Best Practices for Replicable Research on Gender Differences

Shelby Kennedy

Advised by Mark Sheskin, Yale University

Submitted to the faculty of Cognitive Science in partial fulfillment of the requirements for the degree of Bachelors of Arts

Yale University
April 22, 2019

Abstract

Gender has long been at the forefront of psychological research. As such, many opposing viewpoints of the role of gender have developed overtime, ranging from claims that there are no gender differences other than those completely caused by problematic socialization, to the alternative that the two genders are innately different. While these are merely two extremes of the views on how gender differences form, most beliefs and understandings fall in the realm of being a combination of the two. However, it is important to note that understanding where gender differences are rooted is an incredibly complex discussion that has yet to be explicitly answered, despite thousands of years attempting to do so. The ability to continue such a debate is therefore reliant upon having access to high-quality empirical evidence. Unfortunately though, modern psychology is jeopardizing said access but reducing the reliability of much of the available evidence. In this paper I describe the role that gender has played within psychology – moving through major debates and current research. I then describe the current limitations of psychological research and how the specifically apply to gender. Ultimately, I explore potential solutions to mitigate the current limitations. Furthermore, I suggest that gender should always be recorded within psychological research but should only be reported if it was a pre-registered variable. This solution specifically reduces the possibility of finding false-positives within gendered research and creates a strong backbone for current and future research about gender.

Contents

1. Introduction

Scientific research is dependent upon replication as a form of checks and balances to maintain credibility and validity to the understanding of the world – statistically significant data is expected to be able to be replicated in order for it to be considered relevant within the sphere of scientific findings. But what happens when this quality control is no longer fully functional? Unfortunately, modern psychology is currently exploring this.

Papers published within academic journals are often guaranteed as being peer-reviewed, and as such, are assumed "trustworthy and reliable" (Cooper, Donovan, Waterhouse & Williamson, 2007). However, with a decrease in traditional replication practices, this statement no longer is always true. The publication of misrepresentative statistics is becoming a threat to the understanding of the psychological world. The ability to replicate published studies was put to the test when 100 already published experimental and correlational studies were attempted to be replicated. The results did not speak to the aforementioned confidence in "peer-reviewed" publications. Only 36% of the replications showed statistically significant results, to be compared to the original 97% (Open Science Collaboration, 2015). These findings quite clearly point to the reality of a systematic flaw in current research.

Today's society is a world which people are constantly surrounded by new novel ideas. Innovation is not only praised but is required in order to keep pace with the ever-evolving and ever-accessible world. Psychological research is no exception to seemingly needing to fit into this mold. Exclusive sells, even in regards to research (Bohannon, 2015) and this pressure to maintain not only novelty but an impressive pace of publication leaves researchers and reviewers alike seemingly desperate to find their results in publications. With null or merely replication results showing no promise of publication, there is little incentive to then practice these

behaviors. Though less glamorous, replication in turn breeds innovation: “innovation points out paths that are possible, replication points out paths that are likely; progress relies on both” (Open Science Collaboration, 2015). This notion has become seemingly lost and has instead resulted in a hyper-intention to be innovative and a reduced intention to replicate – a dangerous combination. Researchers are looking to “efficiently” manipulate their data to pump out results, while reviewers are hoping to build upon these findings, clearly putting a kink in the checks and balances system.

As previously mentioned, the world is currently a space in which information is constantly at our finger-tips. With access to such a flurry of information, people want to consume information that is easy to digest: they want to have a black and white understanding of the world. A constant “binary” that lends itself to being a possibility of such a way of thinking is that of gender. Everywhere people go they are separated by their gender: boys and girls restrooms, clothing, color schemes. There are huge parties thrown in celebration of distinguishing an unborn child’s classification to one of these two groups. Since this division is so prevalent within people’s social lives, it seems only likely that this concrete categorization carries over into the psychological world – or at least that is what people would like to believe.

Gender is an easy concept to observe, understand, and consequently research since it is always accessible. As such, gender differences often finds itself the content of psychological studies – rightfully so, as understanding the psychological differences has long riddled scientific exploration. However, in the state that modern psychology is currently in, gender serves as the perfect variable to manipulate to produce publishable, and palatable results. There is a clear need for systemic reform in order to salvage the credibility of gender psychology to preserve the ability to understand social roles and individual development. By working through the ins and

outs of the replication crisis and how it relates to gender psychology, a hopeful solution to the current tensions can be concluded. Ultimately, this paper is looking to increase the ability to produce quality findings within gender research so that the aforementioned long debates may continue to be fueled and grounded in appropriate data.

2. Psychology of Gender

The psychology of gender has been a hot topic within psychological research dating far beyond modern studies. As knowledge of gender developments have progressed, a looming question still remains: where do gender differences come from? There is a large disparity between the beliefs of what answers this question. Some extreme cases noting that men and women are innately different beings while others suggest that the two genders are born equal and only brought into gender roles through socialization. Of course, there are many less radical views that fall within this continuum and with an ultimate, underlying, understanding that there is some merger of these two factors that results in gender differences. Even still, solving this debate is seemingly pressing within psychological research as gender remains at the forefront of psychological exploration. Accordingly, in this section I will explore the role that gender has played within psychological research and the core beliefs of what gender differences are rooted in, all the while contextualizing the emphasis that is often times put on understanding gender differences in modern research.

2.1 History of Psychology of Gender

Exploring the cognitive differences between genders is a key area of exploration within psychological research, and has been ranging far back into the history of psychological exploration. Plato accounts for the exploration of the role of women in Ancient Greece, arguing that “their original natures are the same” (Lewontin & Nelson, 1984). From this he concludes

that both men and women are able to pursue the same positions and education, but should "confine themselves to primary social roles for which they are best fitted by temperament and education" (Taylor, 2012; Lewontin & Nelson, 1984). Even still, Plato is deriving these statements all while claiming that despite having the equal freedom to pursue different roles, women continue to be inferior to men (Lewontin & Nelson, 1984). This shows the early contemplations surrounding how gender influenced societal roles and vice versa. Later, early Freudian theories, such as that of the Oedipus and Electra complex, also attempted to explain gender differences through the development of their sexuality. The Oedipus and Electra complex refer to the development of sexual impulses directed at one parent but resentment towards the other. Embedded in these theories are the ideas of castration anxiety and penis envy. Freud uses these theories as justification for each sex taking on their gender roles; boys identify with their father's equal fear of castration, while girls are pushed away from masculine behaviors since they envy their having a penis (Yadav, 2018). Despite the many differences between Plato and Freud, the shared topic highlights the focus there has been on gender as a consistent variable within the exploration of development of roles in society – a trend that has not faltered even in today's research.

Despite gender being of interest dating far back into psychological research, the explicit understanding of where gender differences come from and to what extent they develop has yet to be conclusively settled. As such, the desire to establish concrete distinctions between men and women has deep roots. With this in mind, partnered with the large disparity between what people believe to account for gender differences gender remains prevalent within modern research.

2.2 Vocabulary

Before moving forward in the discussion of the psychology of gender research, it is important to draw a clear distinction in relevant language use. 'Gender' is associated with the behavior and social identifications of an individual, as opposed to 'sex', which refers to the biological characteristics (Pryzgoda & Chrisler, 2000). Despite many people believing these terms to be synonymous, this distinction will be acknowledged throughout the remainder of this paper. As such, in order to keep the distinction between these two separate the terms 'male and female' will be used to refer to sex, and 'men and women' will be used to refer to gender.

2.3 Desiring Discrete Gender Differences

Many physiological differences between males and females can be observed and understood almost immediately. This allows for clear and distinct categorization between the two sexes, at least in most cases. As such, many people hope to claim a parallel distinct psychological categorization.

Among these clear physiological distinctions, height, for example, clearly follows patterns within gender. On average, adult American males are 5'9'' compared to the near 5'4'' American adult females (Fryar, Gu, Ogden & Flegal, 2016). This allows for greater height to be associated with males, allowing for a very black and white grouping of attributes per group. Importantly, however, not all men are taller than all women. Approximately 2% of women are taller than the average man, and around 4% of men are shorter than the average women (varying slightly by age) (U.S. Census Bureau, 2011). Even in this example, no single feature can be considered a complete binary since there are exceptions to even the most seemingly finite distinctions.

People so strongly associate these qualities with the allotted gender, though, that there are stereotype repercussions that result in cases where these are violated. For men especially, their

violation of being the taller gender impacts the perception of their social attractiveness, professional status, personal adjustment, athletic orientation, masculinity, and physical attractiveness – being short is "more of a liability than being tall is an asset" (Jackson & Ervin, 1992). Women are also perceived differently, though less drastically and not in as many categories, via their height. Believing that gender acts as a discrete stratification allows for justification of stereotypes on the basis of gender – one of the motivating factors in the categorization of gender psychology. Even so, it is never interpreted that just because these stereotypes are violated, such in the case that a female is taller than the average male, that they are not fully in their assigned group; being exceptionally tall for a female does not classify someone as a male. Social classifications account for these outliers distinctions, shifting them from being a strict binary. Instead it is understood that there are many factors outside of an individual's sex that impacts their height. Despite the clear categorization of this feature, there is a universal understanding that a balance of genetics and nutrition among other things makes way for the actual height of an individual, not the gender alone. For this reason, while there are consequences for violating expectations, these outliers are not ostracized from the group as a whole.

