network scale-up estimators: transmission error, barrier error, and recall error.

Transmission error. Transmission error comes from the fact that not all people know everything about every individual in their social network. Left unattended, transmission error can cause the most uncertainty in a network scale-up estimate (Salganik et al., 2012). However, since this study collected data on how relevant information travels between individuals in a social network, transmission error can be accounted for by incorporating the transmission rate, τ, in the generalized network scale-up estimate.

Barrier error. Barrier error comes from the fact that social networks are not randomly assigned. In other words, there is non-random mixing of individuals within networks. This can lead to error in estimation when participants have differing probabilities of know individuals within certain social groups. For example, if I ask a participant how many widowers they know, a 25-year-old participant has a lower probability of knowing a widower than a 75-year-old participant. Another example is if I ask about how many pregnant women the participant knows. Female participants between 20 and 40 years old have a higher probability of having pregnant women in their social network than other participants.

This study aimed to reduce barrier error by asking about groups that should have random mixing: psychologists with particular first names. There is no theoretical reason why psychologists named James or Susan would cluster within certain individuals' social networks or geographic locations. In a professional field such as psychology, first names should be nearly randomly distributed. That said, barrier error cannot be ruled out as a source of bias in the reported estimate completely.

Recall error. Recall error comes from the fact that participants are asked to think about members of their social network and to report about them accurately. It is possible that a participant knows six psychologists named Elizabeth, but on the day of participation in this study, only recalls knowing five. Recall error is a source of estimation bias in many different types of estimation procedures (Killworth, Johnsen, Bernard, Ann Shelley, & McCarty, 1990; McCormick et al., 2010; Zheng et al., 2006).

The current study sought to minimize recall error by refraining from asking participants to recall their entire social network, which could number in the hundreds. Instead, participants were first asked to only consider their social network within the population of interest (as defined previously), and then were only asked to recall social network members with particular first names. Since total recall error is related to group size (Brewer, 2000), asking about multiple small groups should reduce recall bias compared to asking about one all-encompassing group. Recall bias is still an issue with any estimation procedure that relies on participant-generated data, and cannot be ruled out completely in this study's reported estimation.

Confidence intervals. To estimate the variability in the generated estimates of QRP use, bootstrapped 95% confidence intervals were generated. For greater detail on this procedure, see the Methods section of Chapter 2. For the direct estimate, the only observations that were bootstrapped were the responses to the direct estimate question. For the unmatched count estimate, the affirmed items for both the innocuous and sensitive lists were bootstrapped, and for the social network estimate, the observations used to calculate y^i, d^i, τ, and δ were bootstrapped.

Preregistration and data availability. This work was preregistered on May 15th, 2017. The preregistration can be found at the following link: https://osf.io/xu25n/. All data and analytic code associated with this work are available on an osf project page at the following link: https://osf.io/2zwqf/. A preprint of the manuscript produced by this work was made publicly available on August 14th, 2018 and is available at the following link: https://psyarxiv.com/3v7hx/.

This final study deviated from the preregistration in two places. First, the title was changed from the working title of "How Many Psychologists Use QRPs? A Social Network Approach to Estimating Hidden Population Size" to "How Many Psychologists Use Questionable Research Practices? Estimating the Population Size of Current QRP Users". This change was made for two reasons. First, not all readers may be familiar with the acronym "QRP", so defining it properly was deemed important. Second, as the study has multiple estimates, and not just social network estimates, the title was broadened to be inclusive of the direct and indirect estimates as well.

The preregistration called for this study to be conducted using two surveys, when instead three surveys were used. Originally, the direct estimate and game of contacts methods would have followed the innocuous list condition of the UCT. However, to avoid potential priming effects of one estimate affecting how a participant responds to the subsequent direct estimate, the direct estimate (and the subsequent game of contacts method) were moved into a new survey (Survey 3). Therefore, the final survey methodology consisted of three surveys, with one containing the innocuous UCT condition, one containing the sensitive UCT condition, and one containing the direct estimate.

Results

The three estimates of recent QRP use in the population of interest are summarized in Figure 10, and described in detail below.

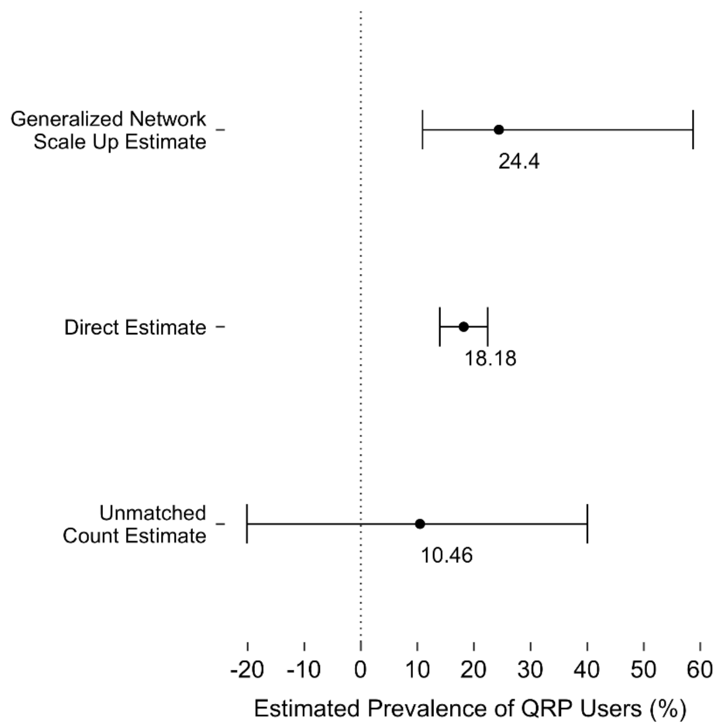


Figure 10. Estimates of the current prevalence of QRP users. The three different estimators are: the Generalized Network Scale Up Estimate, which used the social networks of participants to access the population of interest, the Direct Estimate, which asked participants about their QRP use directly, and the Unmatched Count Estimate, an indirect method of asking participants about their QRP use. Point estimates with 95% bootstrapped confidence intervals. Direct estimate N = 308. Unmatched count estimate N = 279. Generalized Network Scale Up Estimate N = 587.

Direct estimate. To ensure the highest number of participants potentially eligible to participate in our game of contacts method, half of the population of interest was solicited to participate in Survey 3, which contained the direct estimate question. Thus, 3,551 psychologists were solicited, and 308 responses were used for generating estimates. Of these 308 respondents, 56 indicated they had used at least one QRP in the past 12 months. Using the direct estimate equation, I calculated QRP prevalence to be 18.18%, with a bootsrapped 95% confidence interval of [13.96%, 22.40%]. This corresponds to an estimated 1,291 [991, 1590] American psychologists currently using QRPs.

It is possible this estimate underestimates the true number of psychologists using QRPs. For one, social desirability may lead some scientists who have used QRPs to be unwilling to admit it. This estimate is only generated by those participants willing to reveal their identity as a QRP user. Given the somewhat critical social environment that exists for QRP users in the field (Fiske, 2016; Teixeira da Silva, 2018), it is reasonable to believe some participants withheld their identity when asked directly. The following indirect estimation methods sought to mitigate this social desirability bias.

Unmatched count estimate. The remaining 3,550 psychologists contacted were asked to participate in the unmatched count estimate, with 1,775 individuals randomized into the innocuous list condition (list 1 in Table 8) and 1,775 individuals randomized into the sensitive list condition (list 2 in Table 8). From this, 279 responses were received for analysis.

The average number of list items corresponding to participants in the innocuous list condition was 4.28 items. The average number of list items corresponding to participants in the sensitive list condition was 4.39 items. Using the unmatched count estimate, this produced a QRP user prevalence of 10.46% [-20.19%, 22.40%]. This corresponds to an estimated 743 [-1433, 1590] American psychologists currently using QRPs.

It was unexpected that the calculated unmatched count estimate would be lower than the direct estimate. Typically, due to reducing response bias, indirect estimates like the unmatched count estimate are larger than direct estimates when the behavior or identity in question is concealable and potentially stigmatized (Gervais & Najle, 2017; Starosta & Earleywine, 2014; Wolter & Laier, 2014). Given the bootstrapped 95% confidence interval crosses zero, it is likely the relatively low number of participants in our unmatched count estimate (n = 279) led this calculation to be overly sensitive to individual responses. For example, if one additional participant in the innocuous list condition responded by identifying with 6 of the 9 items (1 standard deviation from the mean of 4.28), the calculated unmatched count estimate would change from 10.46% to 9.15%.

This hypothetical 12.5% change in the reported estimate based on the inclusion of one additional participant is worrying and should illustrate the sensitivity the unmatched count estimate has to individual responses with this low number of total participants. Due to this, and due to the confidence interval crossing zero, this unmatched count estimate should not be considered a valid or accurate estimate of the number of current QRP users in psychology.

General network scale-up estimate. All participants who were randomized into the unmatched count estimate were also asked to answer questions regarding their academic social network, and to estimate how many researchers they know who have used at least one QRP in the past 12 months. Participants who were randomized into the direct estimate and who self-identified as a QRP user in that estimate were also asked to answer questions about their academic social network and to participate in the game of contacts method (described previously). Participants in the direct estimate who did not self-identify as a QRP user were asked questions about their academic social network as well, but were not asked how many researchers they know who have used at least one QRP in the past 12 months. Therefore, social network responses were collected from 531 participants from the population of interest (used to calculate δ and d^i), 56 responses from participants who self-identified as QRP users who also completed the game of contacts (used to calculate δ and τ), and 279 responses from participants who estimated the number of researchers they know who have used at least one QRP in the past 12 months (used to calculate y^i).

These 279 individuals identified a total of 664 QRP users and know a total of 46,828 researchers. In other words, the sum of every individual's academic social network resulted in 46,828 researchers. Given that the size of the population of interest is only 7,101 researchers, I am fairly confident all or nearly all members of this population were captured within an individual social network at least once. Using just the network scale-up equation, QRP prevalence is estimated at 1.42% [0.85%, 2.14%]. This estimate serves as a base starting point for our key network estimate, the generalized network scale-up estimate, detailed below.

The equation for the generalized network scale-up estimate relaxes the assumptions of equal network size between QRP identifiers and non-identifiers and total information transmission by incorporating the estimates of τ and δ. Using the 531 responses from the general population of psychologists and the 56 responses from the participants who had indicated they had used at least one QRP in the past 12 months, δ was calculated as 0.97. The social network size of QRP identifiers is very nearly the same size as QRP non-identifiers. Using the game of contacts method, τ was calculated as 0.06. The transmission of QRP identity status to others is very low and will be explored in more detail in Chapter 4.

Using the equation for the generalized network scale-up estimate, QRP user prevalence is estimated to be 24.40% [10.93%, 58.74%]. This corresponds to an estimated 1,733 [776, 4171] American psychologists currently using QRPs.

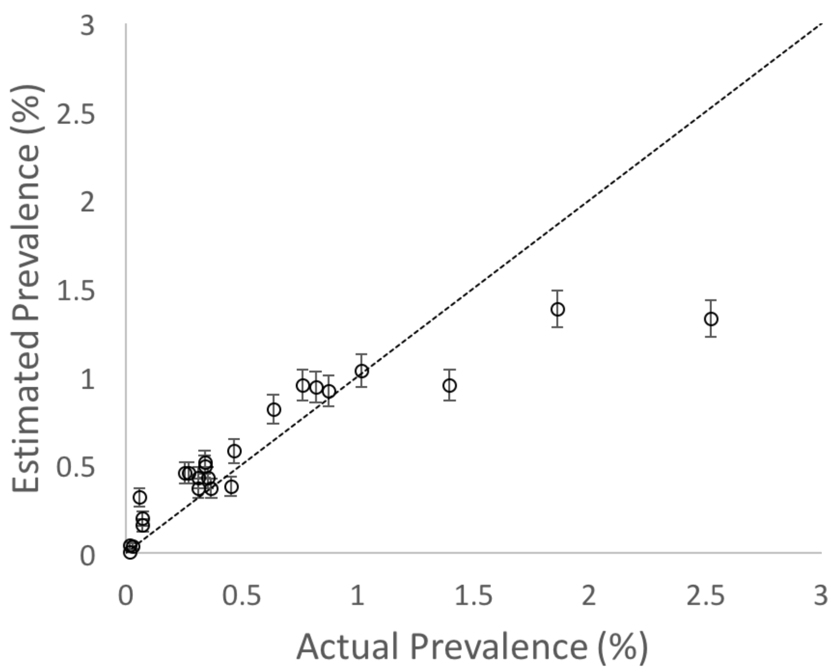


Figure 11. Validation of network scale-up estimates using 24 groups of known size. Each point represents one group, with corresponding 95% confidence intervals. Dotted line represents when estimated group prevalence equals actual group prevalence. N = 531. Correlation between estimated and actual group prevalence, r = 0.91.

Additional analyses were performed to assess the validity of the generalized network scale-up estimate by asking participants how many people they know in populations of known size. Estimates generated by responses were then compared to actual population sizes. If these estimates correspond well with the actual size of these populations, it would suggest the generalized network scale-up method most likely provides a good estimate of population size in this group of participants.

Since participants were asked to report the number of psychologists they know with particular first names to estimate their total social network size (see Table 9), this data was also used to generate generalized network scale up estimates for the population sizes of psychologists with each name.

The estimates made by the participants closely mirror the actual prevalence of these groups within the census of the population of interest – see Figure 11. The correlation between the participant's estimate of group prevalence and actual group prevalence is $r = 0.91$. It cannot be known for certain with these data whether the generalized network scale up estimate accurately identified the true proportion of QRP users in psychology. Nonetheless, that using this same method with these same participants accurately estimated the size of multiple populations of known size is consistent with the conclusion that the generalized network scale-up estimate used here also accurately estimates the proportion of QRP users in psychology.

Discussion

Because of inconsistencies in previous research, this study generated three independent estimates of current QRP use in American psychologists. Depending on the estimator used, it is estimated that 18.18% to 24.40% of American psychologists currently use questionable research practices. The estimate generated from the unmatched count estimate (10.46%) may not be valid or accurate due to the problems with the estimate described previously.

This is the first study to report the prevalence of QRP users in a proximal timespan. As such, it is difficult to draw conclusions about the magnitude of our estimates when compared to previous estimates.

Compared to John et al. (2012), and Agnoli et al. (2017), this study estimates a lower prevalence of questionable research practice use. Compared to Fiedler & Schwarz (2016), however, this study estimates a higher prevalence of questionable research practice use. This study used the same definition of "questionable research practices" as that used by both John et al. (2012) and Agnoli et al. (2017), but was restricted to a timespan of only 15 months, so it is reasonable that this would produce a lower estimate than those with an unrestricted timespan of QRP use. Additionally, since the definitions used were the same, it is also reasonable that this estimate is higher than the Fiedler & Schwarz (2016) estimate, as they changed the definition of each QRP.

Nonetheless, for QRP prevalence estimates to be useful, they must be confined to a timespan of interest. In the present work, that timespan was September 2016 through December 2017. This represents a time when psychology as a field had been introspecting on statistical and methodological issues that may have contributed to the contemporaneous replication crisis. For example, the findings from large-scale attempt at reproducing 100 psychological studies were published in August 2015 (Nosek, 2015). Furthermore, the Society for the Improvement of Psychological Science (SIPS), an academic organization that aims to “bring together scholars working to improve methods and practices in psychological science” had its first meeting in June 2016.

This is also the first study to use the generalized network-scale up estimator to investigate the prevalence of QRP users in psychology. Direct estimates rely on an individual’s willingness to participate and their willingness to honestly share their identity as a QRP user. Bias in either of these dimensions can distort a direct estimate, including the unmatched count technique, as it ultimately relies on an individual revealing potentially concealed information about themselves.

Social network methods, on the other hand, enable researchers to better understand the social processes at work that produce an environment where members vary in their identity and the information they share with others (Zheng et al., 2006). As participants are only reporting on the behavior of unnamed others, social desirability bias and selection bias are both reduced, producing an environment to measure a less bias estimate.

Limitations. The unmatched count estimate was lower than the direct estimate, and had a confidence interval that included zero, neither of which were expected. Due to the reasons stated previously, this estimate may not be valid or accurate. In a review of 101 publications that produced an estimate using the UCT, the median sample size was 1,000, and the lower quartile sample size was 562 (Hinsley, Nuno, Keane, St, & Ibbett, 2019). The estimate reported in this study had a sample size nearly half as small, putting this study at the very low end of sample size compared to other literature using this technique. Although there is no literature declaring that the UCT requires a particular minimum sample size, the presence of so few studies with sample sizes this small raises the possibility that larger samples are required to obtain a valid and reliable estimate. This casts doubt on the credibility of this measure. This essentially leaves this study with two valid estimates: the direct estimate and the generalized network scale up estimate. Future studies using the unmatched count technique may benefit from larger sample sizes, as demonstrated in Gervais & Najle, (2017) and reviewed by Hinsley et al., (2019).

QRPs exist in a grey area of accepted scientific practice. Therefore, it is difficult to interpret the severity of QRP use. This difficulty, along with the high variability among previous estimates of QRP prevalence, has led to several different conclusions. Some have concluded that the problems are overstated (Fanelli, 2018), while others have argued that current QRP use presents a real threat to the viability of several scientific fields, including education and political science (Bosco, Aguinis, Field, Pierce, & Dalton, 2016). Although the work presented in the current study moves the field forward in understanding the prevalence of those that use Type I error inflating behaviors, it provides less guidance on the severity of the consequences of QRP use on the whole.

Science is a globally distributed network, and as such, is difficult to study. The estimates reported in this study were limited to American psychologists, though these issues are not restricted solely to the United States (Agnoli et al., 2017; Fiedler & Schwarz, 2016; Forsberg et al., 2018). Future studies estimating the prevalence of QRP use in other countries or geographic areas will be an important next step to better understanding the global size of this population, as well as investigating the use of QRPs in other scientific fields. Some of this work has already started through the Horizon 2020 framework in the European Union (Forsberg et al., 2018), though more innovative work like this will be required to better understand the full scope of the problems faced.

Implications. These estimates should serve as a baseline to measure the effectiveness of current science reform initiatives, as well as a foundation for new ones. While much work is being done to grow support for initiatives such as pre-registration (E.-J. Wagenmakers & Dutilh, 2016) and Registered Reports (D. Chambers, Feredoes, D. Muthukumaraswamy, & J. Etchells, 2014), it is currently unknown what measurable effect these are having at curbing behaviors associated with inflated Type I error such as QRPs. By performing longitudinal follow-up estimates at future time points, the field can use the baselines estimates presented in this study to measure the effectiveness of these programs at reducing QRP use.

Additionally, the social network scale-up method is an estimation technique new to the field of psychology. This study demonstrates the usefulness of tapping into social networks to estimate the size of hidden populations. While it was used in this study to measure the number of current QRP users in psychology, there are many other uses of interest to psychological researchers. Having the power to measure populations traditionally difficult to study will be a benefit to future psychological research.

Conclusion. By directly asking survey participants about their use of QRPs, 18.18% have used at least one QRP in the past 12 months. By using the generalized network scale-up estimate, 24.40% of American psychologists have used at least one QRP in the past 12 months. This corresponds to between 1,291 and 1,733 individuals. Although some have argued the narrative of the “replication crisis” has become overblown (Fanelli, 2018), this study illustrates how common QRP use is. Although many have called for changes in statistical inference practices to mitigate false-positive findings (Benjamin et al., 2017; Lakens et al., 2018; Mcshane, Gal, Gelman, Robert, &

Tackett, 2017), it is important that the field also focuses on disincentivizing the use of questionable research practices (and other behavioral degrees of freedom) among our peers and coworkers for the betterment of our science.

Chapter 4:

Assessing the Stigmatization of Psychologists Who Use Questionable Research Practices

Most research on the current replication crisis has focused on methodological practices, publication bias, and false-positive findings (Fanelli, 2018). What was been neglected so far is a closer look at the individuals who perform questionable research practices. While the study described in Chapter 3 sought to estimate the number of QRP users in psychology, the study described here seeks to further understand how QRP-using researchers exist within the social structure of the field of psychology, and whether or not they are a stigmatized population.

The term “stigma” was formally defined by Erving Goffman as “an attribute that makes [a person] different from others in a category of persons available for [them] to be, and of a less desirable kind” (Goffman, 1963). Goffman describes two states of stigmatized identity: “discredited”, where the stigmatizing attribute is outwardly identifiable by strangers (i.e., race. gender, physical handicap – sometimes referred to as a “spoiled identity”), and “discreditable”, where the stigmatizing attribute can be concealed from others (i.e., sexual orientation, medical condition, certain mental disorders, behaviors). Since discredited people suffer from a reduced social status, it is potentially beneficial for discreditable people to conceal their stigmatized attribute and to continue being considered “normal” (Goffman, 1963). This is controlled through a process called “impression management”, where the actor (a person with a concealable stigma) communicates with an audience (the “normals” that are unaware of the actor’s true identity) in a manner to convince the viewers of the appropriateness of their assumed role in society. As long as the actor can convincingly portray their role as "normal" by concealing stigmatizing information, they may live and be perceived as a "normal" person (Goffman, 1959).

Impression management serves to change how individuals are perceived in the social environment. Previous work on person perception has demonstrated that observers process the array of stimuli presented to them by an actor in the form of visual cues or information provided in resumes, though these stimuli could take nearly any form (Fiske, 1980; Rubinstein, Jussim, & Stevens, 2018). One uniting characteristic that these stimuli must share is that they are in some way perceptible (Uleman & Kressel, 2013). Goffman (1963) describes a discreditable stigma as one that is imperceptible to others, and may only be revealed by the actor by word or by deed. Actors successfully concealing a discreditable stigma will not generate any stimuli that informs an observer about their discreditably. In this way, impression formation theory, which formalizes person perception and describes how individuals develop impressions of others, fails to adequately address the concealable nature of discreditable stigmas such as QRP use.

Impression formation can be taxing to the actor. Quinn and Chaudoir (2009) found people with concealed stigma have higher psychological distress and worse health outcomes when they believe others may learn about their stigma or when that stigma is central to their self-identity. Additionally, Smart and Wegner (1999) describe individuals concealing a stigma become cognitively overburdened when trying to keep their stigmatized identity a secret. Leary, Tchividjian, and Kraxberger (1994) found active impression management was a risk factor for pregnancy and STI transmission, unhealthy eating and dieting behaviors, legal and illegal drug use, and increased risk of physical injury. Finally, Pachankis (2007) describes the cognitive-affective-behavioral model, which provides a theoretical framework for understanding the psychological and health implications of living with a concealed stigma. In these ways, the social environment can exert a negative influence on those concealing a stigmatized identity, even when population members are unaware of the individual's particular stigma.

Reactions towards stigmatized members of society can differ depending on the perceived controllability the stigmatized individual has over their stigma. For example, people with lung cancer tend to be blamed more for their condition compared to other cancer patients due to the link between cigarette smoking (a controllable behavior) and lung cancer (Chapple, Ziebland, & Mcpherson, 2004). This effect persists even if the individual with lung cancer never smoked. Corrigan (2000) describes differing affective responses by population members towards individuals with stigma depending on whether or not that person is responsible for their stigma. Those seen as responsible for their stigma are met with anger and potential punishment, while those seen as not responsible are met with pity and potential helping behaviors. QRP use could be framed as either externally attributed or internally attributed. One could argue that QRP use is an inevitable outcome of working in a stressful academic career where success is measured in scientific output (here, QRP use is externally attributed to stress). It could also be argued that QRPs are only used by those unfit to be academics who resort to using QRPs to make up for their own inadequacies (here, QRP use is internally attributed to low ability).

There are ways that stigmatized individuals may attempt to manage their identity while minimizing negative effects. One way is through social withdrawal. If an individual fears that others may learn of their stigmatized identity, one way to minimize this threat of exposure is by reducing their number of social contacts. By interacting with fewer people, there are fewer moments when a concealed identity can be accidentally revealed (Ilic et al., 2014). Another way is through selective disclosure of stigmatized identity. Here, stigmatized individuals share their identity with others seen as trustworthy. Often, these people share the same or a similar stigmatized identity. Although both social withdrawal and selective disclosure can both help manage a stigmatized identity, they have diverging outcomes for the individual. Selective disclosure is an adaptive identity management strategy – it allows the stigmatized individual to control their social interactions in a beneficial way and reduces stigmatizing experiences. Social withdrawal, on the other hand, demands more from the stigmatized individual by asking them to continuously monitor their social network and anticipate their potential social interactions. This additional burden results in worse mental health outcomes and no reduction in stigmatizing experiences (Ilic et al., 2014).

Beyond the individual, it is important to consider the possible stigmatization of QRP users for the well-being of the whole of psychology. Determining if QRP use is stigmatizing will enable the development of interventions that either decrease or increase stigmatization. It is generally accepted that increased stigmatization of tobacco smokers has decreased the number of people who smoke (Bayer, 2008), though it is unclear whether the group or the individual should bear more of the stigma burden (Courtwright, 2013). For these reasons, it is important to first understand how QRP users exist within their social environment prior to implementing interventions aimed to reduce QRP use.

To access whether QRP use is stigmatizing, this study assesses the attitudes held by the general population of American psychologists towards QRP users, focusing on four theoretical domains: attribution theory and stigma, social norms and stigma, fear and stigma, and power and stigma (Stuber, Galea, & Link, 2008). These domains are important for understanding if QRP use is stigmatized by psychologists. For instance, population members may fear QRP use will damage the reputation of psychology as a scientific field and thus look down on those who they perceive to be negative contributors. Additionally, Link and Phelan (2001) argue that individuals who are stigmatized must be of lesser power compared to those doing the stigmatizing. This could be conceptualized as social, economic, or political power, any of which allow for the identification of differentness and the separation of those individuals into distinct categories.

In addition to measuring the attitudes of the general population of psychologists towards QRP users, this study also directly observes behaviors characteristic of individuals managing a concealed stigma: social withdrawal and selective disclosure. By using this two-pronged approach, this study will attempt to answer the following research questions:

1) Are QRP users stigmatized by the general population of psychologists?
2) Do QRP users behave as a stigmatized group?

Better understanding of how psychologists view their peers using QRPs will set a foundation for future interventions aimed at reducing QRP use.

Methods

Population of interest. The population of interest for this study was all tenured or tenure-track researchers associated with a PhD-granting psychology department in the United States. As this was the same population of interest for the study detailed in Chapter 3, data was collected for both studies simultaneously.

Survey distribution. Data was collected via three surveys, whose distribution was described previously (Chapter 3, Methods, Survey distribution, page 44). Briefly, three surveys were distributed to the population of interest (N = 7,101). Of these surveys, one survey (Survey 1) included questions to measure attitudes on QRP use and QRP users (see Figure 8). This survey was distributed to 1,775 members of the population of interest. This survey did not ask individuals about their own QRP use.

Survey 3 was distributed to 3,550 members of the population of interest. One question in this survey asked individuals if they had used at least one QRP in the past 12 months. If a participant responded that they had used a QRP, they were given an additional set of questions to estimate how transparent the QRP-user identity was to their peers. This set of questions is called the Game of Contacts. The responses to these questions will be used to investigate the stigma-related behavior of QRP users.

Survey responses. Of the 7,101 email solicitations sent across all three surveys, 214 emails bounced (3.01%). 613 full responses were collected (8.63% full response rate), and 296 partial responses were collected. 130 responses were collected from Survey 1, and of those, 98 were full responses without missing data. 56 participants identified as having used at least one QRP in the past 12 months.

Dependent measure. Because there was no existing measures of QRP-related stigma, questionnaire items measuring stigma related to being a QRP user was developed from a scale designed to assess perceived devaluation and discrimination related to smoking cigarettes (Link & Phelan, 2001; Stuber et al., 2008). The measure assesses respondent perceptions of what most other researchers believe. These items were modified to frame them in terms of QRP use. For example, the item "Most people think less of a person who smokes" was modified to "Most people think less of those who use QRPs". Cronbach's alpha was calculated to assess the reliability of the items as a scale (alpha = 0.78), suggesting acceptable internal consistency (Tavakol & Dennick, 2011). Responses to each question were on a four-point Likert scale that ranged from strongly disagree to strongly agree.

Independent measures.

(i) *Age:* Participants self-reported their age in years.

(ii) *PhD year:* Participants self-reported the year in which they obtained their PhD. Although collected, this measure was not used in subsequent analyses.

(iii) *Acceptability:* To access descriptive and injunctive social norms at a peer level, one question was asked to participants: "How do most of your colleagues feel about using QRPs? Do they think it is acceptable, unacceptable, or that they don't care one way or another?" The 17 participants who responded "they don't care one way or another" were excluded from analyses that included this measure.

(iv) *Attribution:* To assess what participants believe were the causes of QRP use, two questions were asked: “QRP use is due to weak character”, which was used to assess internal attribution. and “QRP use is due to stress”, which was used to asses external attribution.

(v) *Fear:* To access fear related to the academic hazards posed by QRP users in their capacity as mentors, one question was asked: “QRP users are a threat to their students”.

(vi) *Power:* Socioeconomic status was assessed by tenure level (assistant professor, associate professor, or full professor), and by individual income level (measured with six bins: less than $49,999, $50,000-$74,999, $75,000-$99,999, $100,000-$149,999, $150,000-$199,999, $200,000 or more). Although collected, tenure level was not used in subsequent analyses.

Control variables: Racial/ethnic status was assessed by self-identification of categories planned to be used in the 2020 U.S. Census (White, Black or African American, Latino, Hispanic, or Spanish origin, American Indian or Alaska Native, Asian, Middle Eastern or North African, Native Hawaiian or Other Pacific Islander, None of the Above, or Prefer Not to Say). Political orientation (politics) was assessed on a 6-point scale (Very Conservative, Somewhat Conservative, Middle-of-the-road, Somewhat Liberal, Very Liberal, and Not Sure). Gender was assessed as either Female, Male, or Prefer Not to Say.

Behavioral measures. To assess behaviors associated with concealing a stigmatized identity, social withdrawal and selective information transmission were measured. The average social network size of QRP users was measured as described previously (Chapter 3, Methods, page 50). If QRP users socially withdraw, it is predicted their average social network size would be smaller than the average social network size of the general population of psychologists. Selective transmission was assessed by measuring the number of alters in each QRP-user's social network who are aware of the QRP-use identity of the participant and assessing which alters are also QRP users. If a QRP user selectively discloses their identity information, it is predicted that another QRP user is more likely to know compared to a psychologist whose QRP use identity is unknown to the QRP user. In other words, QRP users disclose their QRP-use identity information to other QRP users rather than disclose to individuals with an unknown QRP-use status.

Statistical analyses. For descriptive analyses, responses answered on a four-point Likert scale were reduced to two bins ("agree" and "disagree"). Linear regression was used to assess the direct relationship between independent measures and the dependent measure using the statistical program R. A possible curvilinear relationship between power and QRP stigma was tested by introducing the squared power predictor to an additional model. Data points depicted in linear regression graphs were jittered to provide increased clarity. An odds ratio was calculated to determine the odds of a QRP-using alter knowing the participant's QRP-use identity compared to an alter with unknown QRP-use status knowing the participant's QRP-use identity. An independent samples t test was calculated to determine the mean difference between the average social network size of QRP users compared to the average network size of the general psychologist population.

Results

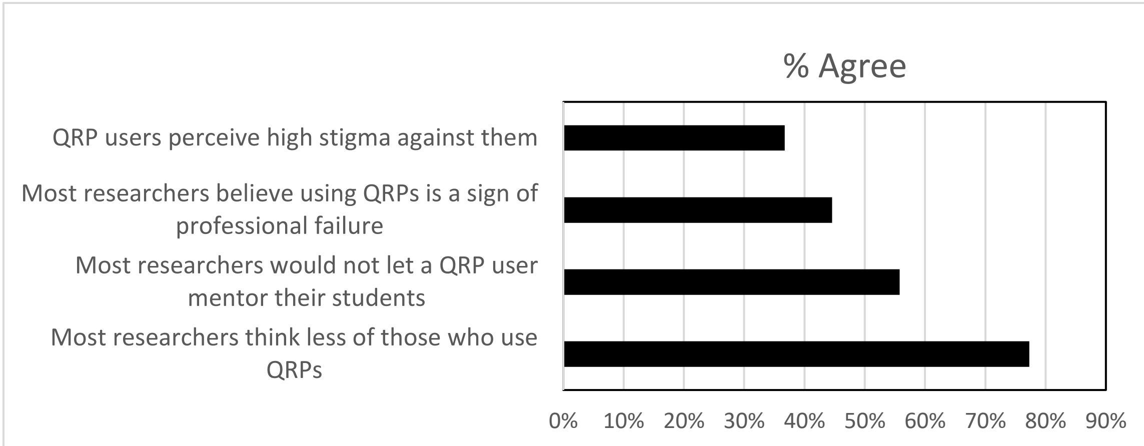


Figure 12. Prevalence of perceived stigma against QRP users. Questions asked to the general population of psychologists. N = 98.

Figure 12 shows the prevalence of perceived stigma against QRP users among the general population of psychologists. Participants agreed that most researchers think less of those that use QRPs (77.3%) and that most researchers would not let a QRP user mentor their students (55.8%). Additionally, 44.6% of participants agreed that using QRPs is a sign of professional failure. Interestingly, only 36.73% of respondents agreed with the statement that QRP uses perceive high stigma against them. It could be argued that the gap between "Most researchers think less of those who use QRPs" and "QRP users perceive high stigma against them" speaks to the nature of stigma itself: that it is a negative process established at the environmental level (as opposed to the individual level) by those free of the stigmatizing mark.

Table 10 reports the multiple regression output of all independent variables of interest regressed on the dependent variable. For this analysis, income was used as the operationalization of power, and age (in years) was used as the operationalization of age (as opposed to PhD conferral year) as these were more interpretable variables and have been used in previous literature (Stuber et al., 2008). This model also included the control variables of gender, ethnicity, and political orientation. Diagnostic plots for this model can be found in Appendix 3.

In this model, age and fear are both significant predictors of stigmatization of QRP users. Here, younger participants gauged QRP use as significantly more stigmatizing than older participants (p = 0.03), and those who feared QRP users as a threat to their students were significantly more stigmatizing to QRP users (p = 0.0069).

Table 10. *All predictor variables regressed on stigma dependent measure. N = 98.*

Coefficients	**Estimate (ß)**	**Estimate (*b*)**	**Std. Error**	**t value**	**p value**	
(Intercept)	---	7.8437	2.4069	3.26	0.0016	**
Acceptability	-0.01056	-0.0494	0.4917	-0.1	0.9202	
external attribution	-0.00675	-0.022	0.3511	-0.06	0.9502	
internal attribution	0.13725	0.4822	0.3803	1.27	0.2083	
Fear	0.31118	0.9136	0.3299	2.77	0.0069	**
Power	0.16157	0.368	0.2328	1.58	0.1178	
Age	-0.23245	-0.0408	0.0185	-2.21	0.03	*
Gender	-0.07151	-0.258	0.4253	-0.61	0.5458	
Ethnicity	-0.05825	-0.1381	0.2852	-0.48	0.6296	
politics	0.02853	0.0508	0.1909	0.27	0.7908	

To check for the possibility of collinearity between the predictor variables, a correlation matrix was calculated for all independent measures. The correlation matrix can be seen in Figure 13.

There were significant correlations between independent measures, providing evidence of possible collinearity. As seen in Figure 13, internal attribution of QRP use is positively correlated with fear of QRP users. Fear of QRP users is also positively correlated with external attribution of QRP use. External attribution of QRP use is also negatively correlated with age.

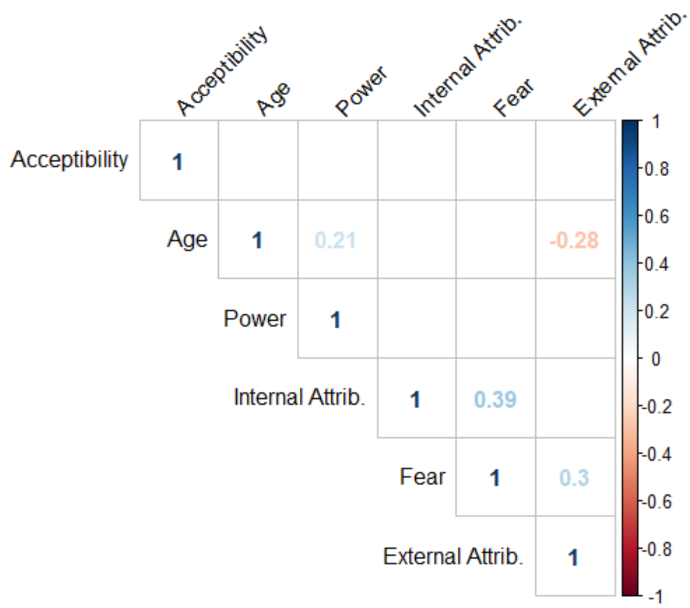


Figure 13. Correlation matrix between independent and dependent measures. Values represent Pearson's r values. Grid squares with a value represent a significant correlation ($p < 0.05$) between the two variables. Blue values represent a positive correlation, and red values represent a negative correlation. "Internal Attrib." and "External Attrib." are abbreviated for Internal and External Attribution, respectively. N = 98.

To better understand the extent of collinearity among independent measures, the variance inflation factor (VIF) was calculated for each. The VIF represents how much higher the variance of a coefficient estimate is due to how strongly the variable is correlated with at least one other variable (O'Brien, 2007). The minimum value for VIF is 1, and there is no maximum value. The VIF values for each variable are in Table 11. VIF values greater than 1 signify higher collinearity. For example, the VIF coefficient of the independent measure "Fear" in Table 11 is 1.28, which means that the variance in the beta coefficient estimate for "Fear" is 28% larger than it would be if it were uncorrelated with the other predictors.