In a similar manner, this discrete categorization is sought after in regards to psychological characteristics as well. Since there are so many seemingly inherent distinctions between men and women in varying categories, it is seemingly logical that the gender of the individual should also influence them psychologically. While gender is often viewed as a taxon, meaning the members "have a greater likelihood of possessing traits that are characteristic of that taxon than nonmembers do" gender overall is a much less rigid metric of differentiation than this "restricted range" where they are "uncorrelated within each group" (Reis & Carothers, 2014). Even so, there

is still a strong perception that there are core differences in genders, which are rooted solely on their biological or sociocultural factors. Though this is not necessarily the case, the drive to seek the validity of the source of these differences motivates a continuation of studying gendered psychology.

2.4 Nature v Nurture Debate

The debate over whether psychological abilities, gendered or otherwise, are rooted in biology or sociocultural influences has been a large source of contention throughout psychology. Although it has been further understood that the most accurate explanation relies on both the two explanations (Eagly & Wood, 2013), supporting one theory over the other has motivated a great deal of psychological research. This is especially the case in regards to understanding gender differences.

On the 'nature' side of the debate it is argued that psychological gender differences are strictly a result of their biological processes. This reasoning is used to "'further strengthen the view that women and men are different human kinds'" (Prentice & Miller, 2007; Reis Carothers, 2014), since it is suggesting that their differences are categorical and innate. The 'nurture' side however, adds weight to the sociocultural influences rather than the biological or cognitive. This follows John Locke's Tabula rasa, considering children's minds to be a blank slate in which their interactions influence their experiences and development (Uzgalis, 2019).

An example of this is marked in regards to the temperaments of young children. Young boys are measured as having surgency in motor control, impulsivity, and high-intensity compared to the strong emergence of control in attention span, shifting of their attention, and perceptual sensitivity in young girls – differences that were large (control) and moderate (surgency) respectively between the two genders (Eagly & Woods, 2013). This speaks to the

nature theory since these characteristics are present within young children, alluding to the presence of these traits being a result of the hardwiring of the children's brains. Alternatively, the nurture view can observe and add power to the interactions between children and their parents. In this way, the encouragement of gender-typical behavior, such as using more supportive speech with daughters, allowing physical risk taking in sons, and creating gendered roles in chores as having influence over how these children develop (Eagly & Wood, 2013) can be explored.

Keeping this example in mind, the intersection of these two influences can more holistically explain these interactions and developments. The merging of nature and nurture suggests that interactions are rooted in the children's response to the type of interaction. This means that boys might receive more physical social interactions since they respond positively to these types of interactions (Eagly & Wood, 2013). This combination of the two theories clearly justifies the behavior as well as the perpetuation of the behavior by showing how they interact; the biological differences reinforce the cultural roles and interactions (Eagly & Wood, 1999).

Despite the prevalence of the extreme beliefs that it is either one end of this spectrum, either nature or nurture that facilitates gender disparities is misguided. Instead, it should be ultimately understood that both are always accredited; one cannot exist without the other. While this debate has and continues to riddle motivations within psychology, it is an unwarranted debate since "every aspect of an organism's phenotype is the joint product of its genes and environment" (Cosmides & Tooby, 1997). Even still, it has motivated much of the research on gender.

2.5 Similarities Between Gender Psychologies

In contrast to the aforementioned height example, there is overlap between the genders, and while there are expectancies and patterns that align more strongly with one gender over the

other, they are not on a rigid binary. Instead, differences between genders can be more strongly linked to "individual differences that vary in magnitude" rather than being solely rooted in their gender (Reis & Carothers, 2014) and therefore occur on a dimensional scale. As such, explaining the overlap of intergroup distributions, as well as the level of intragroup variation becomes feasible; sometimes there can be greater variability within gender than between genders (Hyde, 2007).

Viewing gender in this way also allows for celebration and exploration of the similarities between genders rather than just their differences, a feature that Hyde identifies as being valuable and common in researching between genders in her Gender Similarities Hypothesis (2005).

In a meta-analysis of psychological gender difference data, Hyde explored the effect sizes of the data by categorizing them as being either close to zero, small, moderate, large or very large and found that 78% of the gender differences previously found were small or close to zero (Hyde, 2005). This both brings into question the validity of gender difference studies, as well as sheds light on the frequency of the similarities found between genders. As to be expected, there were still exceptions such as motor behaviors, sexuality, and aggression (Hyde, 2005) just as the value of context was still to be understood (Hyde, 2007). Yet overall, the gender similarities hypothesis shows the critique of the current fashion in which gendered psychology is explored.

3. Major Debates Over the Exploration and Motives of Gender Differences

The fact that each trait is due to both "nature" and "nurture" is merely the beginning of the interesting analysis, rather than the end of it. For example, analyses of "heritability" indicate what percent of variability in a phenotype is due to variability in genes (What is heritability?, 2019). An explicit example can be seen through the exploration of the

heritability of intelligence. There are clearly many environmental causes that can be gathered as influencing intelligence, such as education or parental support. Despite these outside factors though, many biological studies have found that variation in human intelligence is due to genetic variation; one particular study finding that 40-50% of variation of human intelligence being associated with a specific allele frequency (Davies et. al, 2011).This is a particularly interesting instance in which the trait is due to "nature" but can be exacerbated by "nurture". A further exploration of these concepts of trait development is in regards to canalization. Canalization is the tendency of a specific genotype to develop regardless of conditions, such as a different environment or genetic makeup (Hallgrimsson et al., 2002). Exploring how canalized a trait is, indicates the stability of certain traits.

Conceptualizing the development of different traits using biology allows us to approach gender differences with more precise questions: how much, if any, phenotypic variation between psychology of men and women is due to genetic variation between "XX" and "XY" genotypes? How canalized are phenotypic traits that have a strong genetic component?

Understanding the ideas stated here and within the preceding section, as previously noted, is only the beginning to understanding the large role that gender plays within both research and in understanding the world around us. As such, moving forward through this section, I will continue to explore the prevalence of gender in psychological research. However, I will shift away from the underlying motivations for studying these differences, and move instead into a looking at explicit examples of modern research and conversations surrounding gender.

3.1 Nature v Nurture Cont.: Public Debate Over the Origins of Psychological Differences

Before covering recent research on potential gender differences in psychology, it can be instructive to take a close look at a particularly salient example of a back-and-forth discussion on this topic. Although the below example is now over a decade old, the arguments described by the experts involved with the debate are still an important core of the debate today.

On January 16, 2005 the Harvard President, Lawrence Summers, sparked a light under the nature v. nurture debate for justifying psychological gender differences. Summers claimed that the differences between males' and females' cognition, math and science abilities specifically, were innate; coming from an accredited man that offered only four of thirty-two tenured positions to women (Bombardieri, 2005) this was not universally well received. In response to the ripple effect that the claims of the University's President had, Harvard became the facilitator of this debate on a larger scale. Ultimately, it became the breeding ground for the Steven Pinker and Elizabeth Spelke debate: biological vs. socio-cultural.

Within the debate, Steven Pinker sided with Lawrence's original claim that psychological gender differences are an innate result of biological mechanisms. Pinker establishes the baseline understanding that cognitive differences exist between men and women by providing six examples. Among these six categories, Pinker explored the separation of life priorities, the interest in people as opposed to things, the higher rates of risk aversion in females, the male advantage in three-dimensional mental movements, the different strengths within mathematical reasoning and testing, and the greater male variability (Science of Gender). Pinker uses these differences as his initial platform to divulge in discussion about instances in which differences can be understood as having a biological influence greater than zero. Again Pinker lists out six cases that support his belief in the inherent biological differences.

Among his main speaking points, Pinker points to the universality of major sex differences cross-culturally, specifically accounting for child care in women and competitiveness in men. Pinker goes one step farther to even suggest that sex differences are not only universal cross-culturally, but found within other mammals as well – again looking at childcare roles and aggression.

Working to explore and explain the disparity between genders specifically in math and science ability however, Pinker turns to the early emergence of gender stereotypes as an explanation of the biological cause; if a child is presenting differences in response cues at less than a week old there was seemingly not enough social influence to have instilled this in them. According to Pinker, many of these early cognitive separations are only further developed as children move along in development with there being an evident line between play-time choices: boys participate in "rough-and-tumble play" preferring vehicles and weapons, where as girls show cooperative play and look for play parenting and dolls. Pinker surmises that this is the result of the early shown preferences for faces over objects in girls and vice versa for boys – a manifestation of the innate predisposition of preference.

Pinker also works to debunk the negative socializing role of teachers, claiming the idea that teachers "perpetuate gender inequities by failing to call on girls in class" a myth. Instead, teachers are supposedly found to be creating perceptions of the students in reference to their actual accomplishments (Science of Gender; Jussim & Eccles,1992).

Opposing Pinker's strong preference for the biological justification, Elizabeth Spelke pushed for the value of socio-cultural factors in shaping the gender gaps. One central argument pushed by Spelke is that mathematic capability is not an evolved but rather a new acquisition. As such, there has to be some degree of societal influence. Even if there is a biological, cognitive

skill-set involved in performing the skill, since it is not innately based in evolution, mathematic abilities must have developed through social acquisition.