Table 11. *Variance inflation factor (VIF) for each independent measure. N = 98.*

Measure	VIF
Age	1.19
Acceptability	1.02
Internal Attribution	1.23
External Attribution	1.21
Fear	1.28
Power	1.06

Although Hair, Anderson, Tatham, & Black (1995) suggest VIF factors less than 10 are indicative of inconsequential collinearity, it is theoretically important in this instance to look at the direct relationships between the predictors in the multiple regression and the QRP stigma outcome (Mela & Kopalle, 2002). Investigating the direct relationships between each theoretical domain of stigma and QRP stigma will provide additional insight into whether QRP use satisfies conditions predicted by stigma theory: namely, that QRP use breaks social norms, is internally attributed, is feared, and that QRP users are in a lower position of power compared to the general population. Age is an additional predictor that is outside of classic stigma theory, but interesting in this specific context, as QRP use and the resulting reform movement may unequally affect researchers across age (Everett & Earp, 2015).

Model 1: Age. The model of age regressed on stigma is depicted in Figure 14 and described in Table 12 below. Diagnostic plots can be found in Appendix 4.

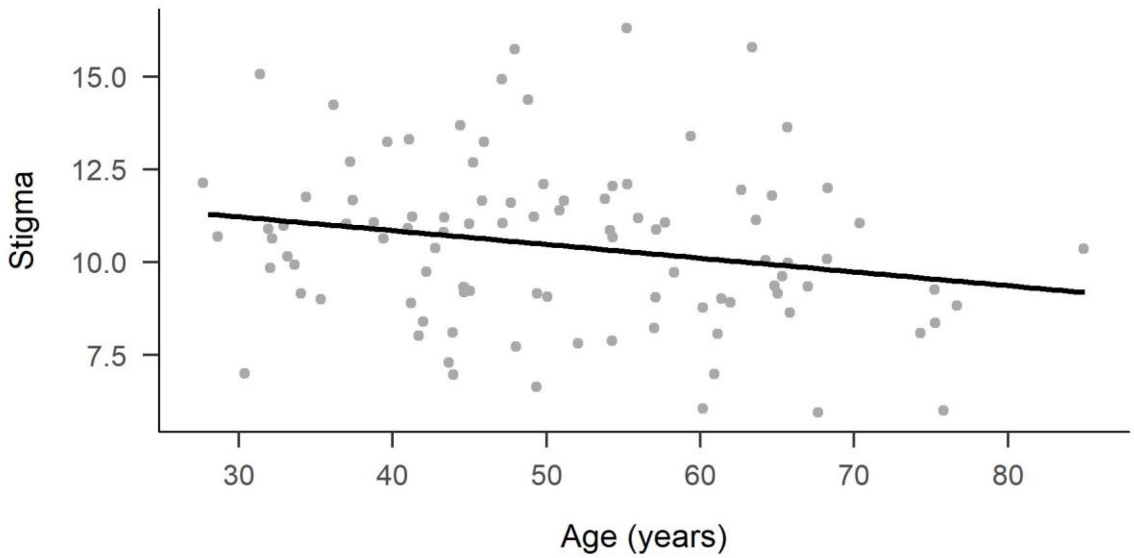


Figure 14. Bivariate relationship between participant age and QRP stigma. N = 98.

Table 12. *Age model regression coefficient estimates. N = 98.*

Coefficients	Estimate (ß)	Estimate (*b*)	Std. Error	t value	p value
(Intercept)	---	13.5863	1.6283	8.34	<0.001 ***
age	-0.2240	-0.0389	0.0178	-2.19	0.031 *
gender	0.0129	0.0436	0.3935	0.11	0.912
ethnicity	-0.1470	-0.3567	0.2826	-1.26	0.21
politics	-0.0665	-0.1208	0.1859	-0.65	0.517

Participant age was a significant predictor of stigma, with younger participants holding greater stigmatizing views of QRP users than older participants (b = -0.04, p = 0.031).

Model 2: Acceptability. The model of acceptability regressed on stigma is depicted in Figure 15 and described in Table 13 below. Diagnostic plots can be found in Appendix 5.

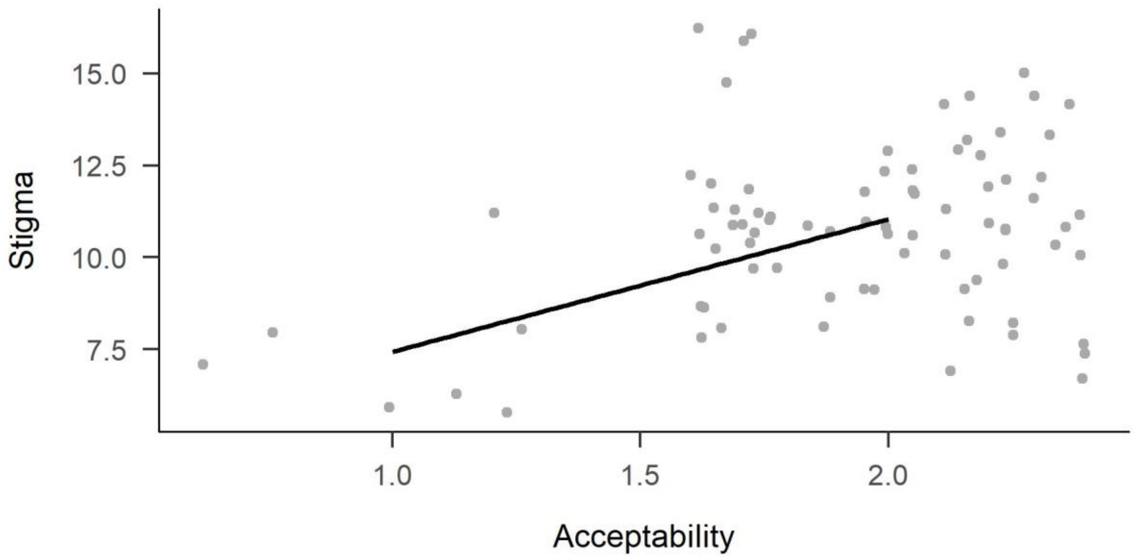


Figure 15. Bivariate relationship between acceptability of QRP use and QRP stigma. Acceptable dummy coded as "1", and unacceptable dummy coded as "2". N = 98.

Table 13. *Acceptability model regression coefficient estimates. N = 98.*

Coefficients	Estimate (ß)	Estimate (*b*)	Std. Error	t value	p value
(Intercept)	---	4.6574	2.1298	2.19	0.0318 *
acceptability	0.4278	3.5079	0.8818	3.98	0.0002 ***
gender	-0.0345	-0.1165	0.4082	-0.29	0.7762
ethnicity	-0.0372	-0.0876	0.2953	-0.3	0.7676
politics	-0.0287	-0.0548	0.2025	-0.27	0.7875

Acceptability of QRP use was a significant predictor of stigma. Those participants who considered QRP use unacceptable (dummy coded as 2) held greater stigmatizing views of QRP users than those who considered QRP use acceptable (dummy coded as 1, b = 3.51, p = 0.0002).

Model 3: Internal Attribution. The model of internal attribution regressed on stigma is depicted in Figure 16 and described in Table 14 below. Diagnostic plots can be found in Appendix 6.

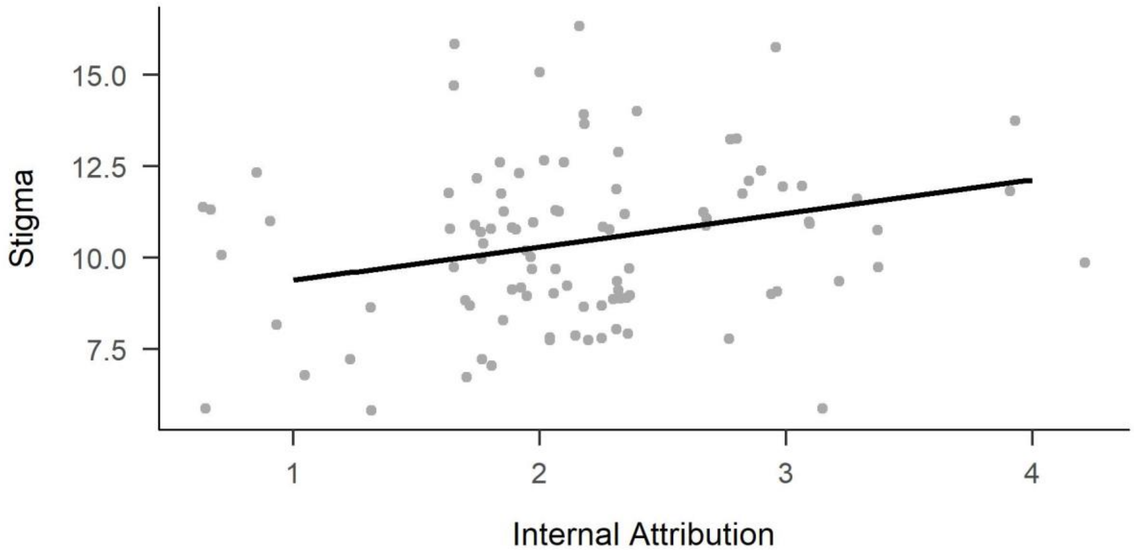


Figure 16. Bivariate relationship between internal attribution of QRP use and QRP stigma. N = 98.

Table 14. *Internal attribution model regression coefficient estimates. N = 98.*

Coefficients	Estimate (ß)	Estimate (*b*)	Std. Error	t value	p value
(Intercept)	---	9.1744	1.57911	5.81	<0.001 ***
internal attribution	0.27787	0.94849	0.34741	2.73	0.0076 **
gender	-0.09311	-0.31424	0.38699	-0.81	0.4189
ethnicity	-0.09525	-0.23111	0.27909	-0.83	0.4097
politics	-0.00124	-0.00225	0.185	-0.01	0.9903

Internal attribution of QRP use was a significant predictor of stigma. Participants who more strongly believed that QRP use was due to a researcher's weak character held greater stigmatizing views of QRP users (b = 0.948, p = 0.008).

Model 4: External Attribution. The model of external attribution regressed on stigma is depicted in Figure 17 and described in Table 15 below. Diagnostic plots can be found in Appendix 7.

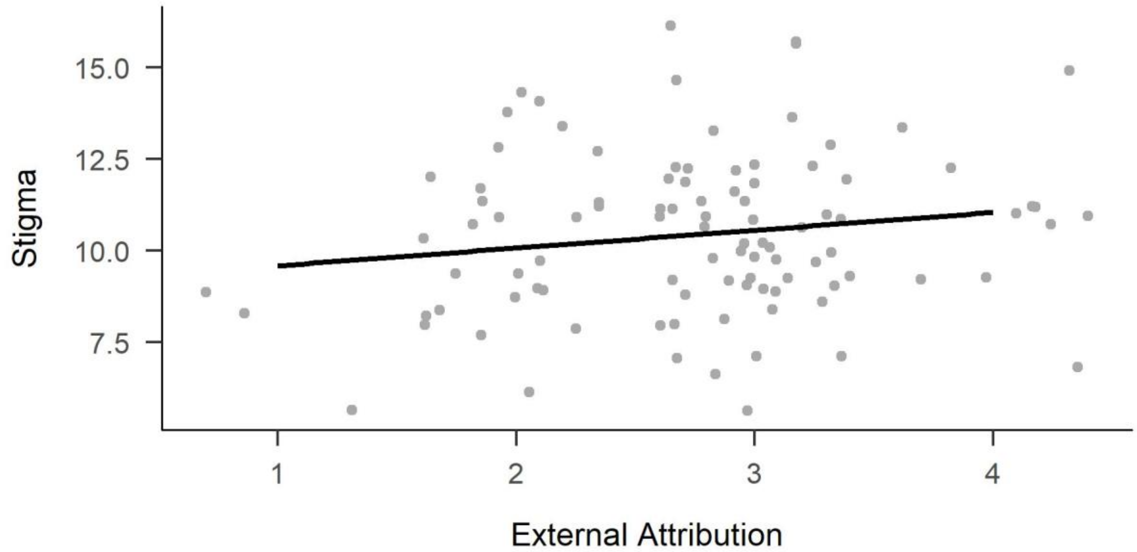


Figure 17. Bivariate relationship between external attribution of QRP use and QRP stigma. N = 98.

Table 15. *External attribution model regression coefficient estimates. N = 98.*

Coefficients	Estimate (ß)	Estimate (*b*)	Std. Error	t value	p value
(Intercept)	---	10.35	1.597	6.48	<0.001 ***
external attribution	0.1449	0.474	0.334	1.42	0.16
gender	-0.0306	-0.103	0.392	-0.26	0.79
ethnicity	-0.1176	-0.285	0.286	-1	0.32
politics	-0.0559	-0.109	0.189	-0.58	0.57

There was insufficient evidence to determine if external attribution was a significant predictor of stigma.

Model 5: Fear. The model of fear regressed on stigma is depicted in Figure 18 and described in Table 16 below. Diagnostic plots can be found in Appendix 8.

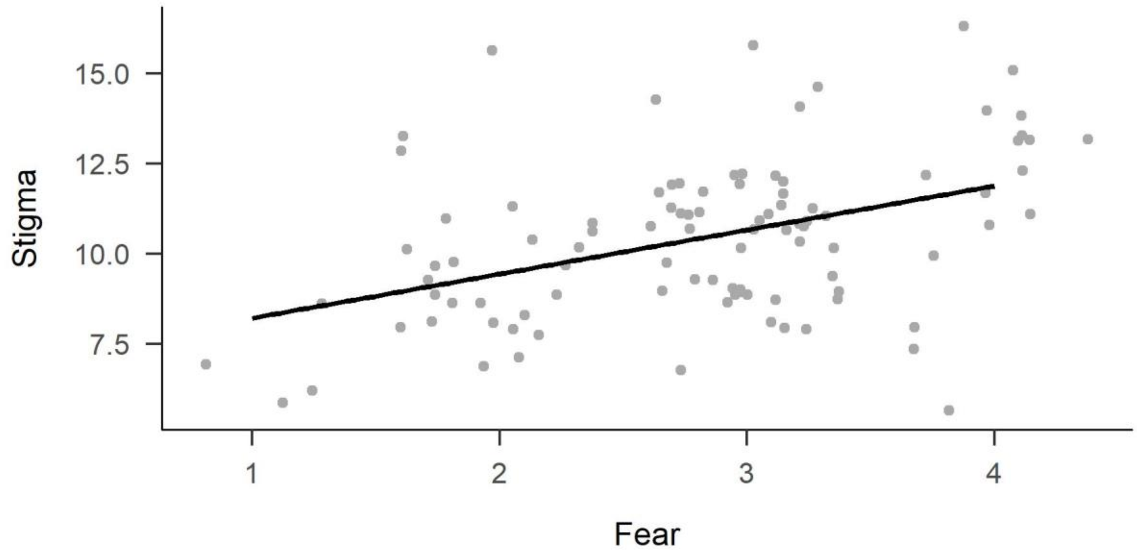


Figure 18. Bivariate relationship between fear of QRP users and QRP stigma. N = 98.

Table 16. *Fear model regression coefficient estimates. N = 98.*

Coefficients	**Estimate (ß)**	**Estimate (*b*)**	**Std. Error**	**t value**	**p value**
(Intercept)	---	7.4884	1.5955	4.69	<0.001 ***
fear	0.4048	1.1873	0.2867	4.14	<0.001 ***
gender	-0.0353	-0.1193	-0.3636	-0.33	0.74
ethnicity	-0.0273	-0.0662	-0.2714	-0.24	0.81
politics	-0.0111	-0.0201	-0.1751	-0.11	0.91

Fear of QRP users was a significant predictor of stigma. Participants who more strongly believed that QRP users were a threat to their students held greater stigmatizing views of QRP users ($b = 1.19$, $p < 0.001$).

Model 6: Power (linear). The linear model of power (operationalized as individual income) on stigma is depicted in Figure 19 and described in Table 17 below. Diagnostic plots can be found in Appendix 9.

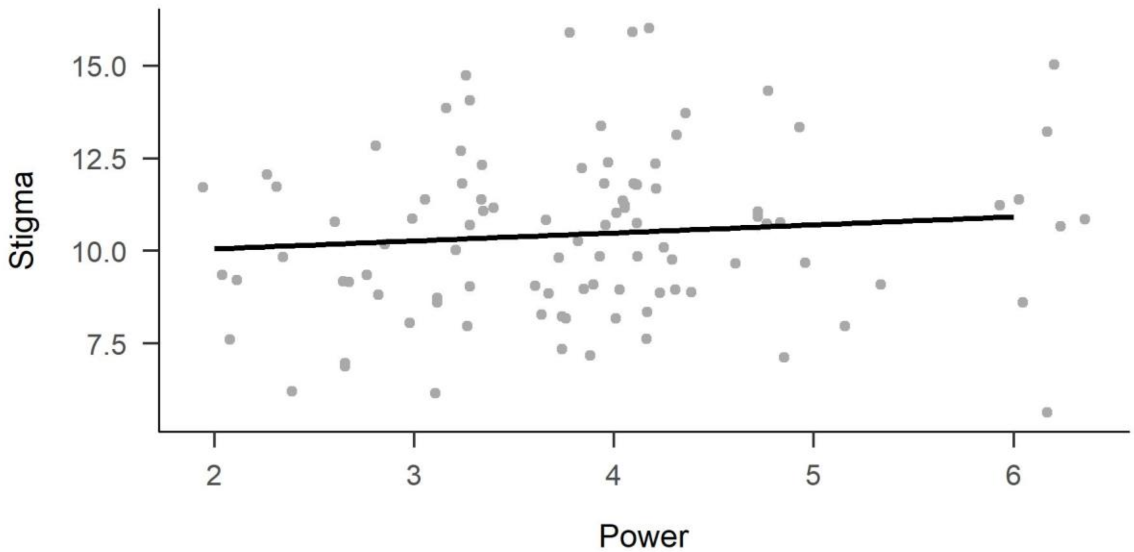


Figure 19. Bivariate relationship between power and QRP stigma. N = 98.

Table 17. *Power model regression coefficient estimates. N = 98.*

Coefficients	Estimate (ß)	Estimate (*b*)	Std. Error	t value	p value
(Intercept)	---	10.3201	1.7214	6	<0.001 ***
power	0.1226	0.2692	0.2311	1.16	0.25
gender	-0.0452	-0.1525	0.3935	-0.39	0.7
ethnicity	-0.1439	-0.3492	0.2893	-1.21	0.23
politics	-0.0206	-0.0375	0.1924	-0.19	0.85

There was insufficient evidence to determine if power was a significant predictor of stigma. However, it is possible that the relationship between power and stigma is curvilinear instead of linear, as tested above. This possibility was tested in Model 7.

Model 7: Power (curvilinear). The curvilinear model of power on stigma is depicted in Figure 20 and described in Table 18 below. Diagnostic plots can be found in Appendix 10.

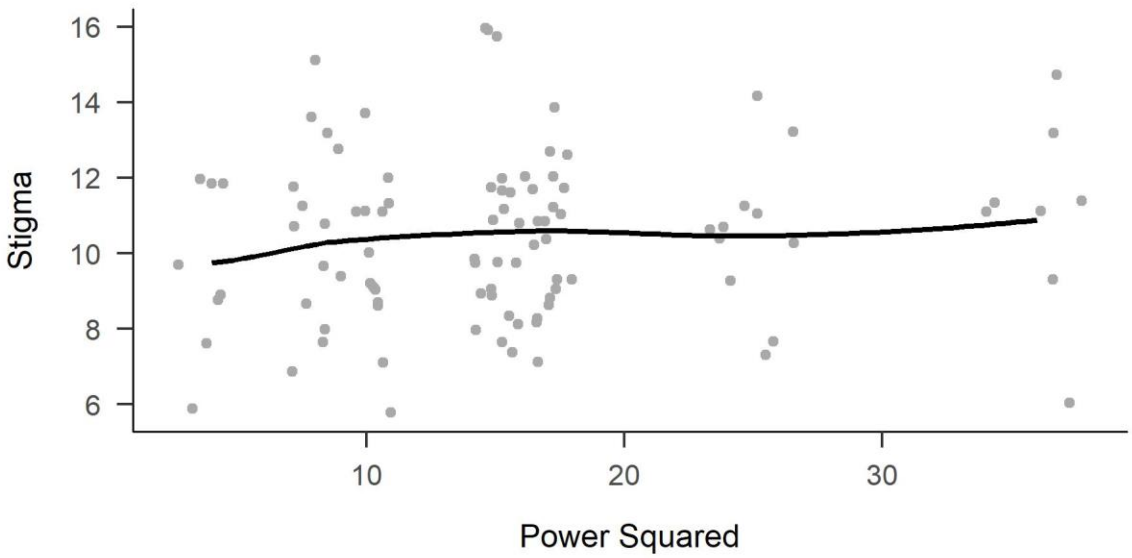


Figure 20. Bivariate relationship between power[2] and QRP stigma. N = 98.

Table 18. *Power model regression coefficient estimates, including power squared. N = 98.*

Coefficients	Estimate (ß)	Estimate (*b*)	Std. Error	t value	p value
(Intercept)	---	9.2536	2.9796	3.11	0.0025 **
power	0.3858	0.8469	1.3348	0.63	0.5274
power squared	-0.2672	-0.0722	0.1643	-0.44	0.6613
gender	-0.0432	-0.1458	0.3956	-0.37	0.7133
ethnicity	-0.1457	-0.3535	0.2907	-1.22	0.227
politics	-0.0221	-0.0402	0.1933	-0.21	0.8356

The line of best fit in Figure 20 was produced by LOWESS (locally weighted scatterplot smoothing), and although there are some minor deviations from linearity, there was not enough evidence to determine if there was a significant curvilinear relationship between power and stigma.

Beyond the bivariate relationships, it is important to consider the frequency of participant responses. Figure 21 depicts the prevalence of agreement with the independent measures used in the previous regression models, and these data are further described in Table 19.

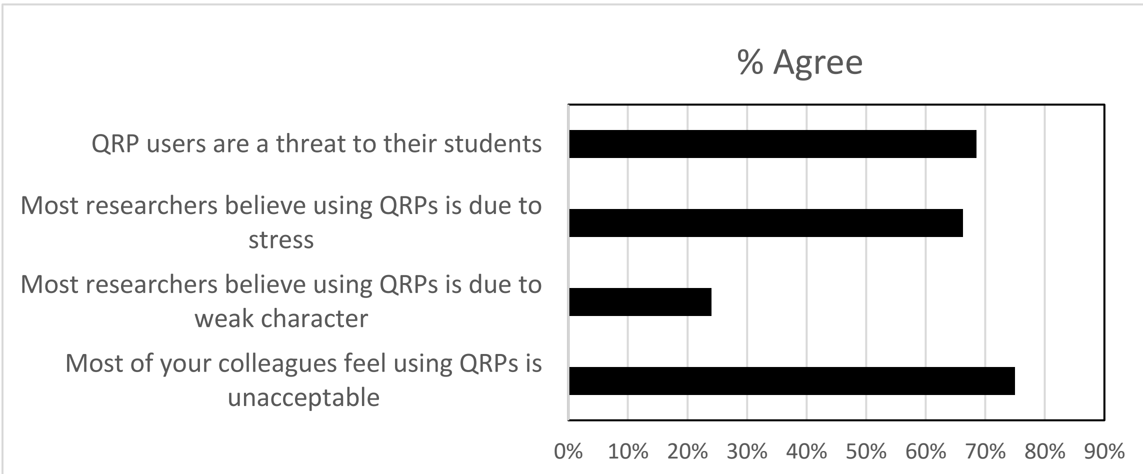


Figure 21. Prevalence of agreement with independent measures. N = 98.

Table 19. *Prevalence of agreement with independent measures by theoretical domain. N = 98.*

Domain	**Item**	**% Agree**
Acceptability	Most of your colleagues feel using QRPs is unacceptable	75.0%
Internal Attribution	Most researchers believe using QRPs is due to weak character	24.0%
External Attribution	Most researchers believe using QRPs is due to stress	66.2%
Fear	QRP users are a threat to their students	68.5%

Although internal attribution was a significant and positive predictor of stigma (see Figure 16 and Table 14), only a small number of participants agreed that QRP use could be internally attributed (24.0%). Most participants agreed that QRP use could be externally attributed (66.2%). Similarly, most participants agreed that QRP use broke social norms (75%) and that QRP use was threatening to students (68.5%).

Stigma-related behaviors. To assess whether QRP-using psychologists behave in ways predicted by social stigma theory, two behaviors were assessed: social withdrawal and selective information transmission.

Social withdrawal. The average social network size for the general population of psychologists was 184.93 individuals. The average social network size for QRP using psychologists was 178.60 individuals. There was insufficient evidence to determine if this difference was significant, $t(70) = -0.2$, $p = 0.8$. See Figure 22 for a density plot of social network sizes for all participants.

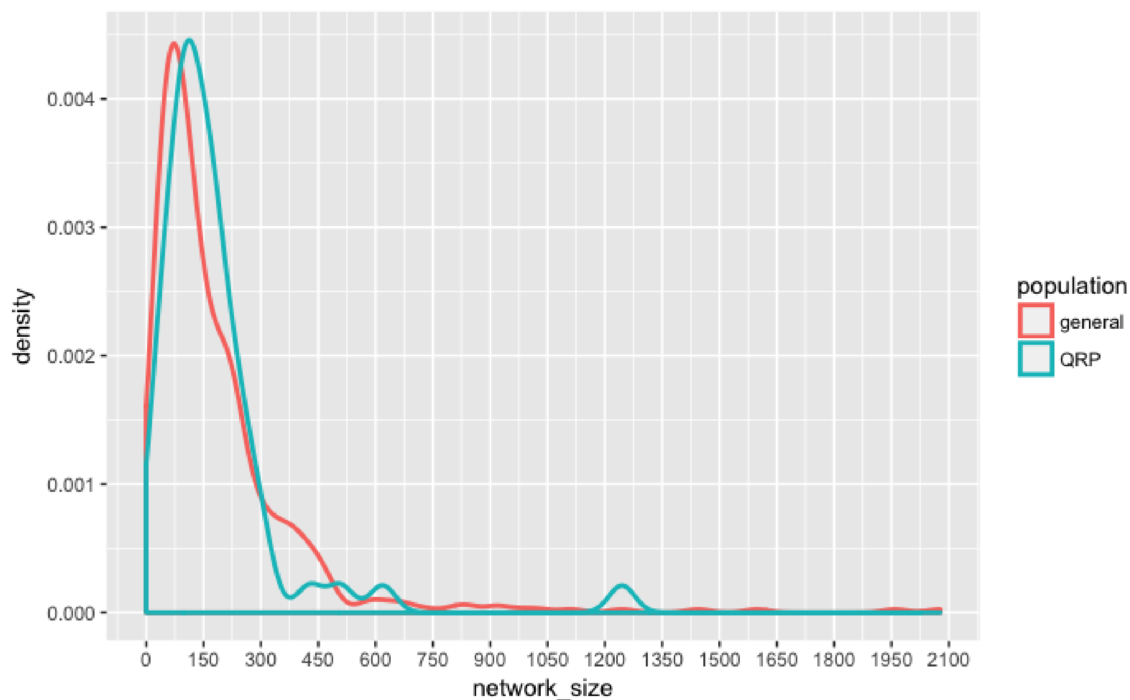


Figure 22. Frequency of participants' social network sizes. The blue line represents QRP users (N = 56) and the red line represents the general population of psychologists (N = 531). Scaled density plot.

Selective Transmission. The 56 QRP users in this study produced a total of 1,230 alters from the game of contacts procedure (described previously, Chapter 3 page 54). One hundred of these alters were considered in-group members. These were alters that were identified as QRP users by participants in this study who self-identified as QRP users. The other 1,130 alters were out-group members, or psychologists with an unknown QRP-use status to the 56 QRP users in this study.

Participants, or egos, were asked for each alter whether or not that person knew of the participant's QRP-user identity status (either "this person knows I have used a QRP in

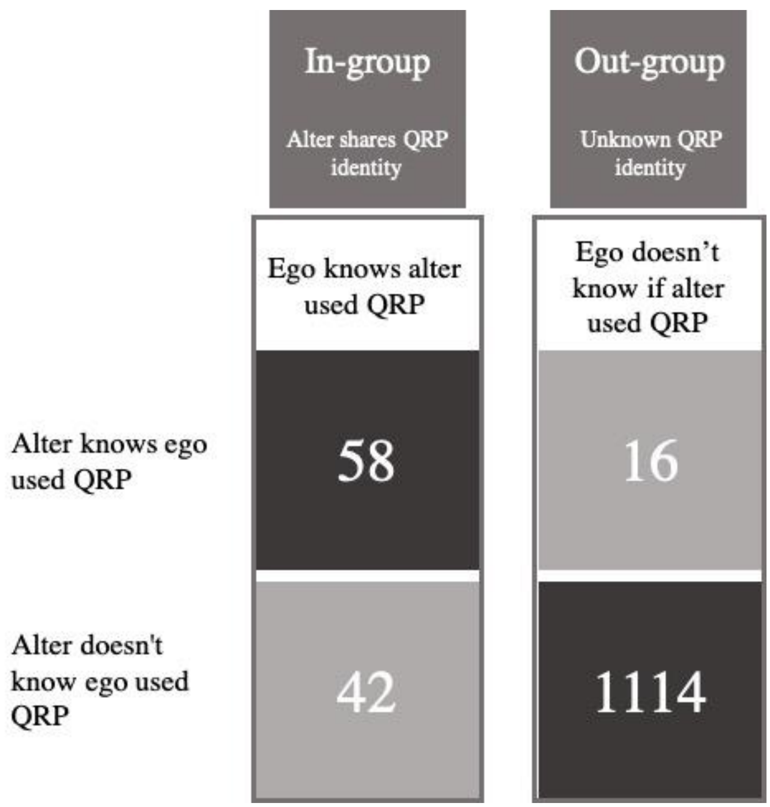


Figure 23. Frequencies of alters knowing an ego's QRP-use identity status, split by whether the alter is known to be a QRP user by the ego. The left column represents when the ego and alter are both QRP users; they are in-group members. The right column represents when the ego is a QRP user and the alter has an unknown QRP-use status; they are out-group members. A total of 1,230 alters were produced from 56 QRP user egos.

the past 12 months" or "I do not know if this person knows I have used a QRP in the past 12 months"). The counts of these responses are depicted in Figure 23.

As seen in Figure 23, 58 out of 100 in-group alters generated know the ego's QRP-use identity (58%). Conversely, when the alter's QRP-use status is unknown to the ego, only 16 out of 1,130 alters generated know of the ego's QRP-use identity (1.44%). This produces an odds ratio of 96.14 [51.03, 181.15], indicating that the odds of an in-group alter knowing the ego's QRP-use status is 96.14 higher compared to out-group alters. This provides evidence of selective transmission to in-group members over out-group members.

Discussion

This study focused on the relationships between groups of research psychologists and whether QRP-using psychologists were stigmatized by their peers. All analyses except those focused on power (model 6 and model 7) support the hypothesis that QRP-users are a stigmatized subpopulation of psychologists.

In model 1 there was a significant negative relationship between stigmatizing views against QRP users and age. Younger psychologists were more willing to express stigma against QRP users compared to older psychologists. While age does not factor specifically into a larger theory of stigma, this result may stem from the relationship between early-career researchers, who are typically untenured, and older, tenured psychologists.

Everett & Earp (2015) conceptualized the current replication crisis as a social dilemma, meaning the good of the many (a credible and robust field of science) opposes the good of the individual (generating high impact research and achieving career security through tenure). The current replication crisis has generated an extra burden for early-career researchers, who are stuck having to choose between doing what is best for their personal career ("publish or perish") and what is best for the field (performing or contributing to replication attempts). As discussed in Chapter 3, QRP use has generated increased false-positive findings in the published literature and has contributed to failed replications. It is possible the stigmatizing views of the younger participants on QRP users observed in this study stem from this social dilemma.

Models 2, 3, 4, 5, 6 and 7 tested four domains of stigma outlined in Stuber et al. (2008): acceptability, attribution, fear, and power.

Acceptability deals with the social norms of the group, with behaviors that are outside of the social norms considered unacceptable (Bernstein, Galea, Ahern, Tracy, & Vlahov, 2007; Stuber et al., 2008). In model 2, there was a significant positive relationship between views of QRP unacceptability and QRP stigma. Higher belief of QRP use breaking social norms was associated with higher stigma against QRP users. This is consistent with the societal use of social norms to extract conformity from its constituents (Phelan, Link, & Dovidio, 2014). Stigmatizing views against those breaking social norms may also be used as an example to other society members of the consequences of breaking norms, furthering conformity (Erikson, 1966).

Attribution is the way people attempt to understand and explain reasons for their own behavior and the behavior of others. Personality or dispositional reasons for behavior are considered internally attributed. Environmental or situational reasons for behavior are considered externally attributed (Tetlock, 1985). Attribution theory has been previously applied to stigma. In mental health stigma, for conditions with symptoms seen as uncontrollable, the individual with mental illness is deemed not responsible and leads to more helping behaviors and fewer punishing behaviors (Boysen, 2008). Model 3 and model 4 tested internal and external attribution as predictors of QRP stigma. As seen in Model 3, those participants who more strongly believed QRP use was internally attributed ("due to weak character") showed significantly more stigma against QRP users. Conversely, there was no significant relationship observed between external attribution ("QRP use is due to stress") and QRP stigma. This finding suggests that QRP users are stigmatized, at least by those who believe QRP use is internally attributed.

Model 5 tested the relationship between fear of QRP users and stigma against them, operationalized as participants believing QRP users are a threat to their students. Fear has been shown to contribute to stigmatizing attitudes towards individuals with mental illnesses (Link, Phelan, Bresnahan, Stueve, & Pescosolido, 1999) and smokers (Stuber et al., 2008), as well as those with health conditions such as leprosy (Bainson & Van Den Borne, 1998) and HIV/AIDS (Herek & John, 2002). Fear was a significant predictor of QRP stigma, with those who most strongly agreed that QRP users were a threat to their students showing significantly more stigma against QRP users compared to those who did not.

Finally, theories of stigma propose that it is not possible to fully stigmatize a particular group unless they lack social, economic, or political power relative to those doing the stigmatizing (Link & Phelan, 2001; Stuber et al., 2008). The present study operationalized power economically, and investigated whether those academics with higher salary displayed stronger stigmatizing attitude towards QRP users. Model 6 tested the linear relationship between power and QRP stigma and did not find a statistically significant relationship.

It is also possible that the relationship between economic power and stigma is not linear, but parabolic. Those individuals with the lowest and highest power could hold similar stigmatizing attitudes towards QRP users, while those more central in the power scale may be less stigmatizing. Model 7 tested this relationship between power and QRP stigma, but did not find a statistically significant relationship.

Taken together, these models suggest that QRP-using psychologists are stigmatized by the general population of psychologists. QRP users are seen as breaking social norms and are feared as a threat to their students, and when QRP use is internally attributed, stigmatizing attitudes are higher. However, when asked directly, most participants agreed that QRP use was more attributable to external variables (like stress, see Figure 21) than internal variables (like weak character).

There are a couple of potential reasons why power was not a significant predictor of QRP stigma in this study. It could be that economic power is a poor operationalization of power in the academic social environment. It is possible instead that the number of published papers, citation count, tenure, or years in a prestigious position could serve as better proxies of power in the academic social setting than income. It could also be that there is no difference in power between QRP users and the general population of psychologists. Academia is unlike the typical social environment in some key ways. For instance, success as an academic psychologist has relied more and more on working with others. Collaboration rates in psychology have been rising over the past 90 years (Zafrunnisha & Pullareddy, 2009), and this selective pressure to collaborate may serve as a vehicle for high income and lower income academics to intersect. The academic model is also based on a mentor-mentee relationship, where professors who make an adequate salary often closely work with graduate students, who are either unpaid, paid a modest stipend, or are economically insecure (Ehrenberg & Mavros, 1992). Academia may not support a social environment where those of higher economic power can stigmatize those of lower economic power.

Beyond just investigating the attitudes of the general population of psychologists on QRP users, this study also directly observed stigma-related behaviors of QRP users themselves. This is a step forward in determining if QRP users are stigmatized because we can ask the question "Do QRP users act like other stigmatized groups?". There were two stigma-related behaviors observed in this study: social withdrawal and selective disclosure.

Figure 22 shows the comparison in social network size between QRP users and the general population of psychologists. Although QRP users have a slightly smaller average social network size (178.6 versus the general population of psychologist's average social network size of 184.93), this difference was not statistically significant. Here too, it is possible that the nature of academic psychology inhibits QRP users from socially withdrawing. As mentioned previously, success as a psychologist has relied more and more on collaboration, therefore restricting one's academic social network directly inhibits success. This outcome may also be due to selection bias, where those QRP-using psychologists who have socially withdrawn no longer found success in academia and moved on to other careers.

The other stigma-related behavior directly observed was selective transmission of QRP-use identity. Figure 23 shows the number of people in QRP users' social networks that either do or do not know about that person's QRP-use identity, given that the social network member either is or isn't a QRP user themselves. When the QRP identity of a social network member is unknown, it is highly unlikely they know about the QRP user identity of the QRP user. However, if the social network member is known to be a QRP user, there is a much greater chance they know the QRP use identity of the QRP user. This means that QRP users selectively disclose their QRP use to other known QRP users.