Surprisingly, Spelke draws on many of the same theoretical suggestions of gender gaps as Pinker. She too explores the timeline these disparities and the role of the education system in their development. First looking at the divergence of the math scores and pursuit of math within education, Spelke refers to a large-scale study exploring mathematic breakdowns in college in which it was found that there was an equal aptitude for mathematics between men and women. However, despite the two genders getting equal grades, there are still significantly less women majoring in mathematics (Science of Gender; Halpern (get year)). This points to an increasing role of social factors in discouraging females from pursuing a career in mathematics, since, as noted here, it is not a skill-based decision. Spelke also accounts for the disparity in SAT-M test scores by suggesting that it is not the mathematic ability that is being reflected, but rather the difference in preferred solution techniques. Males and females prefer, or have a tendency to, solve problems using different approaches. Some of these different approaches may either take more time or be more frequently included within particular versions of the SAT-M. Naturally as such, some scoring of the tests may be inflated to favor a particular solving method over the other, not necessarily overall mathematical skill-set. Again, the assumption that the difference in scores is an automatic representation of innate skill speaks to the demand to fulfill the stereotypic expectation that mathematics is a man's world.

Referring back to the timelines on which these become apparent, Spelke suggests a much different interpretation than did Pinker. Pinker suggested the emergence of gender differences beginning at merely a week old should be interpreted as indication that these differences are innate. Spelke specifically explores the development of mathematic abilities. Mathematic skill-

set can be broken down into a subset of five core systems. In exploration of the development of these systems, no differences were found between genders. This equal development of the abilities that predetermine mathematic ability then supports that the gap in the self-selection into mathematics. Since these differences emerge later in life, paralleling the logic behind Pinker's timeline justification, then there is room for, and there will likely be found, social influence.

In her portion of the debate, Spelke refers to a "snowball effect" alluding to a reinforcement of in-group preferences, noting that people "have an easier time imaging ourselves in careers where there are other people like us". While her example follows the suit of women in math and science, this concept can be generalized within the entirety of the nature v. nurture debate; even if there are biological differences, social factors would perpetuate them based on expectations and the desire to be a member of a group. Gendered in-group preferences can start as young as two years old, with girls and women showing particularly strong own gender preferences (Dunham, Baron & Banaji, 2015). Being vulnerable to gender variation so early on only leads to an expected deep loyalty to one's in-group and desire to fit into its mold. The strength of in-group psychology alone may be enough to validate continuation of gendered behaviors and the engrained belief of what spaces they are able to fit into.

3.1.1 Application of Nature v. Nurture Beliefs

These debates over the origin of gender differences have more real life implications outside of influencing gendered research. Preference for one side, biological or social, over the other actually has influence over the development of the very differences that they are looking to explain. As alluded to within the description of Spelke's pro-socialization argument, in-group psychology has a great power over personal identity. Therefore, it is to no surprise that an individual's beliefs in regards to where their psychological gender differences come from is

dependent, or at least aligns with that of the gendered personality traits they embody. Coleman and Hong (2006) found that women who believed in a biological gender theory (that differences are rooted in a biological disparity between the genders) are more likely to self-stereotype into their gender role. This was seen in a greater presence and a greater endorsement of even negative feminine traits (Coleman & Hong, 2006). It can be concluded, then, that the desire to understand the root of gender differences not only motivates continual debates within the realm of academia but also has the power to motivate gendered behavior in the real-world.

3.2 New Exploration of Gender Differences

While the debate over nature v nurture as justification for psychological gender differences has, and will continue to, permeate within social psychology, other debates of gender studies have simultaneously emerged. Understanding the extent to which gender motivates and influences the role of an individual is becoming more pragmatic.

For example, though Hyde (2005) indicated that aggression was one trait that was an exception to the meta-analyses indicating gender similarities (as per section 2.5), more current research has been conducted that has seemingly dismantled the gender of aggression. Björkqvist (2018) ran a meta-analysis of 148 studies on gender differences in aggression and ultimately found that while boys are more directly aggressive there is an equal presence of indirect aggression within both genders. Björkqvist critiqued the general attribution of aggression to men. He suggests that this common association is the result of not utilizing the most efficient way to report aggression, not accounting for the influence of social intelligence, and overly accounting for the association with testosterone.

Another recent exploration of gender psychology that also piggy-backs off of findings in Hyde (2005), is Hoff et al.'s (2018) meta-analyses of longitudinal vocational interests. In this

analysis, Hoff et al. explored the role of gender in interests, especially in terms of how this influence manifested overtime. Understanding individual interests can be pivotal is predicting occupations, and overall environment specific behaviors. With this in mind, understanding the role that gender plays in interests consequently indicates, to some degree, the implications that gender poses on larger life experiences of the individual. Previous studies had shown a stronger association between men and Realistic interests, and women to be more strongly interested in Social interests (Sue et al., 2009), and Hoff et al. looked to elaborate upon this study and to explore it conceptually in association with a developmental timeline. Looking at the variation in the interest intensity level throughout development from adolescence into adulthood found that the aforementioned distinction in interest between realistic and social interests remained, and even widened, during adolescence. Moving towards adulthood however, these differences reduced and the participants actually showed an interest in cross-gender interest expectations. This aligns with the idea that adolescents are much more concerned with their peers' opinions and again points to the power of in-group psychology that was introduced in section 3.1. Alternatively, adults are much more confident in asserting themselves as an individual. This analysis shows not only the immediate role of gender, but also the dynamic way that it can influence an individual – thus pointing to gender as a key variable in psychological exploration.

As indicated within these research examples, the exploration of the role that gender plays in the individual is continuous. These also are few of many examples of modern research that have transgressed the underlying debate of where gender differences come from, and instead explore how they develop and continue to impact the individual overtime.

3.2.1 Shifting Gender Research: Introduction of Non-binary Psychology

A motivation behind gender research is the hope of establishing an explicit binary between men and women. Section 2.3 is almost fully dedicated to this concept of searching for discrete differences between genders, as the idea of a gender binary has shaped the role of gender in scientific research (Hyde., 2019). Recently, however, a new approach to gender research has emerged as a pressing new avenue of research opportunity: the exploration of non-binary psychology. Today's society has become increasingly more accepting of disparate identity communities. As such, it is also becoming more wildly encouraged that "psychologists challenge the dominant binary assumption about gender and create environments that include and affirm non-binary individuals" (Matsuno & Budge, 2017). This critique of the systematic gender binary is not necessarily an entirely new concept, as seen by Mead (1935) declaring that most societies arbitrarily "divide the universe of human characteristics into two" (Delphy, 1993), alongside other early suggestions that gender precedes sex in regards to social roles (Delphy, 1993). Modern explorations of binary deviant behavior supports that this gender binary is culturally created (Hyde et al., 2019) and builds off these prior suggestions that there was a range in identification and is working to actually explore the psychological differences of individuals on this identification range. Such recent explorations have found, for example, low gender determinism within non-binary transgender individuals (Bradford & Catalpa, 2019). Additionally, a higher correlation between friend support and life satisfaction within cisgender and binary transgender individuals compared to non-binary transgender individuals suggests that there are further psychological differences within the different self-identification groups, thus showing even more support for continuing exploration of different identity groups (Bradford & Catalpa, 2019). Even though the exploration of non-binary gender identification research has

presented itself as a new avenue of research opportunity, there will still continue to be research is still rooted in finding concrete differences between men and women.

4. Replication Crisis

The previous section focused on the role of gender within psychological research, establishing an understanding of the major debates regarding gender. Moving forward, though, throughout this section I will describe the current limitations of psychological research and the replication crisis, as it has been labeled. Ultimately, this will lead me to a discussion of how this is particularly pertinent to research regarding gender differences.

4.1 Research in the Media

"7 Proven Health Benefits of Dark Chocolate" (Gunnars, 2018) and other articles of the sort frequent today's online media, consequently flooding consumers' heads with inflated information. Everybody is familiar with these click-bait headlines, and yet they cannot help but be drawn to the scientific justification to indulge; scientific proof seemingly trumps any former knowledge or intuition especially since people are prone to favoring easy and specific answers (Bohannon, 2015). What is allowing titles such as these to continue to riddle our sphere of information? While headlines are catering to both the desires of the consumer and the pressures of the industry, the science behind them is also caving to the pressure to produce palatable results.

4.1.1 How Research Becomes Headlines

Research is constantly being threatened by the influence of biases. While many of the concerns stem from the side of the participants and their responses, biases in research can also stem from the researcher's end. In the case of the studies such as the aforementioned chocolate example, much of the conducted research can be traced to having a heavy involvement from the

associated industry. Not only are the corporations the ones interested in the results, they are also the ones funding majority of the research (Cooper, Donovan, Waterhouse & Williamson, 2007). For obvious reasons, industries are interested in releasing data that promotes their agenda and will likely bring in more revenue and will invest accordingly. Furthermore, there is less likely to be an exploration of potential negative results. It does not behoove a chocolate company, for example, to perpetuate the negative health implications of their product. Accordingly, there is a pressure to publish positive results (Cooper, Donovan, Waterhouse & Williamson, 2007). This can also explain why so many more headlines that declare the novel, "healthy" side to chocolate are given more publicity than those that are antithetical to this cause. Even when these oppositional pieces do exist, they are not the articles that "get clicks" (O'Connor, 2018; Bohannon, 2015). This pressure to produce a specific range of qualifying results in order to be considered publishable leaves researchers a very small window of opportunity to accommodate the demands of their funding. As such, statistical manipulation serves as a solution in order to cater to such demands. The ability to make so many ongoing claims on the impacts of chocolate, for the sake of consistency, is grounded in the industries desire to find a result in the directionality they desire, not necessarily pertaining the specific content.

This concept is able to be transferred to psychology as well. While there are different actors at play, the same underlying dynamics permeate within psychological research. The demand for novel findings, the pressure to publish, and statistical manipulation are all also existent within psychology.

4.2 What the Replication Crisis is not: Extreme Cases of Data Malpractice and Fraud in Research

Despite the consequences that the replication crisis poses, there remains a distinction between the practices to be further explored within the next section and explicit fraud. To be clear: the replication crisis does not refer to explicit cases of fraudulent behavior. It instead, refers to malpractice that is often unintentionally harmful. While many instances in which the integrity of data collection, and therefore the resulting findings, has been compromised occur because of these naïve practices fueled by the motivation to publish, there are still cases in which the malpractice is drastic and intentional. These extreme instances, while they do not encapsulate the entirety or even the majority of the crisis at hand, still point to the overall flaws in the system and the detrimental effects that data misrepresentation can have. Dutch psychologist Diederik Stapel, for example, shook the world of psychological research when it was revealed that he had manipulated his data while completely fabricating research on human attitudes and behaviors (Verfaellie & McGwin, 2011). Stapel was accused, and reprimanded for, creating entire data-sets that matched his hypothesis rather than actually running studies. This behavior originally stemmed from Stapel's inability to get the desired significant results when studying the impact of priming on rating personal attractiveness, but continued throughout the remainder of his career in which he counterfeited research about human behavior (Bhattacharjee, 2013). With sound knowledge of what makes a reasonable study and what the ideal threshold of believable significance is – meaning he created data that had a significant result, but one that was small enough to not raise suspicion – while also being shielded by the safety of his growing academic status, Stapel survived a career built entirely on fraud until he, and his 55 fraudulent papers, became discovered in 2011 (Bhattacharjee, 2013).

Stapel's case closely followed the Harvard professor Marc Hauser who was also found having fabricated data in numerous of his studies on cognition in monkeys earlier in 2011

(Carpenter, 2012; Harvard, 2012). Unlike Stapel who admitted his wrongdoings, Hauser merely surrendered that he had "'let important details get away from my control'" (Carpenter, 2012) as his justification for his malpractice. Hauser recorded false values and altered coding in order to fit his personal theoretical predictions, and was reprimanded for malpractice even within some of his unpublished works on the grounds that he "misled collaborators in unpublished studies shows that this is a recurring pattern of behavior" (Carpenter, 2012). Another notable case of such extreme research fraud manifested in 2018, in which Brian Wansink was forced to resign from Cornell after having 6 (now 13) papers on food and nutrition retracted for "mistaken reporting, poor documentation, and some statistical mistakes" (Servick, 2018).

These cases of fraud, while they do exist, do not necessarily dominate the case studies in which data has been corrupted. In these extreme cases, however, they are "symptomatic of a broader problem" (O'Connor, 2018). Researchers are intentionally taking advantage of, and exposing the flaws in the current system. Ironically enough, Wansink was quoted in one of his papers years prior stating that "misinformation can have harmful effects on the health and economic status of consumers" (Wansink, 2005), showing the utter intentionality behind his behaviors; he was fully aware of the ramifications of skewing data representation and chose to do so anyway.

4.3 What is the Replication Crisis?

Modern psychology has become bogged down by this recurring question of what is reliable data? All fingers seemingly point to the replication crisis, a term coined to refer to the "failures to replicate several high-profile studies" (Stroebe & Strack, 2013) to be at fault. The replication crisis, as indicated by the title, refers to the inability to replicate findings and has consequently led to a growing number of published false-positive findings. Despite the ability to

replicate data being a "defining feature of science" (Open Science Collaboration, 2105), this basic function of checks-and-balances to verify what is being published no longer seems to be the basic standard of what is considered quality findings. While it is not a new concept, the replication crisis has become an increasingly popular criticism of much of the scientific world, especially in regards to psychology – noting that these criticisms bring forth implications for psychological studies on both the local and the systemic level.

There is not one sole factor that has allowed for the emergence of this "crisis", but instead a myriad of parts that compound to produce non-reproducible results. As introduced in section 4.1.1, statistical manipulation, lack of acknowledgement of null results, and the pressure to publish all contribute to the current prevalence of the replication crisis within modern psychology.

4.3.1 P-Hacking

Thus far, there have been references to skewing data or statistical manipulation without much further explanation. These aforementioned statistical alterations are more colloquially referred to as p-hacking, a practice that has become one of the predominant enablers of the replication crisis. P-hacking, as the name indicates, involves manipulating gathered data in order to achieve significance – noting that significance is often understood to be $p < 0.05$. There are many practices that p-hacking encapsulates such as: "conducting analyses midway through experiments to decide whether to continue collecting data, recording many response variables and deciding which to report post-analysis, deciding whether to include or drop outliers post analyses, excluding, combining, or splitting treatment groups post-analysis, including or excluding covariates post-analysis, and stopping data exploration if an analysis yields a significant p-value" (Head, Holman, Lanfear, Kahn, & Jennions, 2015). These different

categories that the researcher has to make decisions on are referred to as researcher degrees of freedom and can fully influence the outcome of the study, depending on what and when they are picked. Researcher degrees of freedom increase the probability of gathering a statistically significant result since there is a greater likelihood of at least one potentially observed variable falling within the threshold of being statistically significant (Simmons, Nelson, Simonsohn, 2011). Often times in fact, people are unaware at how significant the shifted outcome that may be as a result of p-hacking. Majority of researchers may be led to believe that they are working efficiently and using their resources effectively, when in turn they are being incredibly detrimental to their research. For example, imagine a researcher has already run their experiment on their originally assigned 50 participants. It can be incredibly attractive to merely add another 10 participants to the already gathered data in order to secure a significant result as opposed to running an entirely new round of their study with 60 participants. Comparatively, re-running the study with more participants may seem to be a waste of time and resources when there is seemingly the option to just retroactively add the 10 participants. However, as described below, running a small number of additional participants greatly increases the chances of false positives.

In an exploration of how much researcher degrees of freedom can impact the likelihood of obtaining a false-positive, Simmons et al. (2011) used a random computer simulator to generate "experimental data" from a normal distribution to then be assessed on predetermined researcher degrees of freedom. The motivation of the study was to observe how often the *p* value went below standard significance levels as the researcher degrees of freedom changed. The researcher degrees of freedom observed were having two dependent variables, increasing sample size by ten per cell, controlling for gender, and dropping or not dropping one of the conditions of the experiment. These researcher degrees of freedom are very common within research and, as to

be explored shortly, have a large impact on the found results. It is common, for example, to want to increase N at the completion of an experiment rather than spending the resources to run a whole other study. However, the Simmons et al. (2011) found that an act such as this, despite being out of want to conserve time and resources, jeopardizes the entire status of the findings.

Noting again that the data is random computer generated numbers, there was still a drastic increase in the probability of obtaining a false-positive when observing the assigned researcher degrees of freedom. This is especially true when the researcher degrees of freedom were combined. While there is a consistent increase in false-positive likelihood across the researcher degrees of freedom, the most notable false-positive rate occurred when all four researcher degrees of freedom were accounted for, in which there was a 60.7% false-positive rate (Simmons, Nelson & Simonsohn, 2011). This means that when these random numbers were subjected to all four behaviors, the likelihood of finding a significant result increased by nearly 61% is over twelve times the original chance (5%) of mistakenly finding a significant effect. Each manipulation of the individual researcher degrees of freedom increases the likelihood of finding a significant result too, showing how even one malpractice can skew data findings. This drastic increase found by compounding them, though, shows just how detrimental these practices can be to the credibility of research.

Additionally, sample size can also be a pawn in establishing the groundwork allowing for p-hacking. Comparatively, smaller sample sizes are considered to be more efficient in producing statistically significant results (Bakker, van Dijk, Wicherts, 2012) as opposed to one larger study. This ensues as a result of the small-scale studies being underpowered samples and being vulnerable to manipulation, again as opposed to the results from a stronger, larger sample;

smaller samples provide more opportunities to alter the design especially when they are recurring tests rather than a one-time large test (Bakker, van Dijk, Wicherts, 2012).

4.3.2 Null Results

The presentation of findings, as well as what findings are being chosen to be presented, has become a factor that enables the replication crisis. Null results are not only not celebrated, but they are often ignored or even rejected within psychological research and publications (Greenwald, 1975). As such, there has a "file drawer" phenomenon has developed meaning that researchers ignore, or file away, certain findings that do not promise publication (Mervis, 2014). This behavior of abandoning, or hiding, results has become a norm within psychological research in order to increase their access to publications, as to be explored within the next section. Unfortunately though, there are many implications surrounding the devaluation of null findings within reporting psychological findings.