Revealing is one significant way individuals can manage an invisible social identity (Goffman, 1963). Being stigmatized is harmful, as it can lead to stereotyping, loss of status, and discrimination (Clair, Beatty, & Maclean, 2005; Link & Phelan, 2001). By selectively revealing an invisible stigmatized identity to in-group members (in this case, other QRP users), one can avoid the harmful effects of stigmatization while minimizing the negative consequences of keeping one's identity a secret (Garcia & Crocker, 2008; Ilic et al., 2014).

Limitations. This study had several limitations. First, it was a survey-based study that relied on self-reports from tenured or tenure-track psychologists. Self-reports are known for to be sensitive to social desirability bias, where participants provide socially desirable responses to study questions (R. J. Fisher, 1993). This bias has the potential to either inflate and attenuate the observed results. Future studies on the perceptions of psychologists on QRP users may benefit from indirect measures to avoid this source of bias.

Another limitation of this study is the relatively small number of participants. This study was run along with the study detailed in Chapter 3. To avoid participants being primed about their own QRP use prior to the measures used in this survey, only participants who received Survey 1 in that study were able to participate in the current study. Out of the 7,101 American tenured or tenure-track psychologists that could have potentially participated, only 1,775 were solicited to participate in this study. Future studies will benefit from utilizing the larger population from which to sample.

Finally, this study did not investigate the perceived stigmatization experienced by QRP users. This study conceptualized stigma as an external force that is exerted on individuals, and asked if that force existed or not. This is consistent with Goffman's framework of stigma as a process by which the reaction of others spoils an individual's identity (Goffman, 1963). However, there is a significant literature on *perceived* stigma, which is defined more as the experience of stigma by an individual and less as the stigma exerted by the social environment (Mickelson, 2001). There is evidence that individuals with the same stigmatized identity vary in the amount of stigma they perceive. It is possible that those most socially stigmatized may perceive the least amount of stigma (Crandall, 1991). Future studies may want to investigate how QRP users perceive stigma in addition to the social environmental stigma that exists.

Implications. Since the 1980s, psychologists and other social sciences have identified with those victimized by stigma: those with mental illnesses, drug users, the obese, and others traditionally marginalized by social norms. This may have partially been due to the HIV/AIDS epidemic, where stigma against homosexual and bisexual men obscured a catastrophic public health crisis that claimed tens of thousands of lives (Bayer, 2008). Simultaneously, however, stigma was being used purposefully to combat the public health risks associated with smoking. By the mid-1980s, most U.S. states had already imposed limits on public smoking, and since then a process of denormalization has "pushed tobacco use out of the charmed circle of normal, desirable practice to being an abnormal practice" (Bayer, 2008; California Department of Health Services, 1998).

In better understanding the social relationships that exist between QRP using psychologists and the greater general population of psychologists, the field can better tackle the ongoing replication crisis. It is possible that increasing stigma on QRP users may lead to a reduction in QRP users, similar to the denormalization of smokers in the past 30 years. It is also possible that reducing QRP stigma may allow QRP users to reveal their identity and open a dialog to increase adoption of best practices in performing and reporting scientific studies. Either way, knowing that QRP users are a stigmatized subgroup of psychologists is a first step in understanding the social dynamics that play a part in the current replication crisis.

Conclusion. Psychologists who use questionable research practices are stigmatized by the general population of psychologists. This is evidenced by negative perceptions of QRP users: they are seen as breaking social norms and they are considered a threat to their students. Additionally, stigmatizing attitudes towards QRP users is a function of attribution. Those psychologists that believe QRP use is internally attributed (i.e., QRP use is due to weak character) hold more stigmatizing attitudes towards QRP users than those who believe QRP use is externally attributed (i.e., QRP use is due to stress).

Beyond just the attitudes of the general population of psychologists, this study directly observed the behaviors of QRP users and found QRP users behaved similar to other stigmatized populations. QRP users selectively disclosed their stigmatized identity to those who shared this identity. In other words, QRP users conceal their identity to those whose identity is unknown, but reveal their identity to those known to be QRP users.

This study demonstrated that QRP users are a stigmatized subgroup of scientists within the larger population of psychologists. Moving forward from the current replication crisis, knowing this about QRP users and their social environment can help to better craft interventions that minimize the use of questionable research practices.

Chapter 5:

General Discussion and Conclusion

The previous three chapters assessed the role of the researcher in the process of research. Idealistically, a scientist is an objective observer of truth in the world, who takes note of what they see and reports it for the betterment of all. Realistically, however, a scientist is a human with a variety of wants and needs, living in a world that rewards hard work and measurable productivity. Because of this, the researcher may not always be an objective observer, and may influence their subject of observation. The three studies reported in this dissertation assessed the role of the researcher in two domains: the consumption of knowledge, operationalized as the search for and downloading of journal articles, and the production of knowledge, operationalized as the production of journal articles describing novel empirical research.

In Chapter 2, metadata associated with a digital journal article precipitated social herding: more researchers chose to download articles that had a higher number of previous downloads. While it is true that many researchers would read from the same small pool of journals in the past (Chen, 1974), this per-article level of granularity is a feature of the academic literature being digitized, and has the potential to limit the universe of scientific literature read by scientists. Furthermore, the journal article that becomes the most downloaded is not constant over multiple replicates. The most downloaded article varied across the four replicate worlds with download count metadata in the study detailed in Chapter 2. Finally, participants subjectively rated journal article abstracts as higher quality in the presence of metadata. Thirty-four out of forty-eight abstracts were rated higher when the study condition included download counts, suggesting that knowing other scientists have downloaded a paper increases the perceived quality of a paper's abstract.

While Chapter 2 focused on the domain of knowledge consumption, Chapter 3 and Chapter 4 focused on the domain of knowledge production, specifically the use of questionable research practices when collecting, analyzing, and disseminating research results. These practices have been a primary culprit for creating the large number of false-positive findings in the published literature. The study detailed in Chapter 3 sought to measure the size of the QRP-using population, and did this with three different estimators. Of the two estimates that were significantly different from zero, one estimated that 18.18% of American psychologists have used a QRP in the past 12 months, and the other estimated 24.4%. Since we know scientists have used QRPs in the past (Leslie K John et al., 2012; Motyl et al., 2017), this finding is an important benchmark for the size of the current QRP using population. In order to know if interventions to reduce QRP use are effective, knowing how many researchers used QRPs is necessary.

Chapter 4 took a closer look at how psychologists viewed QRP users and how QRP users managed their identity and asked if QRP use was stigmatized. Based on the negative attitudes psychologists held towards QRP users, and based on the observation that QRP users share their identity with other QRP users but not ambiguous peers, QRP use was considered a stigma. This is again important to know for the generation of effective interventions aimed to reduce the number of QRP users. By knowing that QRP use is stigmatized, interventions can strive to either decrease stigma and provide a path for QRP users to reveal their identities and reform their research behaviors, or increase stigma to limit the number of researchers who believe QRPs are acceptable research practices.

Overall, the results show that researchers are not completely removed from the subject of their research. As this is the case, it is critical that academics reflect on their role in the scientific process and how their motivations, biases, and shortcomings may shape what they observe, report and ultimately archive as knowledge.

Weaknesses and Strengths

All of the reported studies used academic researchers as subjects. This is an extremely unique population to study, as they are keenly aware of experimental procedures, manipulations, and are generally trained to be critical and inquisitive. It is possible that participants attempted to uncover the motivation underlying some of the studies and altered their responses accordingly. It is also possible that many chose not to participate as a function of their position as an academic, resulting in selection bias. Strictly speaking, one cannot generalize from a non-random sample to a population, as biases in sampling or participants could result in non-representative findings. However, given the exploratory nature of this work and the relative lack of research on this population, this work represents a strong first step in the metascientific investigation of researchers. Of course, this work will only become stronger with larger, random samples. Future studies would benefit from longitudinal approaches to both track the size of the QRP user population over time and to observe trends in literature consumption. Finally, all of these studies were restricted to psychologists. While this is a strength it that it allows for a focused analyses of one field of inquiry, future studies would benefit from expanding the sample criteria to include other research fields. Although the current "replication crisis" is centered on psychology, it is by no means the only scientific field with these issues.

Additionally, for Chapters 3 and 4, the population of interest was limited to tenure or tenure-track psychologists associated with an American PhD-granting psychology department. Given the heterogenous nature of academia, this frame does not include many other psychologists, such as those at teaching-focused institutions, graduate students, post-doctoral researchers, or psychologists in other countries. Since graduate students and post-docs tend to be closer to the process of data collection and analysis, it is possible that excluding these populations led to an underestimate in the number of current QRP users, for example. It is also possible that attitudes towards QRP users may be different in these populations, and their inclusion in future studies will help further elucidate the social environment that exists between career psychologists.

These studies have multiple other strengths. One of these strengths is the naturalistic experiment design of the literature search study detailed in Chapter 2. While "download count" could have been a researcher modified independent variable, allowing this variable to be generated by the actual participants more realistically modeled how academics actually leave their imprint on what they read, and how that imprint further manipulates the literature consumption of others. This design also allowed for the ability to measure unpredictability of success, a measure that could not be measured otherwise. Perhaps the greatest strength of this dissertation is the multidimensional approach to understanding the different ways in which researcher influence their research. It is not only at the data analysis stage, but in how they treat other researchers which in turn influences the now-digitized and metadata rich published literature. The relationships are a complicated tangle, and while there is much more to learn, this dissertation begins to scratch the surface of understanding the intricate relationship between researcher and research.

Future Directions and Conclusion

This dissertation lays the groundwork for many exciting future studies. One area that needs further study is the content of journal abstracts and associated metadata. Scholars rely on tools such as Google Scholar, PubMed, and Web of Science, all of which provide different sets of metadata along with scientific articles. Additionally, alternative metrics such as the number of social media mentions, blog posts, and news articles add to the complex stimuli available to researcher when they find an academic paper. There is great potential in better understanding this digital environment and the role it plays on literature decision making. Another potential future direction is in the testing of different interventions that either increase or decrease stigmatization of QRP users. It is unclear which potential avenue would produce the greatest reduction in QRP use. Developing interventions that can lead to long-lasting reductions in QRP-use will be critical to produce a radical shift in academic culture.

Metascience is a small but growing area of inquiry. This dissertation highlights the importance of understanding the scientist as a means of better understanding the science. Robert Oppenheimer once said,

> *"we do not believe any group of men adequate enough or wise enough to operate without scrutiny or without criticism. We know that the only way to avoid error is to detect it, that the only way to detect it is to be free to enquire. We know that the wages of secrecy are corruption. We know that in secrecy error, undetected, will flourish and subvert"*

Oppenheimer, 1951.

In this time of unrest and change within the scientific community, taking the researcher's critical eye and turning it inward to detect the errors and to rout the corruption will only serve to better the craft and increase trust in published results. This work is a first step in the direction of a more accurate base of psychological knowledge.

References

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS ONE*, *12*(3), 1–17. https://doi.org/10.1371/journal.pone.0172792

Ahrnsbrak, R., Bose, J., Hedden, S., Lipari, R., & Park-Lee, E. (2017). Key Substance Use and Mental Health Indicators in the United States: Results from the 2016 National Survey on Drug Use and Health. *Substance Abuse and Mental Health Services Administration*, *7*(1), 877–726. https://doi.org/10.1016/j.drugalcdep.2016.10.042

Arentoft, A., Van Dyk, K., Thames, A. D., Sayegh, P., Thaler, N., Schonfeld, D., … Hinkin, C. H. (2016). Comparing the unmatched count technique and direct self-report for sensitive health-risk behaviors in HIV+ adults. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV*, *28*(3), 370–375. https://doi.org/10.1080/09540121.2015.1090538

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership and Men*, 119–124. https://doi.org/10.1017/CBO9781107415324.004

Atkinson, A. B. (1970). On the Measurement of Inequality. *Journal of Economic Theory*, *2*, 244–263.

Bainson, K. A., & Van Den Borne, B. (1998). Dimensions and process of stigmatization in leprosy. *Leprosy Review*, *69*, 341–350.

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. https://doi.org/10.1177/1745691612459060

Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Editorial : Evidence on Questionable Research Practices : The Good , the Bad , and the Ugly. *Journal of Business and Psychology*, *31*(3), 323–338. https://doi.org/10.1007/s10869-016-9456-7

Bayer, R. (2008). Stigma and the ethics of public health: Not can we but should we. *Social Science and Medicine*, *67*(3), 463–472. https://doi.org/10.1016/j.socscimed.2008.03.017

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. https://doi.org/10.1037/a0021524

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., & Johnson, V. E. (2017). Redefine Statistical Significance. *PsyArxiv*, (July 22), 1–18. https://doi.org/10.17605/OSF.IO/MKY9J

Bernard, H. R., Hallett, T., Iovita, A., Johnsen, E. C., Lyerla, R., McCarty, C., … Stroup, D. F. (2010). Counting hard-to-count populations: the network scale-up method for public health. *Sex Transm Infect*, *86 Suppl 2*, ii11-5. https://doi.org/10.1136/sti.2010.044446

Bernstein, K. T., Galea, S., Ahern, J., Tracy, M., & Vlahov, D. (2007). The built environment and alcohol consumption in urban neighborhoods. *Drug and Alcohol*

Dependence, *91*(2–3), 244–252. https://doi.org/10.1016/j.drugalcdep.2007.06.006

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's Threat to Organizational Research: Evidence From Primary and Meta-Analytic Sources. *Personnel Psychology*, *69*(3), 709–750. https://doi.org/10.1111/peps.12111

Bourdieu, P. (1988). *Homo Academicus*. Stanford, CA: Stanford University Press.

Boysen, G. U. Y. A. (2008). EDUCATION AND MENTAL HEALTH STIGMA : THE EFFECTS OF ATTRIBUTION , BIASED ASSIMILATION , AND ATTITUDE POLARIZATION. *Journal of Social and Clinical Psychology*, *27*(5), 447–470.

Brewer, D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, *22*(1), 29–43. https://doi.org/10.1016/S0378-8733(99)00017-9

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Cabin, R. J., & Mitchell, R. J. (2000). To Bonferroni or Not to Bonferroni: When and How Are the Questions. *Bulletin of the Ecological Society of America*, *81*(3), 246–248. Retrieved from http://www.jstor.org/stable/20168454

California Department of Health Services. (1998). *A Model for Change: the California experience in tobacco control*.

Canty, A., & Ripley, B. D. (2017). boot: Bootstrap R (S-Plus) Functions.

Chapple, A., Ziebland, S., & Mcpherson, A. (2004). Stigma, shame and blame experienced by patients with lung cancer: qualitative study. *BMJ*, *online fir*(June), 1–5. https://doi.org/10.1136/bmj.38111.639734.7C

Chen, C. (1974). How Do Scientists Meet Their Information Needs? *Special Libraries*, *65*(7), 272–280.

Clair, J. A., Beatty, J. E., & Maclean, T. L. (2005). Out of sight but not out of mind: Managing invisible social identities in the workplace. *Academy of Management Review*, *30*(1), 78–95. https://doi.org/10.5465/AMR.2005.15281431

Connelly, N. A., & Brown, T. L. (1995). Use of Angler Diaries to Examine Biases Associated with 12-Month Recall on Mail Questionnaires. *Transactions of the American Fisheries Society*, *124*(3), 413–422. https://doi.org/10.1577/1548-8659(1995)124<0413:uoadte>2.3.co;2

Corrigan, P. (2000). Mental Health Stigma as Social Attribution : Implications for Research Methods and Attitude Change Mental Health Stigma as Social Attribution : Implications for Research Methods and Attitude Change. *Clinical Psychology Science and Practice*, *7*(1), 48–67. https://doi.org/10.1093/clipsy.7.1.48

Courtwright, A. (2013). Stigmatization and public health ethics. *Bioethics*, *27*(2), 74–80. https://doi.org/10.1111/j.1467-8519.2011.01904.x

Crandall, C. S. (1991). AIDS-related stigma and the lay sense of justice. *Contemporary Social Psychology*, *15*(2), 66–67.

D. Chambers, C., Feredoes, E., D. Muthukumaraswamy, S., & J. Etchells, P. (2014).

Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, *1*(1), 4–17. https://doi.org/10.3934/Neuroscience.2014.1.4

Davidson, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press.

Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A.-W., Cronin, E., … Williamson, P. R. (2013). Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias - An Updated Review. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0066844

Ehrenberg, R. ., & Mavros, P. G. (1992). *Do Doctoral Students' Financial Support Patterns Affect Their Times-To-Degree and Completion Probabilities*.

Elms, A. C. (1975). The crisis of confidence in social psychology. *American Psychologist*, *30*(10), 967–976. https://doi.org/10.1037/0003-066X.30.10.967

Erikson, K. T. (1966). *Wayward Puritans: A Study in the Sociology of Deviance*. New York, NY: John Wiley & Sons, Ltd.

Everett, J. A. C., & Earp, B. D. (2015). A tragedy of the (academic) commons : interpreting the replication crisis in psychology as a social dilemma for early-career researchers, *6*(August), 1–4. https://doi.org/10.3389/fpsyg.2015.01152

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

Fanelli, D. (2018). Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences of the United States of America2*, *in press*, 1–4. https://doi.org/10.1073/pnas.1708272114

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, *7*(1), 45–52. https://doi.org/10.1177/1948550615612150

Fisher, R. A. (1925). *Statistical methods for research workers*. *Biological monographs and manuals*. https://doi.org/10.1056/NEJMc061160

Fisher, R. J. (1993). Social desirability bias and the vailidity of indirect questioning. *Journal of Consumer Research*, *20*(September 1993), 303–315. https://doi.org/10.1086/209351

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*(6), 889–906. https://doi.org/10.1037/0022-3514.38.6.889

Fiske, S. T. (2016). Mob Rule or Wisdom of Crowds. *APS Observer*.

Forsberg, E. M., Anthun, F. O., Bailey, S., Birchley, G., Bout, H., Casonato, C., … Zöller, M. (2018). Working with Research Integrity—Guidance for Research Performing Organisations: The Bonn PRINTEGER Statement. *Science and Engineering Ethics*, 1–12. https://doi.org/10.1007/s11948-018-0034-4

Forte, M. (2015). *Pierre Bourdieu's Homo Academicus: Objectifying the Academic Object*. Concordia University.

Garcia, J. A., & Crocker, J. (2008). Reasons for disclosing depression matter: The consequences of having egosystem and ecosystem goals. *Social Science and Medicine*, *67*(3), 453–462. https://doi.org/10.1016/j.socscimed.2008.03.016

Garfield, A. E. (1996). What Is The Primordial Reference For The Phrase â€™ Publish Or Perish '? *The Scientist*, *10*(12), 11.

Gelman, A., & Loken, E. (2014). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Psychological Bulletin*, *140*(5), 1272–1280. https://doi.org/dx.doi.org/10.1037/a0037714

Gervais, W. M., & Najle, M. B. (2017). How many atheists are there? *Social Psychological and Personality Science*, 1948550617707015.