First off, there is an initial aversion to finding null results in the first place which may lead to running analysis that favor one's own hypothesis (Ferguson & Heene, 2012). This implicates the findings with confirmation and publication biases that make any findings less reliable. Outside of this however, failure to acknowledge null results within the report of a study leads to future studies that are similar in nature. While some see this as a waste of resources, there are greater ramifications. With the right to hide their null findings, researchers allow room to re-run the study using more rudimentary approaches that enable themselves (or others) to find significant results, such as using a smaller sample size (Mervis, 2014). All the while, the research that does find significant results pertaining to the subject are likely to seem stronger since there are not null results counteracting the new findings (Mervis, 2014). Additionally, not publishing null findings produces a limitation on replicability since the replications are not always

meaningful when the failed attempts are not acknowledged (Ferguson & Heene, 2012). This partnered with the increased strength of future similar studies (Mervis, 2014) makes for a problematic duo in terms of the quality control of the research.

Ultimately, this failure to report null findings grants researchers the ability to cherry-pick what findings they want to report and be associated with. This allows them to completely change the way they frame their experiment based on the outcomes rather than their intentionality. Discrimination to null hypothesis in publications is detrimental to research progress (Greenwald, 1975) and points to the severity of the pressure to publish within the psychological world.

4.3.3 Pressure to Publish

The large number of teams involved in running social psychology research shifts the nature of the field to revolve around "beating the competition" (Ioannidis, 2005); researchers want to be the first to publish something notable, as well as publish the most notable theories. This rush to produce novel results can consequently jeopardize the accuracy of the said results. Since there is an emphasis on obtaining a position of power in comparison to their competitors, research is rooted in finding either new positive results, or negative results that counterbalance that of the competition (Ioannidis, 2005), as acknowledged in the previous section, there is a lack of celebration of null results despite them also being informative. As such, again similar to in the food research example early in this section, the urgency to produce publishable data allows for desperate researchers to be more prone to make statistical errors.

This pressure to publish also influences the research from not only the production standpoints, but also the reviewers' perspective. The ability to reproduce data is vital in holding research accountable and for validating studies, other words there would be few regulations to what qualifies as publishable data. In order to replicate a data-set, another set of researchers must

conduct an identical study – noting that this requires an expenditure of equal resources with little to no chance of getting published. The inability to publish replicated data is rooted in the desire to only publish novel results, almost directly paralleling the concepts articulated in the exploration of how null results are handled. Because of this inability to publish, some view these direct replications to be a waste of funds and resources (Stroebe & Strack, 2013). Furthermore, the little promise of publishing any type of direct replication incredibly reduces the incentive of other researchers to continue to replicate studies (Simmons, Nelson, Simonsohn, 2011). This viewpoint from the reviewers creates a complication in a basic but necessary precautionary step in publication, and is jeopardizing the accuracy and validity of research for selfish motivations. This is directly a result of the increasingly competitive nature of the social psychology field: "these problems have been exacerbated in recent years as academia reaps the harvest of a hypercompetitive academic climate and an incentive scheme that provides rich rewards for overselling ones work and few rewards at all for caution and circumspection" (Pashler & Wagenmakers, 2012). As a recent and potential solution to mitigating the dichotomy between wanting to publish and wanting to review, there is a value to conceptual replications over direct (Pashler & Harris, 2012). In doing so, review researchers can have the essence that they are checking the research while they are in reality just expanding upon it; there is much less actual review. These practices appease the reviewers feeling of obligation to replicate the data while also still granting them the opportunity to publish. However, these conceptual replications do not accurately test or portray to ability to reproduce the data, and are therefore not acting as proper quality control.

The higher degree of push-back within the replication process has disrupted the checks and balances system. As such, results that typically would have been deemed invalid remain

viable in the psychological field. This is, consequently, essentially increasing the longevity of the relevancy of these studies. Many people still believe that skewed data will be filtered out by new research overriding it and nudging it out of relevance (Pashler & Harris, 2012; Bakker, van Dijk, Wicherts, 2012). However, without immediate, and appropriate, replication tests the likelihood of a study getting flagged as problematic significantly reduces. Efforts to replicate and verify data are most commonly exclusively aimed at modern studies. Therefore, a false-positive report will remain published if it is not de-bunked within relative approximation to when it is originally published (Bakker, van Dijk, Wicherts, 2012).

4.4 Contradictory Views on the Severity of the Replication Crisis

Despite the prevalence of the ongoing discussion of the replication crisis within modern psychology, not everyone shares the sentiment that it is a risk to the field.

There is first a question of whether the replication crisis is even relevant enough to be considered a major concern. Between the belief that because of the low alpha level ($p=.05$) there is a reasonable boundary on the production and published rate of errors, suggesting that no more than 5% of positive findings include errors (Pashler & Harris, 2012), and the small percentage of researchers expected to perform questionable research practices (Fanelli, 2018), there seems to be little belief that this is more than a minor issue. As such, the language surrounding the 'replication crisis' which makes it sound impending and significant, is inaccurate since many believe that most literature is not actually distorted (Fanelli, 2018), as justified by the aforementioned reasons, and should therefore not be taken as such a serious threat. Another qualm surrounding the language of the replication crisis is that it is presented as being generalizable. Some fields may be impacted by the statistical manipulations more heavily than others, since statistical power varies between subfields and low statistical power increases risk of

false-positives (Fanelli, 2018). It is therefore unfair to use apply the idea of a crisis to subfields that are not hardly at all likely to be affected.

Other instances that discredit the severity of the replication crisis include objecting the scenario depicted within 4.3.2 in which null findings are not reported. Instead of believing this idea of cherry-picking what results are being presented, others believe that null hypothesis are in fact included within publications just not within the abstracts or titles of the papers. In this way, the papers are still being presented as interesting and the null results are still being accounted for and reported so it is not a problematic feature of the publication (Fanelli, 2018). Additionally, the idea that spoiled results will be reduced and "pruned out" over time is recurring. Previously introduced in section 4.3.3 in regards to the severity of such inclinations, this lack of concern about the mere presence of misrepresentative research is justified by arguing that new research will eventually override it and nudge it out as time goes on (Pashler & Harris, 2012). As previously stated though, this is not the case since replication efforts are tailored to modern research, so once it misses its threshold of being a study of interest, it is no longer really at risk of being de-bunked.

There are also debates on the role of the replication crisis in the future of science and psychology. While some believe that it is motivating overall better practices, others believe that the current situation is actually highlighting the already in use best practices: "rather than undermining science it is reaffirming the best practices of the scientific method" (Loken, 2019). This is then suggested the positive impact that the replication crisis is having without suggesting or demanding an explicit need for reform. Conversely, some believe that calling attention to the replication crisis is consequential to the future of scientific research. Instead of motivating change and innovation in the future of research, it is believed that the modern emphasis on the

replication crisis will project negativity onto future generations of researchers, "fostering in them cynicism and indifference" (Fanelli, 2018).

Another argument against the gravity of the replication crisis is actually formulated in regards to replication as means to validate results. While some view direct replication as a waste of funds and resources (Stroebe & Strack, 2013) since you are not seemingly obtaining anything new from re-running an identical study, others value conceptual replications over direct replications (Pashler & Harris, 2012). Advocating for conceptual replication is disguised as revolving around strengthening validity and generality when in reality it is most likely the result of positive conceptual replication studies being more publishable than positive replication in a direct study; there is nothing novel about a direct replication study and publications are drawn to surprising data (Pashler & Harris, 2012). This creates an interesting dichotomy in that it is not that other researchers are neglecting to attempt to replicate the study, they are just doing so in a manner that does not necessarily reproduce the data but rather expand upon it. These motivations to find publication, though committed under the pretenses that the replication crisis is a mundane worry of the psychological field, ultimately are feeding into the very problem; the people that do not acknowledge the severity of the replication crisis are actually among the perpetuators.

Despite the controversial opinions surrounding the severity of the replication crisis, studies such as the one run by the Open Science Collaboration (2015) in which only 36% of 100 already published findings were replicated as significant, as depicted within Section 1, show the reality and severity that the replication crisis poses on modern psychology.

4.5 Gender Studies Reliability at Risk

The depiction of the parameters and ramifications of the replication crisis throughout this section thus far have set the stage for understanding the way such data manipulation occurs. The

particular impact of the replication crisis on gender research, though, is pinnacle to the scope of this paper; in order to further understand how to better protect gender research from falling victim to the replication crisis, it must first be understood the extent to which the crisis particularly affects it.

A large motivating factor in cases affected by the replication crisis is the large number of researchers in the field. The prevalence of gender within psychology has been accredited numerous times as being a hot topic of exploration. Unfortunately, this allows for it's data to be particularly vulnerable to not being replicated: "the hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true" (Ioannidis, 2005). This is because with so many people pining to find significant data, eventually a number of them will and will therefore report it accordingly, failing to the report the null findings.

Not only is gender a variable of interest since it has such a grand history of being thought to influence social roles, but it is also an omnipresent variable to explore. This meaning that the gender of participants is easily accessible to be collected, and often is, even when there are no hypotheses about it – it is an easily accessible, but powerful researcher degree of freedom. Further subsequent analyses may be run and reported when found significant, creating an abundance of literature rooted in false-positives. This follows the prior discussion of Simmons et al. (2011) in which it was shown that additional analyses and reporting for variables not originally intended for exploration can substantially increase the potential of false positives; controlling for gender in fact led to a more than double potential of finding a false-positive.

5. Solutions to Mitigate the Impact of the Replication Crisis in Gendered Studies

Thus far the severity and the ramifications of the replication crisis have been explored. With the reputation and validity of social psychology as a whole (Simmons, Nelson, Simonsohn,

2011; Stroebe & Strack, 2013) being threatened there is an on ongoing demand for a solution to the replication crisis in order to minimize further discrediting within the field. Not only is there a demand to stop the spread of non-replicable data, but there is also a demand to increase the quality of research as a whole. Social psychology in particular has a dense number of researchers and a high frequency of published works, which have been identified as features that make it vulnerable to falling victim to the replication crisis. Tightening the scope to re-focus on gender specifically, there is an exceptional need to produce more quantity of more quality research. Gender has been explained as being particularly vulnerable to having false-positive reports since it is both always present and there is a high demand to find and understand concrete gender differences (Science of Gender). The ongoing debates surrounding gender research demand quality research to serves as the grounds of these arguments. As such, while some of the more general systemic suggestions remain valid, there is the need to be more specific with the regulation of future studies to produce more reputable – and replicable – data.

5.1 Existing Solution Suggestions

Despite the public scrutiny that psychological research has faced, researchers have yet to self-motivate a solution to mitigate their malpractices. With so much at stake – the credibility of the entire field of psychological research – if a solution does not decelerate the current upward trajectory of invalid results, there is a pressure to disrupt the recurring behaviors that allow for researchers to slip into the replication crisis. Before moving towards a more pointed solution aimed specifically at increasing the quality of gender research specifically, it is valuable to create an understanding of the current core practices that are being suggested as solutions to the replication crisis. It is important to note that this section is not attempting to solve the replication crisis in its entirety or depict one single solution as being the "perfect solution." Instead,

understanding the basis of how the replication crisis is being approached can create the infrastructure for understanding the gender specific suggestion.

5.1.1 Pre-Registration and Improving the Statistics of Reported Data

One suggested solution is to tighten the constraints on researchers prior to undergoing research. This solution refers to the practice of pre-registration, which involves researchers reporting their research plan prior to running the actual study. The early submission includes a thorough articulation of the intended study: the hypothesis, the design, the number of subjects, the analytical plans, the rules for excluding data, the research rationale, etc. (Gonzales & Cunningham, 2015; Association for Psychological Science, 2016). For all intents and purposes, pre-registration requires the research to submit their research in as much detail as possible prior to actually undergoing the planned research.

Pre-registration proves to be a helpful solution in reducing the variability of researcher degrees of freedom. Since the intentionality and the explicit rules for data collection and data termination are gathered prior to undergoing actual research (Simmons, Nelson, Simonsohn, 2011), there is little room to allow researchers to perform behaviors such as add more people to their sample without running a completely new study, run for effects on not originally accounted for variables. These examples of behaviors are understood as having significant impact on the outcome of the data, as explored within section 4.3.1. Therefore, reducing the ability to perform such behaviors consequently reduces the ability to find false-positives as a result of delayed account of these researcher degrees of freedom. There is also the expectation that having data be able to be rejected by not following their originally presented plan rather than on the basis of the presented results would also decrease the number of false-positives (Gonzales & Cunningham,

2015). The idea behind this is that there will be a stronger emphasis on the research, and presented sound justifiable explorations, rather than just on the results.

Having the intention of the researcher on hand prior to the study also increases the reviewers' ability to hold them to these research plans. By doing so, there is an unambiguous standard and expectation of the researcher's technique both to be used by the reviewer as the standard on what the research should include. However, the concept of further holding reviewers accountable in the replication process will be further explored later in the paper.

Shifting from reporting prior to running the experiment to the report of the research after it has been conducted, it has also been recommended that all variables that were collected to any degree are listed in the report. In the implementation of this suggestion, all conditions, including those that failed, are included in the report; all results prior to deleting observations must be included in the report; and all results found without using a covariate will be included in the report, even if the ultimate analysis includes a covariate (Simmons, Nelson, Simonsohn, 2011). The reporting of these features is not meant to hinder the ability to properly analyze data or diminish results that rely on some of the mentioned techniques. Rather, they can be seen as highlighting the need or benefit of some analytical features such as using a covariate or finding what data exclusion is warranted, since the raw results will show the necessity of undergoing these statistical processes (Simmons, Nelson, Simonsohn, 2011). These reports are suggested with the intention of facilitating more transparent communication of what the findings are – the researcher is in essence forced to account for every step that they took in order to get to the results that the ultimately published. Encapsulating what the research actually found, rather than merely what the desirable findings were, allows for transparent recognition of overly manipulated data since it is clearly juxtaposed next to its raw counterpart.

Relating directly to the actual practice of running studies, however, there is a demand to increase the sample size expected within each study. As previously discussed in Section 4.3.1, smaller sample sizes are "more efficient" in producing statistically significant results and provide more opportunities to manipulate the design (Bakker, van Dijk, Wicherts, 2012; Simmons, Nelson, Simonsohn, 2011): "if you measure a large number of things about a small number of people, you are almost guaranteed to get a 'statistically significant' result" (Bohannon, 2015). Larger sample size conversely strengthens the power of the evidence – as long as the size is in respect to what is being explored. Running large-scale research for every posed research question is both impractical and illogical. It burns through resources and is not valuable when observing narrow markets, (Ioannidis, 2005). Even so, the expected sample sizes should be increased to some degree. Regardless of the inability to run "large-scale" projects, it would behoove the psychological field to implement the use of larger, more randomized studies to help inherently avoid false-positive (Ioannidis, 2015). As seen here, these solutions are not meant to work in isolation but rather in conjunction with one another to increase researchers responsibility for their results.

5.1.2 Equal Accountability for the Editing Party

While the researchers that are falsely reporting their inflated data findings are to blame for much of the replication crisis, the reviewers are not innocent. Instead, they can be viewed as enabling the malpractice behaviors and should also require reform in their role in the replication crisis. 'It takes two to tango' as they say, and while it is the researchers that are doing the manipulations, it is the reviewers that are being negligent; it is the onus of the reviewers to put aside their personal desire to publish and hold researchers accountable (Bakker, van Dijk, Wicherts, 2012). This responsibility first and foremost manifests in the reviewers enforcing

requirements (Simmons, Nelson, Simonsohn, 2011). Reviewers have an authority within psychological research. As such, researchers will aim to please their reviewers when they are held to consistent, systematic requirements. This is where again the benefits of pre-registration come in, since there are explicit standards for the reviewer to hold the research to. Along this vein then, reviewers should also demand more from the researchers, not just in following the guidelines but in terms of exemplifying that the results are not arbitrary and that their analysis are compelling. In order to ensure such behaviors, reviewers should not shy away from the value of exact replications over conceptual replications (Simmons, Nelson, Simonsohn, 2011).

5.1.3 Changing the Role of Journals and Publications in Research

A explored within the previous sections 4.3.2 and 4.3.3, there is a common misconception that only novel and significant findings are valuable. While this notion permeates amongst researchers and reviewers alike, it is ultimately rooted in the way that publishers and journals approach research. Publication is one of the ultimate levels of status to be reached by a researcher, so naturally they strive to fit their research within the confines of what journals want: exclusive, significant findings. Again, as previously explored, this leaves researchers desperate to appease the publishing journals, "yielding to the pressure to do whatever is justifiable to compile a set of studies that we can publish" (Simmons, Nelson, Simonsohn, 2011). Researchers justify their behaviors of malpractice and resort to p-hacking out of desperation to achieve this high status of significance.

Clearly, the publishers carry a large amount of clout within the psychological world, and should therefore also be accounted for in the step towards mitigating the replication crisis. How can this pressure from the publishers be channeled to a healthy, productive power within research?

One suggested solution is that publications be based on out-come neutral criteria (Gonzales & Cunningham, 2015). The impact of this was alluded to within section 5.1.1 as it can be understood as being a feasible standard from the implementation of pre-registration. If reviewers were more prone to reject a study based on its failure to abide by its pre-registered terms rather than the ultimate findings, the entire focus of the study would shift. Currently, studies are rejected, or turned away from publishing, on the basis of their results. In this alternative solution however, they would be turned away because of the core of the research. This would have a subsequent impact of increasing the emphasis back towards the research – the motivations behind it, the core practices involved, the intrigue of the questions asked (Gonzales & Cunningham, 2015). With less emphasis put on strictly finding significant results, there would be less motivation for researchers to manipulate their data in such strenuous ways. Publication would be more accessible, and the research would be more pure.

Additionally, the world of psychology would have to become "more tolerant of imperfections in results" (Simmons, Nelson, Simonsohn, 2011). It is one thing to not automatically reject research on the grounds of the findings, but a completely different thing to accept that not all research questions and findings will procure interesting results. There are still things to be learned and celebrated from psychological findings of all levels, and in order to revert the emphasis of research back on the research itself, the emphasis on the results much truly undergo a cultural shift.