Goffman, E. (1959). The Presentation of Self in Everyday Life. *Garden City NY*. https://doi.org/10.2307/258197

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Simon and Schuster.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate Data Analysis* (3rd ed). New York, NY: Macmillan Publishing Company.

Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, *107*(2), 455–476. https://doi.org/10.1007/s11192-016-1849-x

Herek, G. M., & John, P. (2002). HIV-Related Stigma and Knowledge in the United States : Prevalence and Trends , 1991 – 1999. *American Journal of Public Health*, *92*(3), 1991–1999.

Hinsley, A., Nuno, A., Keane, A., St, F. A. V, & Ibbett, H. (2019). Asking sensitive questions using the unmatched count technique : Applications and guidelines for conservation, *2019*(November 2018), 308–319. https://doi.org/10.1111/2041-210X.13137

Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews*, (1). https://doi.org/10.1002/14651858.MR000006.pub3

Hullett, C., Park, H. S., & Lindsey, L. L. M. (2008). A Critical Assessment of Null Hypothesis Significance Testing in Quantitative Communication Research, *34*(May 2003), 171–187. https://doi.org/10.1111/j.1468-2958.2008.00317.x

Ilic, M., Reinecke, J., Bohner, G., R??ttgers, H. O., Beblo, T., Driessen, M., … Corrigan, P. W. (2014). Managing a stigmatized identity-evidence from a longitudinal analysis about people with mental illness. *Journal of Applied Social Psychology*, *44*(7), 464–480. https://doi.org/10.1111/jasp.12239

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), 0696–0701. https://doi.org/10.1371/journal.pmed.0020124

Jing, L., Qu, C., Yu, H., Wang, T., & Cui, Y. (2014). Estimating the sizes of populations at high risk for HIV: A comparison study. *PLoS ONE*, *9*(4), 1–6. https://doi.org/10.1371/journal.pone.0095601

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. https://doi.org/10.1177/0956797611430953

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. https://doi.org/10.1177/0956797611430953

Killworth, P. D. P., Johnsen, E. C. E., Bernard, H. R. H., Ann Shelley, G., & McCarty, C. (1990). Estimating the size of personal networks. *Social Networks*, *12*(4), 289–312. https://doi.org/10.1016/0378-8733(90)90012-X

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., … Babalola, M. T. (2018). Many Labs 2 : Investigating Variation in Replicability Across Samples and Settings. https://doi.org/10.1177/2515245918810225

Krumme, C., Cebrian, M., Pickard, G., & Pentland, S. (2012). Quantifying Social Influence in an Online Cultural Market, *7*(5), 0–5. https://doi.org/10.1371/journal.pone.0033785

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., J Apps, M. A., Argamon, S. E., … Lino de Oliveira, C. (2018). Justify Your Alpha: A Response to “ Redefine Statistical Significance .” *Nature Human Behavior*, *2*, 168–171.

Landen, D. D., & Hendricks, S. (1995). Effect of recall on reporting of at-work injuries. *Public Health Reports (Washington, D.C. : 1974)*, *110*(3), 350–354. https://doi.org/10.1016/s0022-4375(97)90342-x

Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS ONE*, *10*(6), 1–15. https://doi.org/10.1371/journal.pone.0127502

Leary, M. R., Tchividjian, L. R., & Kraxberger, B. E. (1994). Self-Presentation Can Be Hazardous to Your Health : Impression Management and Health Risk, *13*(6), 461–470.

Lee, S., & Bozeman, B. (2005). The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science*, *35*(5), 673–702. https://doi.org/10.1177/0306312705052359

Link, B. G., & Phelan, J. C. (2001). Conceptualizing Stigma. *Annual Review of Sociology*, *27*, 363–385. https://doi.org/10.1146/annurev.soc.27.1.363

Link, B. G., Phelan, J. C., Bresnahan, M., Stueve, A., & Pescosolido, B. A. (1999). Public Conceptions of Mental Illness : Labels , Causes , Dangerousness , and Social Distance. *American Journal of Public Health*, *89*(9).

McCarty, C., Killworth, P., & Bernard, H. (2001). Comparing two methods for estimating network size. *Human Organization*, *60*(1), 28–39. Retrieved from http://sfaa.metapress.com/index/efx5t9gjtgmga73y.pdf

McCormick, T. H., Salganik, M. J., & Zheng, T. (2010). How many people do you know? Efficiently estimating personal network size. *J Am Stat Assoc.*, *105*(489), 59–70. https://doi.org/10.1016/j.immuni.2010.12.017.Two-stage

Mcshane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2017). Abandon

Statistical Significance, 1–12.

Mela, C. F., & Kopalle, P. K. (2002). The impact of collinearity on regression analysis : the asymmetric effect of negative and positive correlations. *Applied Economics*, *34*, 667–677.

Mickelson, K. D. (2001). Perceived Stigma , Social Support , and Depression. *Society for Personality and Social Psychology*, *27*(8), 1046–1056.

Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Mueller, A. B., … Skitka, L. J. (2017). The State of Social and Personality Science: Rotten to the Core, Not so Bad, Getting Better, or Getting Worse? *Journal of Personality and Social Psychology*.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, *341*(6146), 647–651. https://doi.org/10.1126/science.1240466

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology*, *15*(6), 1044–1045. Retrieved from http://dx.doi.org/10.1093/beheco/arh107

Newcombe, R. G. (1987). Towards a reduction in publication bias. *British Medical Journal*, *295*(September), 656–659.

Neyman, J., & Pearson, S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses, *231*, 289–337.

Nickerson, R. S. (1998). Confirmation Bias : A Ubiquitous Phenomenon in Many Guises, *2*(2), 175–220.

Nosek, B. A. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716-aac4716. https://doi.org/10.1126/science.aac4716

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Oppenheimer, J. R. (1951). The Scientist in Public Affairs: Encouragement of Science. *Bulletin of the Atomic Scientists*, *7*(1), 6–8.

Pachankis, J. E. (2007). The psychological implications of concealing a stigma: a cognitive-affective-behavioral model. *Psychological Bulletin*, *133*(2), 328–345. https://doi.org/10.1037/0033-2909.133.2.328

Pain, E. (2016). How to Keep Up with the Scientific Literature. Retrieved from http://www.sciencemag.org/careers/2016/11/how-keep-scientific-literature

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of Social Distance ? Failure to Replicate Effects on Social and Food Judgments, *7*(8). https://doi.org/10.1371/journal.pone.0042510

Perneger, T. V. (2004). Relation between online "hit counts" and subsequent citations: Prospective study of research papers in the BMJ. *Bmj*, *329*(7465), 546–547. https://doi.org/10.1109/MMIT.2008.206

Phelan, J., Link, B. G., & Dovidio, J. F. (2014). " Stigma and Prejudice : One Animal or Two ?," *67*(3), 358–367. https://doi.org/10.1016/j.socscimed.2008.03.022.

Qualtrics. (2005). Qualtrics Survey Software. Retrieved July 20, 2009, from www.qualtrics.com

Quinn, D. M., & Chaudoir, S. R. (2009). Living with a concealable stigmatized identity: The impact of anticipated stigma, centrality, salience, and cultural stigma on psychological distress and health. *Journal of Personality and Social Psychologysychology*, *97*(4), 634–651. https://doi.org/10.1037/a0015815

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R. a. (2015). Assessing the Robustness of Power Posing. *Psychological Science*, *26*(5), 653–656. https://doi.org/10.1177/0956797614553946

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

Rubinstein, R. S., Jussim, L., & Stevens, S. T. (2018). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology*, *75*(July 2017), 54–70. https://doi.org/10.1016/j.jesp.2017.11.009

Russell Bernard, H., Johnsen, E. C., Killworth, P. D., & Robinson, S. (1991). Estimating the size of an average personal network and of an event subpopulation: Some empirical results. *Social Science Research*, *20*(2), 109–121. https://doi.org/10.1016/0049-089X(91)90012-R

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, *311*(5762), 854–856. https://doi.org/10.1126/science.1121066

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, *311*(5762), 854–856.

https://doi.org/10.1126/science.1121066

Salganik, M. J., Fazito, D., Bertoni, N., Abdo, A. H., Mello, M. B., & Bastos, F. I. (2011). Assessing network scale-up estimates for groups most at risk of HIV/AIDS: Evidence from a multiple-method study of heavy drug users in Curitiba, Brazil. *American Journal of Epidemiology*, *174*(10), 1190–1196. https://doi.org/10.1093/aje/kwr246

Salganik, M. J., Mello, M., Abdo, A., Bertoni, N., Fazito, D., & Bastos, F. (2012). The Game of Contacts: Estimating the Social Visibility of Groups, *100*(2), 130–134. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations

Salganik, M. J., Mello, M. B., Abdo, A. H., & Bastos, F. I. (2011). Web Appendix Assessing network scale-up estimates for groups most at risk for HIV / AIDS : Evidence from a multiple method study of heavy drug users in Curitiba , Brazil, 1–17.

Sen, A. (1973). *On Economic Inequality*. Oxford: Clarendon Press. Retrieved from http://www.amazon.com/Economic-Inequality-Radcliffe-Lectures/dp/019829297X/ref=sr_1_1?s=books&ie=UTF8&qid=1310677285&sr=1-1

Sijtsma, K. (2016). Playing with Data???Or How to Discourage Questionable Research Practices and Stimulate Researchers to Do Things Right. *Psychometrika*, *81*(1), 1–15. https://doi.org/10.1007/s11336-015-9446-0

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

https://doi.org/10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2013). P-curve : A Key To The File Drawer. *Journal of Experimental Psychology: General*, *xx*(xx), 1–38. https://doi.org/10.1037/a0033242

Smart, L., & Wegner, D. M. (1999). Covering up what can't be seen - Concealable stigma and mental control.

Starosta, A. J., & Earleywine, M. (2014). Assessing base rates of sexual behavior using the unmatched count technique. *Health Psychology and Behavioral Medicine*, *2*(1), 198–210. https://doi.org/10.1080/21642850.2014.886957

Steneck, N. H. (2006). Fostering Integrity in Research : Definitions , Current Knowledge , and Future Directions, *12*(1), 53–74.

Stuber, J., Galea, S., & Link, B. G. (2008). Smoking and the emergence of a stigmatized social status. *Social Science and Medicine*, *67*(3), 420–430. https://doi.org/10.1016/j.socscimed.2008.03.010

Subramanyam, R. (2013). Art of reading a journal article: Methodically and effectively. *J Oral Maxillofac Pathol*, *17*(1), 65–70. https://doi.org/10.4103/0973-029X.110733

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Teixeira da Silva, J. A. (2018). Freedom of Speech and Public Shaming by the Science Watchdogs. *Journal of Advocacy, Research, and Education, 5*(1).

Tetlock, P. E. (1985). Accountability: A Social Check on the Fundamental Attribution

Error. *Social Psychology Quarterly*, *48*(3), 227–236.

Uleman, J. S., & Kressel, L. M. (2013). A Brief History of Th eory and Research on Impression Formation, 53–73.

United States Census Bureau. (2018). *National health Interview Survey: CAPI Manual for NHIS Field Representative*.

Value all research products. (2013), 2013.

Van Noorden, R. (2014). Scientists may be reaching peak in reading habits. Retrieved from https://www.nature.com/news/scientists-may-be-reaching-a-peak-in-reading-habits-1.14658

Veldkamp, C. L. S., Hartgerink, C. H. J., van Assen, M. A. L. M., & Wicherts, J. M. (2017). Who Believes in the Storybook Image of the Scientist? *Accountability in Research*, *24*(3), 127–151. https://doi.org/10.1080/08989621.2016.1268922

Wagenmakers, E.-J., & Dutilh, G. (2016). Seven Selfish Reasons for Preregistration. Retrieved from https://www.psychologicalscience.org/observer/seven-selfish-reasons-for-preregistration/comment-page-1

Wagenmakers, E., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (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. https://doi.org/10.1037/a0022790

Wang, H., Lu, Y., & Zhai, C. (2010). Latent Aspect Rating Analysis on Review Text Data : A Rating Regression Approach, 783–792.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, *7*(NOV), 1–12. https://doi.org/10.3389/fpsyg.2016.01832

Wilson, L. (1942). *The Academic Man: A Study in the Sociology of a Profession*. New York, NY: Oxford University Press UK.

Wolter, F., & Laier, B. (2014). The Effectiveness of the Item Count Technique in Eliciting Valid Answers to Sensitive Questions. An Evaluation in the Context of Self-Reported Delinquency. *Survey Research Methods*, *8*(3), 153–168.

Zafrunnisha, N., & Pullareddy, V. (2009). Authorship pattern and degree of collaboration in psychology. *Annals of Library and Information Studies*, *56*(December), 255–261.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science*, *22*(7), 894–900. https://doi.org/10.1177/0956797611411057

Zeileis, A., & Kleiber, C. (2015). R Package "ineq." *CRAN Package Library*, 15.

Zhang, T. Y., Hellstrom, I. C., Bagot, R. C., Wen, X., Diorio, J., & Meaney, M. J. (2010). Maternal care and DNA methylation of a glutamic acid decarboxylase 1 promoter in rat hippocampus. *J Neurosci*, *30*(39), 13130–13137. https://doi.org/10.1523/JNEUROSCI.1039-10.2010

Zheng, T., Salganik, M. J., & Gelman, A. (2006). How Many People Do You Know in Prison? *Journal of the American Statistical Association*, *101*(474), 409–423.

https://doi.org/10.1198/016214505000001168

Appendices

Appendix 1: End User Agreement for AbstractFindr.

The following outlines the agreement between the user ("you") and the developers ("the developers") of abstractfindr ("the product"). The following will outline definitions, terms of use, ownership and licensing, third-party distribution, and publishing, to the extent applicable for this product. If you are uncertain about your rights to use, license, publish, or distribute any material, you should contact your legal advisor. Use of the product is at your own discretion, and you may use or stop using the product whenever you'd like within one visit. Once you have used the product, you may not use it again.

A. abstractfinder is provided to you to use free of charge. The developers may make available future upgrades and advancements to the product for your future use. The developers may provide you any such upgrade or advancement for free. For example, if you originally use the product in 2018, the developers may update or advance the product in 2019. Future use of the product in 2019, with the updates and/or advancements, would also be free to use. Upgrades and updates, if any, may not necessarily include all existing features or new features that the developers release on other platforms. If they do, they may be provided without charge. These terms will govern any upgrades, updates, or advancements provided by the developers that replace and/or supplement the original product, unless such upgrade or update is accompanied by a separate end user agreement in which case the terms of that agreement will govern.

B. Title, license, and intellectually property rights to any and all content displayed by or accessed through the product belongs solely to the respective content owner. Such content may be protected by copyright or other intellectual property laws and treaties, and may be subject to terms of use of the third party providing such content. This agreement does not grant you any rights to use such content nor does it guarantee that such content will continue to be available to you.

C. Eligible users of the product are individuals that identify themselves as being employed in the scientific field of psychology, either current or retired. This means full professors, associate professors, assistant professors, emeritus professors, adjunct professors, post- doctoral researchers, research assistants, graduate students, lab managers, lecturers, senior lecturers, or any other such position where the individual identifies as conducting scientific research in the field of psychology. Eligible users may or may not conduct work at an educational institution, but should identify as being associated with a particular educational institution (i.e. college or university).

D. The developers of the product are Nicholas Fox, a graduate student at Rutgers University, Lee Jussim, a professor of psychology at Rutgers University, and Steven Mattia, an undergraduate student at Rutgers University. They may be contacted at the following email addresses: abstractfindr@gmail.com, nwf7@psych.rutgers.edu, jussim@psych.rutgers.edu. Nicholas Fox can also be contacted at the following phone number: 631-682-1343.

E. Any material downloaded from the product may not be transferred to other parties unless you obtain consent from the content owner. All material being provided through the product is done so with full knowledge of the ownership of the material belonging to the material creator, and not to the developers of the product.

F. You agree that the developers may collect, maintain, process, and use usage and related information that is gathered periodically to facilitate the provision of product updates, product support, and other services to you (if any) related to the product, ad to verify compliance with the terms of this agreement. All information collected is anonymous, meaning there are no personal identifiers collected. The developers may use this anonymous information to provide and improve future products and services. No user information will be transferred to third parties: the only parties able to access your anonymous use data will be the developers and the Rutgers University institutional review board, if necessary. This anonymous data will be saved for at least five years, as long as the information is relevant to the product. If the product changes in the future in such a way so that collected information is no longer relevant, that information will be deleted. At all times your information will be treated in accordance with the Rutgers University institutional review board policies on anonymous data collection, which is incorporated into this agreement.

G. Use of the product requires internet access. Certain services may require a mouse for interaction with the product. Any information displayed by any feature of the product is for general informational purposes only, and should not be relied upon for any particular advice.

H. You agree that the product contains proprietary content, information and material that is owned by the developers as well as the content creators, and is protected by applicable intellectual property and other laws, including but not limited to copyright and creative commons, and that you will not use such proprietary content, information, or materials in any way whatsoever except for permitted use of the product or in any manner that is inconsistent with the terms of this agreement or that infringes any intellectual property rights of a third party or the developers. No portion of the product may be reproduced in any form or by any means. You agree not to modify, rent, lease, loan, sell, distribute, or create derivative works based on the product, in any manner, and you shall not exploit the product in any unauthorized way whatsoever, including but not limited to, using the product to burden network capacity or artificially inflate usage data. You further agree not to use the product in any manner to harass, abuse, threaten, defame or otherwise infringe or violate the rights of any other party, and that the developer is not in any way responsible for any such use by you, nor for any harassing, threatening, defamatory, offensive, infringing or illegal messages or transmissions that you may receive as a result of using the product.

I. In addition, the product and third party materials that may be accessed through the product are not available in all languages. The developers make no representation that the product and such third party materials are appropriate or available for use in any particular location. To the extent you choose to access the product or such material, you do so at your own initiative and are responsible for compliance with any applicable laws, including but not limited to local laws. The developers reserve the right to change, suspend, remove, or disable access to the product at any time without notice. In no event will the developers be liable for the removal of or disabling of access to the product. The developers may also impose limits on the use of or access to the product, in any case without notice or liability.

J. This agreement is effective until terminated. Your rights under this agreement will terminate automatically or otherwise cease to be effective without notice from the developers if you fail to comply with any term(s) of this agreement. Upon termination of this agreement, you shall cease to use the product.

K. You expressly acknowledge and agree that, to the extent permitted by applicable law, use of the product is at your sole risk and that the entire risk as to satisfactory quality, performance, accuracy and effort is with you.

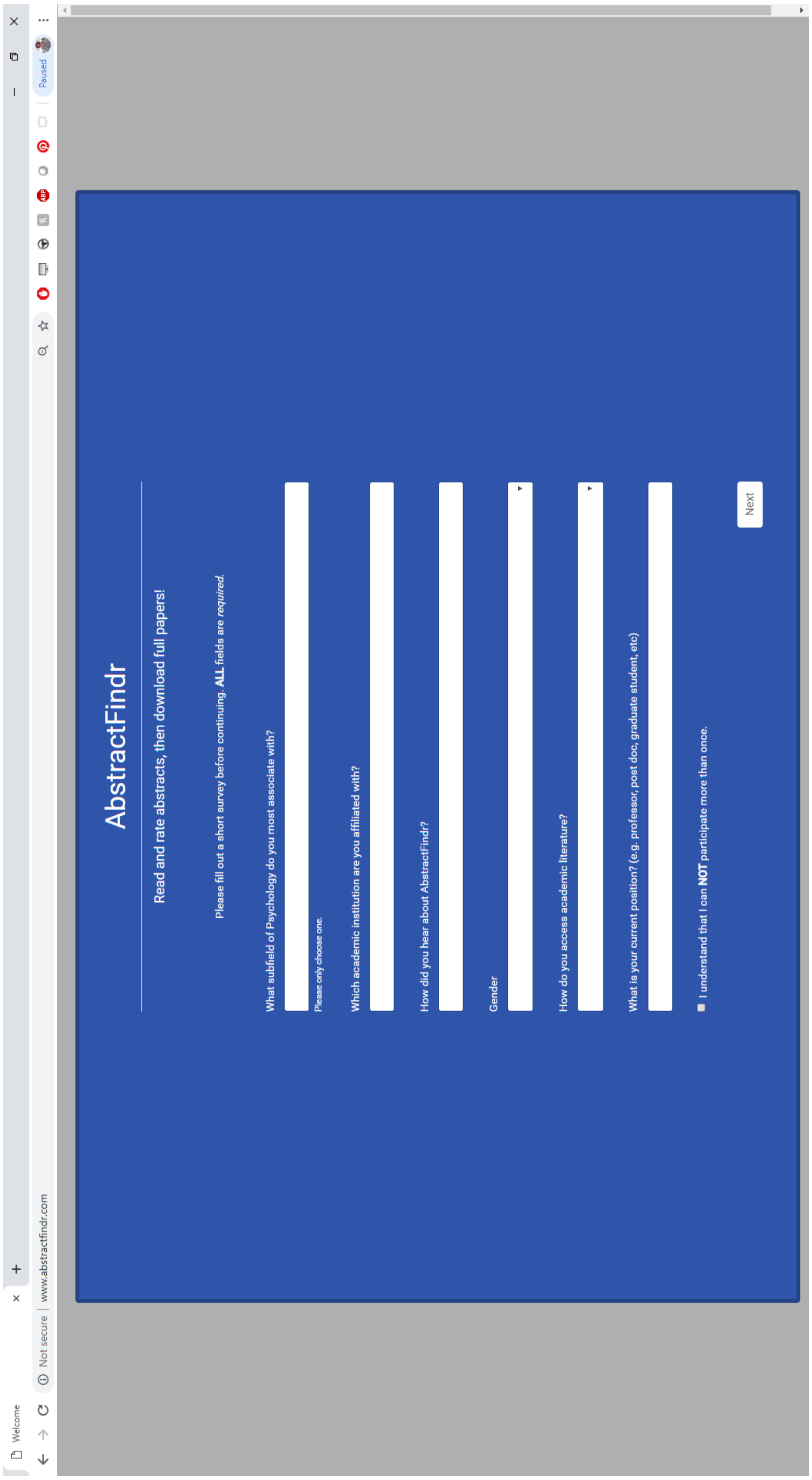
L. To the maximum extent permitted by applicable law, the product is provided “as is” and “as available”, with all faults and without warranty of any kind. The developers do not warrant against interference with your use of the product, that the operation of the product will be uninterrupted or error-free, that any services will continue to be made available, or that the product will be comparable to any other third party product.

M. No oral or written information or advice given by a developer or a person speaking on behalf of the developers shall create a warranty. Should the product prove defective, the developers will not be held liable for any repairs to the product. Some jurisdictions do not allow the exclusion of implied warranties or limitations on applicable statutory rights of users, so the above exclusion and limitations may not apply to you.

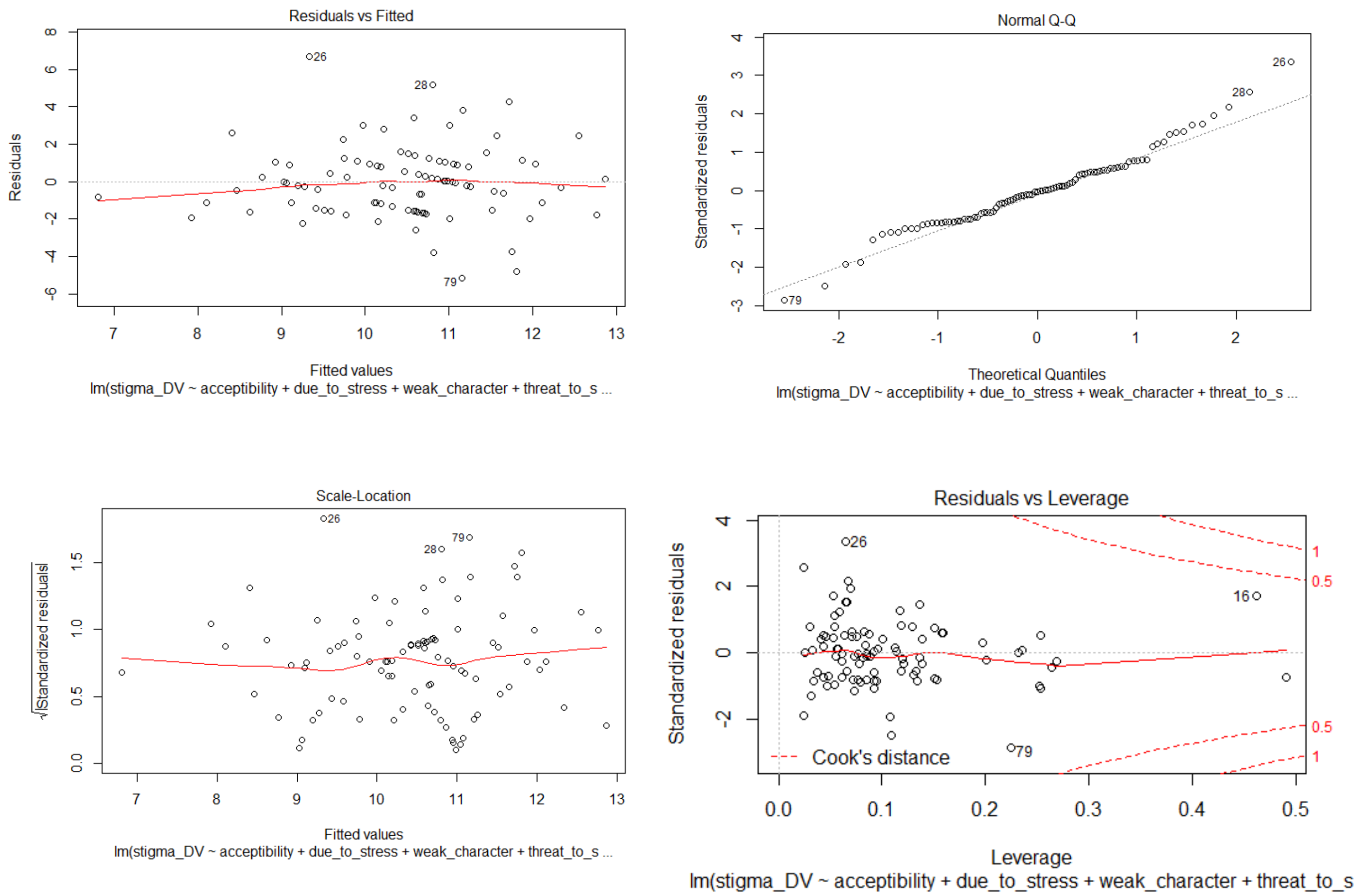
N. There are no foreseeable risks to using the product. However, to the extent not prohibited by applicable law, in no event shall the developers be liable for personal injury, or any incidental, special, indirect or consequential damages whatsoever, including, without limitation, damages for loss of profits, loss of data or information, loss of intellectual property, business interruption or any other commercial damages or losses, arising out of or related to your use or inability to use the product, however caused, regardless of the theory of liability and even if developers have been advised of the possibility of such damages. Some jurisdictions do not allow the limitation of liability for personal injury, or of incidental or consequential damages, so this limitation may not apply to you. In no event shall the developer's total liability to you for all damages (other than as may be required by applicable law in cases involving personal injury) exceed the amount of fifty dollars. The foregoing limitations will apply even if the above stated remedy fails of its essential purpose.

O. This document constitutes the entire agreement between you and the developers relating to the use of the product and supersedes all prior or contemporaneous understandings regarding such subject matter. No amendment to or modification of this agreement will be binding unless in writing and signed by the developers. Use of the product is entirely voluntary, and you may choose with stop using the product at any time.

Appendix 2: Screenshots of each front page of AbstractFindr.

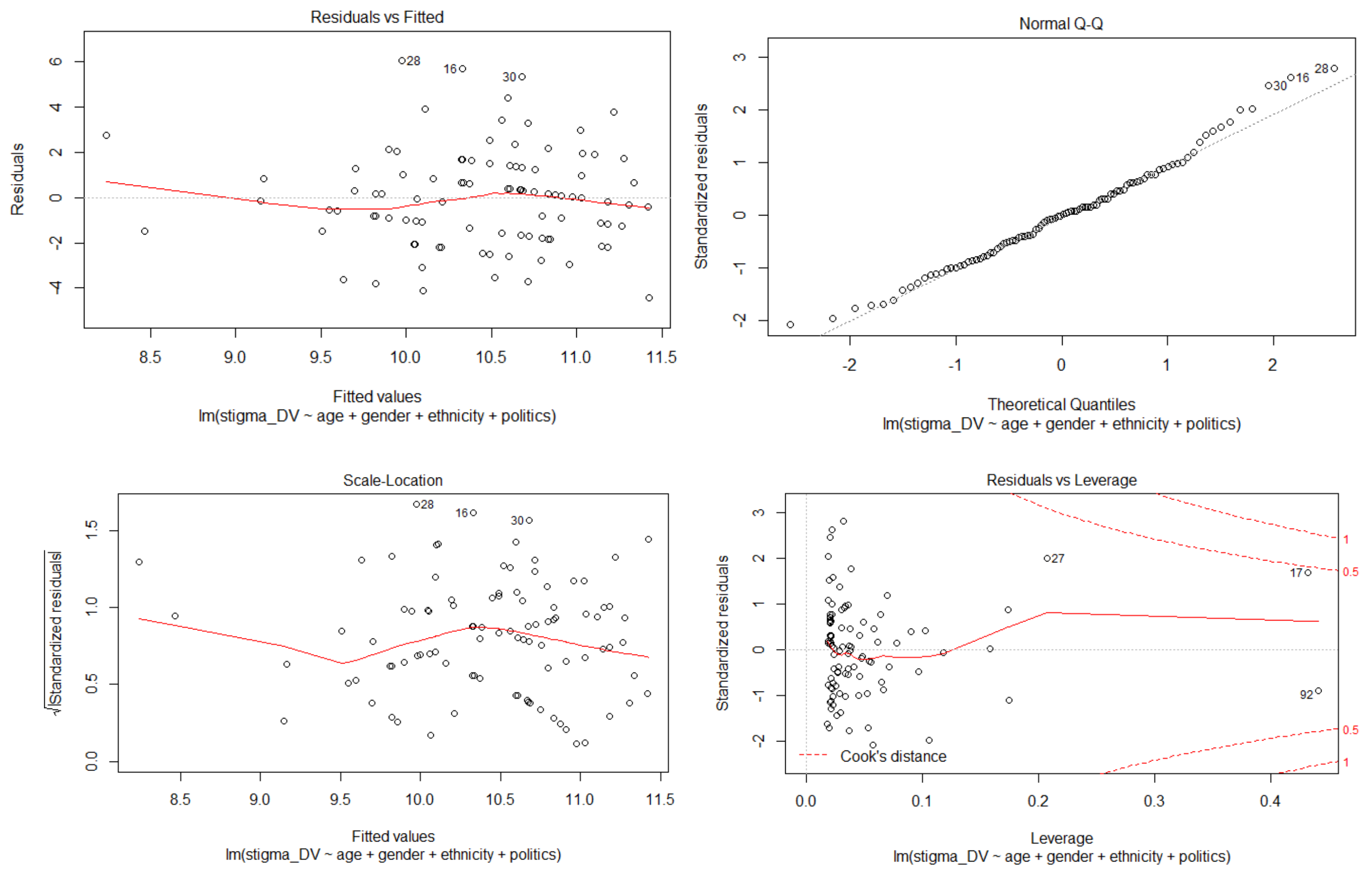

Welcome
Not secure | www.abstractfindr.com
Paused
AbstractFindr
Read and rate abstracts, then download full papers!
Please fill out a short survey before continuing. ALL fields are required.
What subfield of Psychology do you most associate with?
Please only choose one.
Which academic institution are you affiliated with?
How did you hear about AbstractFindr?
Gender
How do you access academic literature?
What is your current position? (e.g. professor, post doc, graduate student, etc)
I understand that I can NOT participate more than once.
Next

Appendix 3: Diagnostic plots for "all variable" regression model (Table 10).



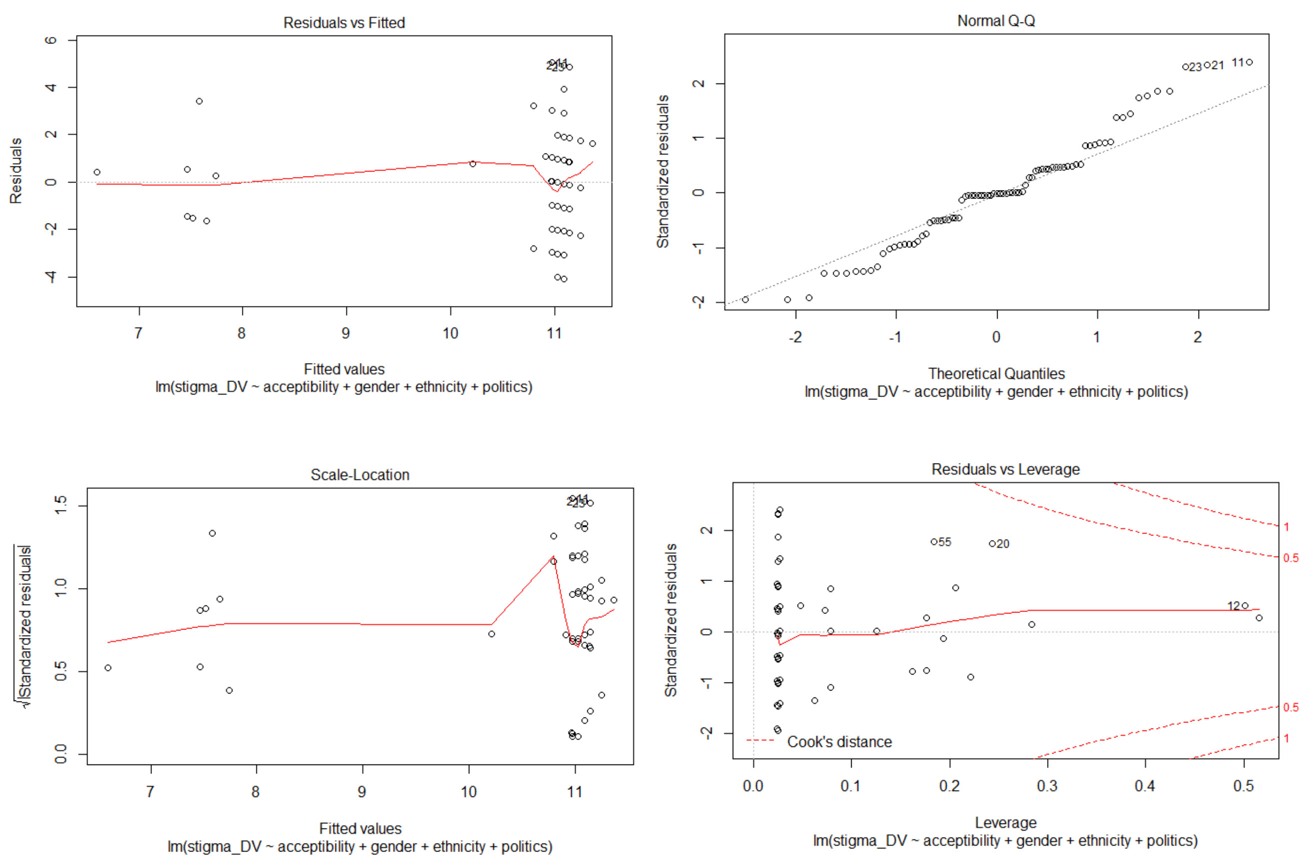
The "Residual vs Fitted" plot tests the assumption of linearity. The "Normal Q-Q" plot tests the assumption of normally distributed residuals. The "Scale Location" pot tests the assumption of homogeneity of variances. The "Residuals vs Leverage" plot is used to identify overly influential datapoints.

Appendix 4: Diagnostic plots for "age" regression model (Table 12).