5.1.4 Increase Political Diversity in Social Psychology

In addition to adjustments made on the individual project level, there are also potential benefits of altering the scope of the psychological field as a whole. Increasing the political diversity within social psychology proves to behoove the field and the validity of its research.

Psychology, and social psychology specifically, is at risk of becoming a homogenous, liberal, community as indicated by the 84% psychology professors identifying as liberal (Duarte, Crawford, Stern, Haidt, Jussim, & Tetlock, 2015). This manifests itself as a threat to social psychology in that vigilant critics from a refreshed perspective will be replaced with reviews from people that have the same expectations and motives. Confirmation Bias, or the tendency to support or be partial to pre-existing beliefs (Nickerson, 1998), has the ability to affect even empirical data; data can be manipulated and interpreted according to the desired outcome – as seen recurring through the replication crisis. Diversifying the perspective and motivations of not only the researchers but also the reviewers, then, would in turn reduce this straightforward homogenous motivation and work as a checks and balances system both on the larger and local scale; results would be more surely protected from errors and there would be a faster correction time, if errors were to occur (Duarte, Crawford, Stern, Haidt, Jussim, & Tetlock, 2015).

While this proposition alone shows potential, understanding the impact of politics on psychology is vital in appreciating this as a potential solution, or part of the solution rather, of deconstructing the vulnerability to the replication crisis. Political biases are most relevant within certain sectors of social psychology that parallel party agenda's and are surprisingly enough not limited to the participants of the experiments. While obviously biases that are more salient within certain political associations may prove to influence responses or behaviors within studies even when this is not what is being explicitly observed (i.e. responses to people of color by conservatives or Christians by liberals), these same salient biases influence the researchers. This manifests both in the questions they are exploring and the interpretation of the results (Duarte, Crawford, Stern, Haidt, Jussim, & Tetlock, 2015). Meaning, since social psychology has presented as a densely liberal space, many of the research questions will be motivated to support

the liberal agenda, and as previously alluded to, the results are victim to confirmation bias. This is relevant specifically to gender studies since it can parallel traditional views compared to more modern opinions of gendered stereotypes, even to go as far as to parallel sexist behaviors (Duarte, Crawford, Stern, Haidt, Jussim, & Tetlock, 2015) – a topic that is clearly a point of tension and disparity between liberals and conservatives. Therefore, as previously stating, by diversifying the pool of social psychologists there would be more questions being asked, more perspectives being used in analysis, and more criticisms of results all of which would reduce the directional manipulation of the research, subsequently improving the validity of said results (Duarte, Crawford, Stern, Haidt, Jussim, & Tetlock, 2015).

5.2 Gendered Research Specific Solution

Most of the aforementioned solutions are applicable to general models of shifting the dynamics within psychological research that make it vulnerable to being manipulated. While the implementation of these solutions have promising attributes and potential in minimizing the effects of the replication crisis, the exploration of gender in psychological research should require further more specific requirements to reduce future and further exploitation of gender as a variable for statistical significance inflation. It is important to note that the following suggested solution is not attempting to solve the replication crisis as a whole, nor is it posing as being the sole perfect solution. Simply put, the following suggestion is looking to improve the quality, and the frequency at which it is able to be obtained, of data within gender research. Using the previously discussed solutions as motivation, I propose the following solution be implemented in order to do so.

Moving forward, it should be required that gender/ sex always be recorded within psychological research but can only be reported if it was pre-registered. Additionally, these findings should be reported regardless of significance.

This solution can be seen as the most viable solution for a number of reasons. First, by instilling this permanence to sex and gender as a variable within psychological research, it is stripping researchers the opportunity to run these analyses after data collection as a last-ditch effort to find significance. As such, the increased rate of finding a false-positive when gender was controlled for after the fact, as shown within Simmons et al. (2011), would be diminished.

Requiring pre-registration in order to report analyses of gender is also a key factor in the success of this suggested solution. This requirement will help keep the studies aligned with their intended purpose, as only studies that intended to explore the role of gender will be able to report on it. Even though gender and sex should always be recorded as a way to reduce p-hacking, not every psychological study will be, nor should it be, centered around finding gendered results – only those intended to explore gender should be focused on doing so. By requiring the pre-registration then, it forces the researcher to hone in and focus on the core purpose of the research they are running rather than allowing them to look to report any variable they find significance with. This also adds a validity to gender studies since they will have been run intentionally so.

This requirement is also vital from the reviewers' perspective. If a researcher were to report every variable that was accounted for in their study, the publications would be bogged down with unnecessary data. This would significantly lengthen reports without adding much value in many cases, since there are infinite more potential variables to collect compared to ones of value within a study. By only being allowed to report on the variables that are pre-registered,

researchers are forced to keep their publications concise and to the point, rather filled with arbitrary findings.

Using sex and gender as a mandatory base variable also increases the potential to further comparative explorations of these variables. This allows for the possibility of transcending original nature v nurture debates, ultimately resulting in a much more modern understanding of the different psychological implications of gender compared to sex – the social identity v the biological identity. Research could include, for example, the exploration of psychological effects when there is a lack of congruency between the two identities, or whether traits and behaviors are more greatly associated with a gender compared to a sex. The constant recording of both variables could also give insight on when gender identity becomes salient to individuals, especially when it does deviate from their sex.

Additionally, sex and gender being accounted for on every psychological study creates a large database for potential future meta-analysis to further explore the differences. Meta-analysis is beneficial in testing the quality of past designs and reducing errors (Crombie & Davies, 2003), especially Type II errors, being the failure to detect a present effect which are exceedingly present within traditional data analysis (Schmidt 1992). Importantly though, meta-analysis minimize biases in the synthesis of data (Crombie & Davies, 2003). This again parallels the motivation of the solution of reducing biases in social psychology. Studies of differences in males and females and men and women are privy to biases from both the researchers and the participants, which could skew the data in a traditional sense. As seen within the studies depicted within section 3.2, meta-analyses are increasing valuable the ongoing assessment of gender differences, and thus as much data as can be surmised to allow for such efficient analyses should be compiled and made available.

Unfortunately, while this suggestion is an especially viable one it still has flaws in its application. Many researchers may not want to, or will find it unnecessary to record for gender when it is not pertinent to their specific research. Additionally, some may question how gender obtained such status over other variables. While this solution is being used here specifically in terms of gender, it is plausible that it also be applicable to other variables such as race and age, or even attributes like height and eye color – the number of variables that researchers can collect is endless. It is beyond the scope to of this paper to weigh in on any traits besides sex and gender specifically. However, since this paper is advocating for the mandatory recording of gender/ sex, it is important to ascertain why this variable above others is worthy of perpetually being recorded. Why does gender matter so much to always record for it as opposed to other variables? While the other factors may also have interesting implications and influences on the individual, gender is unique in that it has many dynamics that make it subject to affecting many avenues of the individual. Gender, as previously enforced is omni-present, and can have both biological and socio-cultural roots while also affecting both the psychology of the individual and the group. For this reason, the large reach that gender has allows for it to be a viable variable to always be recorded.

There is expected push-back to this permanence of recording gender as opposed to other variables, but the understanding of the potential role that gender has in so many aspects of human lives, as well as the high demand of gender research compared to other variables can point to the reason why gender is relevant enough to always be recorded for. Additionally, for those skeptical of always having to record for a variable that may not be pertinent to their specific psychological research, recording gender takes little to no additional resources since it is

in fact omni-present. It also can easily be recorded at the end as to avoid any potential priming or biases that may influence the actual intentionality of their study.

Understanding the validity of this particular solution can be reinforced when comparing it to other possible solutions. It is important to note here, that the number of potential solutions is ongoing, as many different niche stipulations can be drawn within each one, so only relevant ones similar in nature to the one proposed will be discussed.

Another potential solution would be to always record and also always report gender within research. This solution, similar to the aforementioned, reduces the possibility of false-positive gender results since it physically cannot be run as a variable after As such, this increases the reliability of gendered data. However, this solution in particular, threatens to significantly bog down the research. If a study is not about gender, there is no reason to include it as such in the report. Doing so would simply muddle the report, burying the actually relevant significant findings behind a mass report of irrelevant data. This not only increases the length of the reports but also takes away the ability to emphasize the actual findings of the study. It also poses as a threat to what qualifies as gendered research and additionally complicates the ability to find quality gender research. As such, this is not considered a particularly applicable solution.

Alternatively, researchers could be required to always record gender but only report it if significant findings were found. This solution also reduces the use of gender as an “afterthought” variable in seeking significance. However, it does not increase the intentionality of gender reports. Instead, this fosters an approach that is relatively similar to the current approach to null results (section 4.3.2); the researchers would have access to pick and choose what variables and what findings they want to present. This would skew the realm of gender research to include studies that were not developed with gender in mind, and therefore validate methodologies that

happened to find a significant role of gender. Previously, the many stipulations that each solution leaves room for was acknowledged, and this particular suggestion lends its hand to an interesting number of them. For example, within this solution gender could be required to always be recorded, only reported when significant, but only allowed to be used in the title if it were pre-registered. A stipulation such as this is still subject to the researcher picking and choosing what they are reporting, but it is increasing the accountability in regards to the intentionality of the study. Research would have to be presented in terms of what it is intended to find, rather than publicize itself based on what it finds. Even still, this particular solution leaves too much variability in how researchers are able to present their data, again leaving gender vulnerable to being misrepresented.