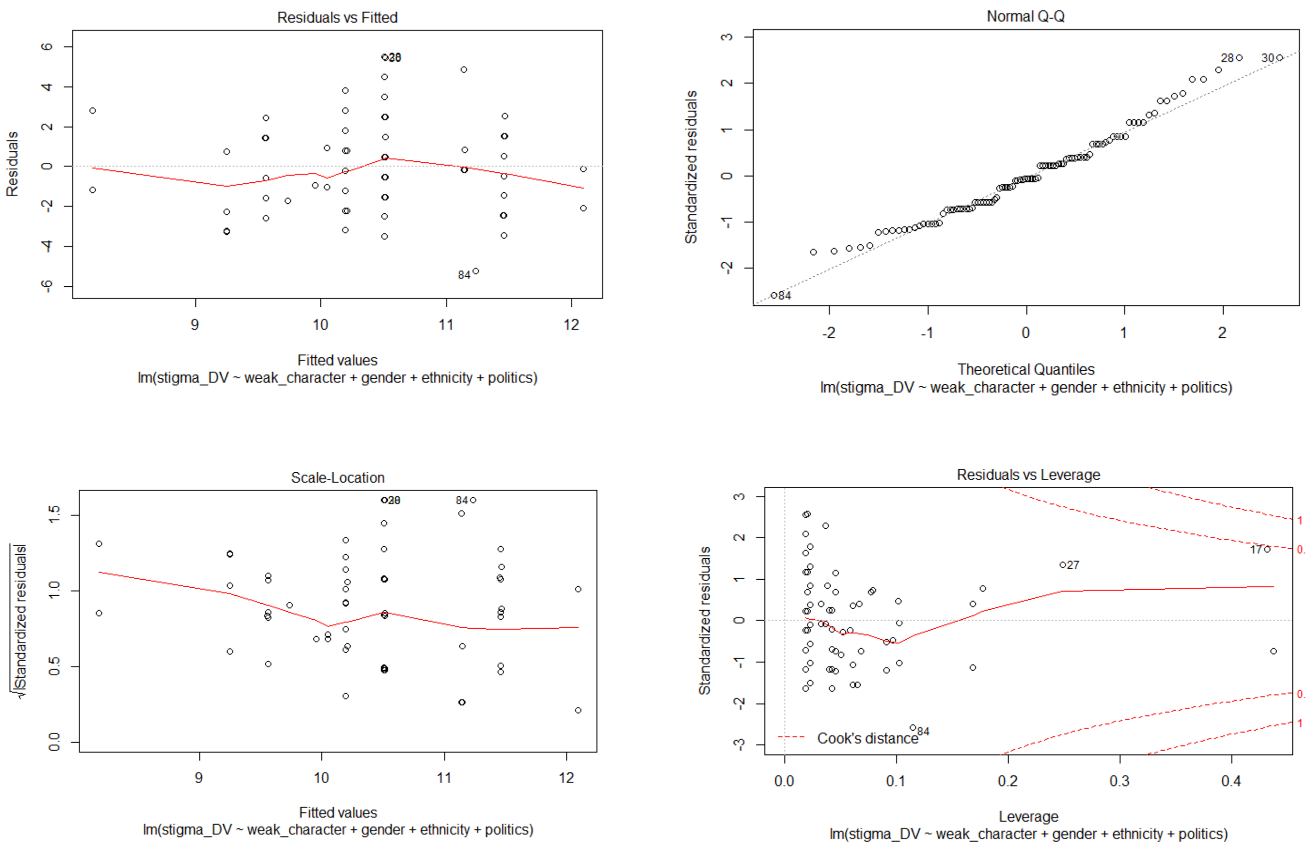
The "Residual vs Fitted" plot tests the assumption of linearity. The "Normal Q-Q" plot tests the assumption of normally distributed residuals. The "Scale Location" pot tests the assumption of homogeneity of variances. The "Residuals vs Leverage" plot is used to identify overly influential datapoints.

Appendix 5: Diagnostic plots for "acceptability" regression model (Table 13).



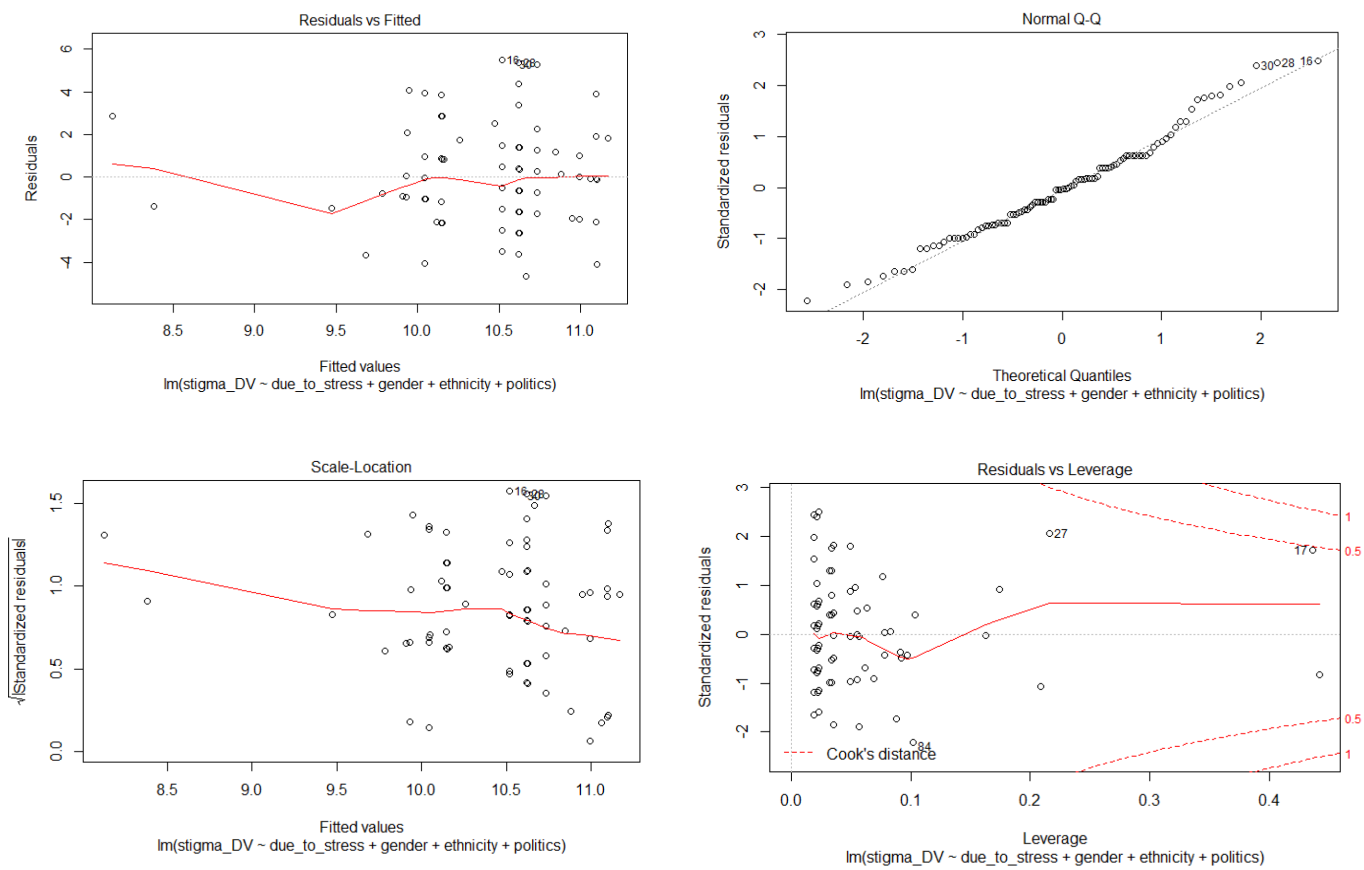
The "Residual vs Fitted" plot tests the assumption of linearity. The "Normal Q-Q" plot tests the assumption of normally distributed residuals. The "Scale Location" pot tests the assumption of homogeneity of variances. The "Residuals vs Leverage" plot is used to identify overly influential datapoints.

Appendix 6: Diagnostic plots for "internal attribution" regression model (Table 14).



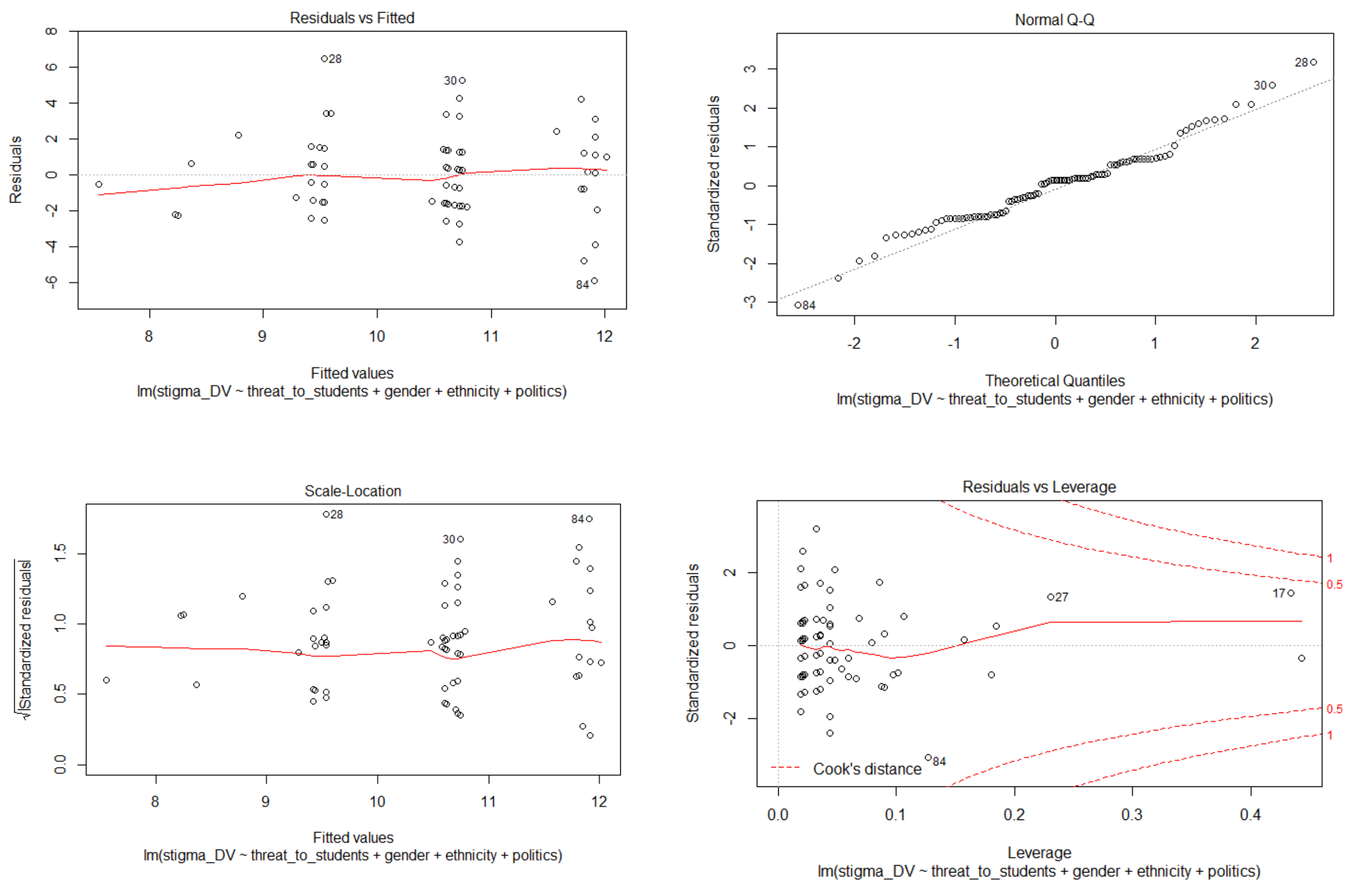
The "Residual vs Fitted" plot tests the assumption of linearity. The "Normal Q-Q" plot tests the assumption of normally distributed residuals. The "Scale Location" pot tests the assumption of homogeneity of variances. The "Residuals vs Leverage" plot is used to identify overly influential datapoints.

Appendix 7: Diagnostic plots for "external attribution" regression model (Table 15).



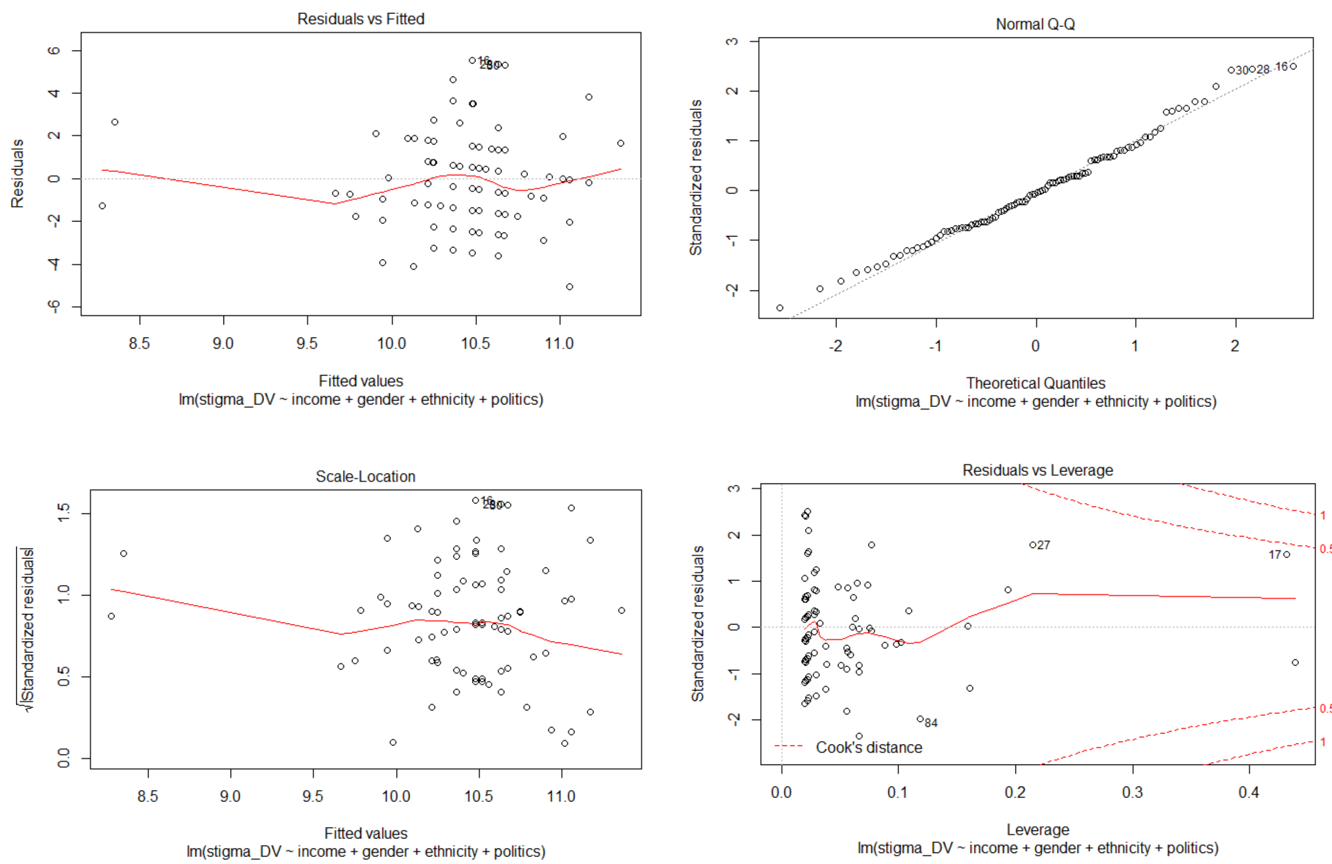
The "Residual vs Fitted" plot tests the assumption of linearity. The "Normal Q-Q" plot tests the assumption of normally distributed residuals. The "Scale Location" pot tests the assumption of homogeneity of variances. The "Residuals vs Leverage" plot is used to identify overly influential datapoints.

Appendix 8: Diagnostic plots for "fear" regression model (Table 16).



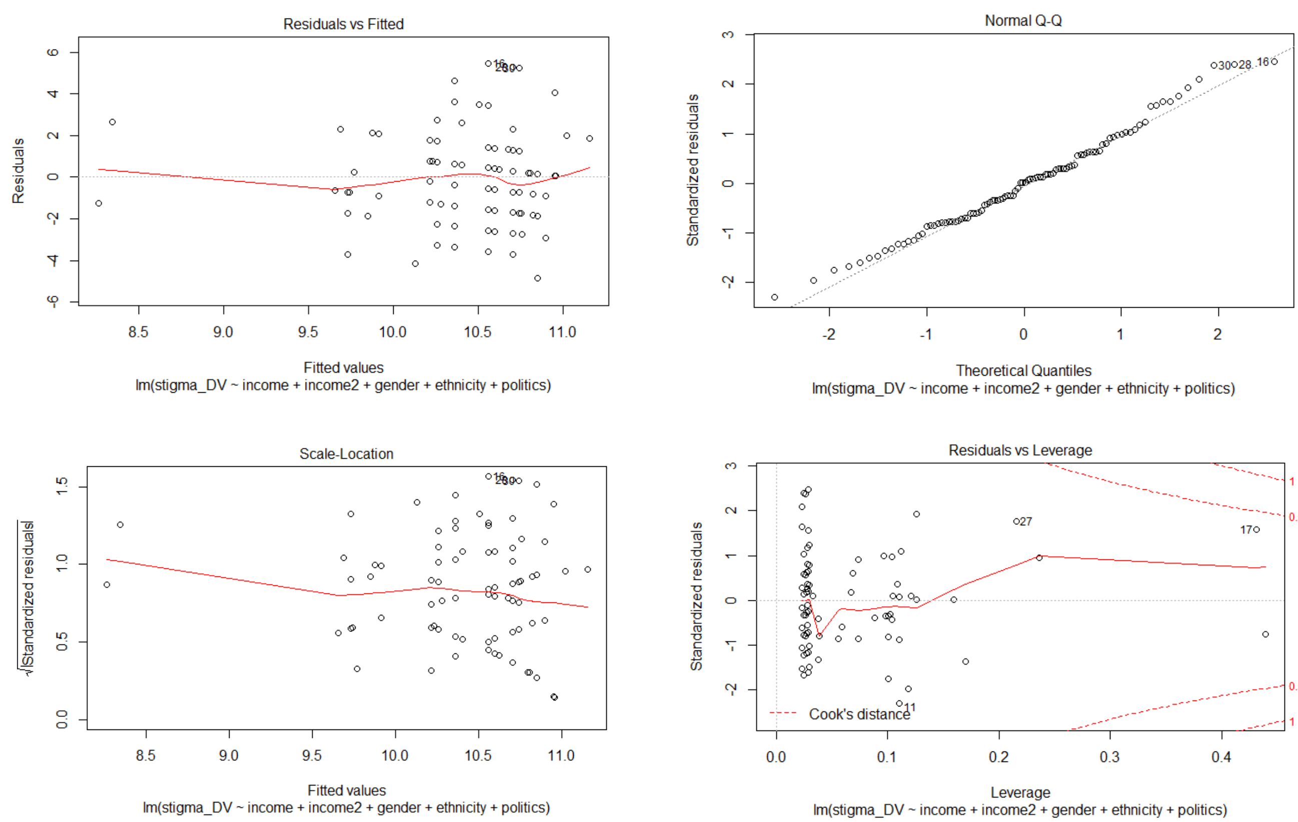
The "Residual vs Fitted" plot tests the assumption of linearity. The "Normal Q-Q" plot tests the assumption of normally distributed residuals. The "Scale Location" pot tests the assumption of homogeneity of variances. The "Residuals vs Leverage" plot is used to identify overly influential datapoints.

Appendix 9: Diagnostic plots for "power" regression model (Table 17).



The "Residual vs Fitted" plot tests the assumption of linearity. The "Normal Q-Q" plot tests the assumption of normally distributed residuals. The "Scale Location" pot tests the assumption of homogeneity of variances. The "Residuals vs Leverage" plot is used to identify overly influential datapoints.

Appendix 10: Diagnostic plots for "power squared" regression model (Table 18).



The "Residual vs Fitted" plot tests the assumption of linearity. The "Normal Q-Q" plot tests the assumption of normally distributed residuals. The "Scale Location" pot tests the assumption of homogeneity of variances. The "Residuals vs Leverage" plot is used to identify overly influential datapoints.