After comparison to these recently described solutions, the initially discussed suggestion of always recording gender but only reporting the results if it was a pre-registered variable, presents itself as an incredibly viable solution in an attempt to mitigate the impact of the replication crisis and instill a future of quality gender findings. Furthermore, this solution points to instilling a future of continuing psychology's legacy of exploration of the role of gender on the individual and society.

6. Conclusion

Throughout this paper, the significance of gender as well as the implications of the modern replication crisis was explored. The relationship between the prevalence of gender within psychological research and the ramifications of the limitations of current psychological research on research about gender specifically, point to the need for a systemic reform. The current state of psychological research has left the complicated debates surrounding gender at risk of being grounded in unreliable evidence. As such, the need for a new solution was

highlighted. Ultimately, the benefits of the new, gender specific solution to increase the number of quality results in face of the replication crisis were discussed. Within this new expected framework, it is expected that gender always be recorded within psychological research, but is only reported if it was pre-registered as a variable of interest. The implementation of this particular solution suggests a hopeful shift in psychological gender research away from being susceptible to the replicable crisis, and instead allows for the debates surrounding gender to be motivated in sound, reputable data.

Author Contributions

The concept of this thesis was developed from discussions between Shelby Kennedy and Mark Sheskin. Sheskin continuously provided Kennedy with feedback and supplied suggested literature to include in her review. Sheskin also created the title.

Acknowledgements

I would like to thank Mark Sheskin who advised me through this process for ideation of the topic to the completion of the paper and every step along the way. Without his ongoing assistance throughout this process, this project would not have been possible.

References

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.

Bhattacharjee, Y. (2013, April 26). The Mind of a Con Man. Retrieved from https://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html

Bohannon, J. (2015, May 27) I Fooled Millions Into Thinking Chocolate Helps Weight Loss. Here's How.

Bombardieri, M. (2005, January 19). Harvard Women's Group Rips Summers. Retrieved from http://archive.boston.com/news/education/higher/articles/2005/01/19/harvard_womens_group_rips_summers/

Bradford, N. J., & Catalpa, J. M. (2019). Social and psychological heterogeneity among binary transgender, non-binary transgender and cisgender individuals. *Psychology & Sexuality*, 10(1), 69-82.

Carpenter, S. (2012, September 6). Harvard Psychology Researcher Committed Fraud, U.S. Investigation Concludes. Retrieved from https://www.sciencemag.org/news/2012/09/harvard-psychology-researcher-committed-fraud-us-investigation-concludes

Coleman, J. M., & Hong, Y. (2008). Beyond nature and nurture: The influence of lay gender theories on self-stereotyping. *Self and Identity,7*(1), 34-53. doi:10.1080/15298860600980185

Cooper, K. A., Donovan, J. L., Waterhouse, A. L., & Williamson, G. (2007). Cocoa and health: A decade of research. *British Journal of Nutrition,99*(01). doi:10.1017/s0007114507795296

Cosmides, L., & Tooby, J. (1997). Evolutionary psychology: A primer.

Crombie, I.K. , & Davies, H.T. 2003. What is meta-analysis? Evidence-based medicine: What is...? series, second edition. Retrieved from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/Meta-An.pdf

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., ... & McGhee, K. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular psychiatry*, *16*(10), 996.

Delphy, C. (1993, January). Rethinking sex and gender. In *Women's Studies International Forum* (Vol. 16, No. 1, pp. 1-9). Pergamon.

Duarte, J., Crawford, J., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. (2015) Political diversity will improve social psychological science. *Behavioral and Brain Sciences, 38,* E130. doi:10.1017/S0140525X14000430

Dunham, Y., Baron, A. S., & Banaji, M. R. (2015). The development of implicit gender attitudes. *Developmental Science*, **19**, 781–789. https://doi.org/10.1111/desc.12321

Eagly, A.H., & Wood, W. (2013). The nature-nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspectives on Psychological Science*, 8, 340-357.

Fanelli, D. (2018, March 13). Opinion: Is science really facing a reproducibility crisis, and do we need it to? Retrieved from https://www.pnas.org/content/115/11/2628

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*(6), 555-561.

Fryar CD, Gu Q, Ogden CL, Flegal KM. Anthropometric reference data for children and adults: United States, 2011–2014. National Center for Health Statistics. Vital Health Stat 3(39). 2016.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin,82*(1), 1-20. doi:10.1037/h0076157

Gonzales, J. E., & Cunningham, C. A. (2015, August). The promise of pre-registration in psychological research: Encouraging a priori research and decreasing publication bias. Retrieved from https://www.apa.org/science/about/psa/2015/08/pre-registration

Gunnars, K. (2018, June 25). 7 Proven Health Benefits of Dark Chocolate. Retrieved from https://www.healthline.com/nutrition/7-health-benefits-dark-chocolate

Hallgrímsson, B., Willmore, K., & Hall, B. K. (2002). Canalization, developmental stability, and morphological integration in primate limbs. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5217179/

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The Extent and Consequences of P-Hacking in Science. *PLOS Biology,13*(3). doi:10.1371/journal.pbio.1002106.

Hoff, K. A., Briley, D. A., Wee, C. J., & Rounds, J. (2018). Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies. Psychological bulletin, 144(4), 426.

Hyde, J.S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, 581-592.

Hyde, J.S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science*, 16, 259-263.

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist,74*(2), 171-193. doi:10.1037/amp0000307

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

Kanazawa, S. (2008, November 16). Common misconceptions about science I: "Scientific proof". Retrieved from https://www.psychologytoday.com/us/blog/the-scientific-fundamentalist/200811/common-misconceptions-about-science-i-scientific-proof

Lewontin, R. C., & Nelson, R. (1984, October 25). Plato's Women. Retrieved from https://www.nybooks.com/articles/1984/10/25/platos-women/

Linda A. Jackson & Kelly S. Ervin (1992) Height Stereotypes of Women and Men: The Liabilities of Shortness for Both Sexes, The Journal of Social Psychology, 132:4, 433-445.

Loken, E. (2019, April 09). The replication crisis is good for science. Retrieved from https://theconversation.com/the-replication-crisis-is-good-for-science-103736

Marc Hauser "Engaged in Research Misconduct". (2012, September 05). Retrieved from https://harvardmagazine.com/2012/09/hauser-research-misconduct-reported

Matsuno, E., & Budge, S. L. (2017). Non-binary/Genderqueer Identities: A Critical Review of the Literature. *Current Sexual Health Reports,9*(3), 116-120. doi:10.1007/s11930-017-0111-8

Mead, M. (1935). *Sex and temperament in three primitive societies* (Vol. 370). New York: Morrow.

Mervis, J. (2014, August 29). Why null results rarely see the light of day. Retrieved from https://science.sciencemag.org/content/345/6200/992

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220. http://dx.doi.org/10.1037/1089-2680.2.2.175

O'Connor, A. (2018, September 29). More Evidence That Nutrition Studies Don't Always Add Up. Retrieved from https://www.nytimes.com/2018/09/29/sunday-review/cornell-food-scientist-wansink-misconduct.html

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

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531-536.

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

Pryzgoda, J. & Chrisler, J.C. Sex Roles (2000) 43: 553. https://doi.org/10.1023/A:1007123617636

Reis, H.T. & Carothers, B.J. (2014). Black and white or shades of gray: Are gender differences categorical or dimensional? *Current Directions in Psychological Science*, 23, 19-26.

Servick, K. (2018, September 21). Cornell nutrition scientist resigns after retractions and research misconduct finding. Retrieved from https://www.sciencemag.org/news/2018/09/cornell-nutrition-scientist-resigns-after-retractions-and-research-misconduct-finding

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

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59-71.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: a metaanalysis of sex differences in interests. Psychological Bulletin, 135(6), 859-884. http://dx.doi.org/10.1037/a0017364

Taylor, C. C. (2012). The Role Of Women In Plato's Republic. *Virtue and Happiness,*74-87. doi:10.1093/acprof:oso/9780199646043.003.0005

The Science of Gender And Science Pinker Vs. Spelke A Debate. (n.d.). Retrieved from https://www.edge.org/event/the-science-of-gender-and-science-pinker-vs-spelke-a-debate

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47*(10), 1173-1181. http://dx.doi.org/10.1037/0003-066X.47.10.1173.

Uzgalis, William, "John Locke", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.) URL = <https:??plato.stanford.edu/archives/spr2019/entries/locke/>.

U.S. Census Bureau Health and Nutrition. (2011). Statistical Abstract of the United States: 2011; Cumulative Percent Distribution of Population by Height and Sex 2007 to 2008.

Verfaellie, M., & McGwin, J. (2011, December). The Case of Diederik Stapel. Retrieved from https://www.apa.org/science/about/psa/2011/12/diederik-stapel

Wansink, Brian, Position of the American Dietetic Association: Food and Nutrition Misinformation (August 14, 2005). Journal of the American Dietetic Association, 106:4 (April), 601-607, 2006. Available at SSRN: https://ssrn.com/abstract=2714473.

What is heritability? - Genetics Home Reference - NIH. (2019, April 16). Retrieved from https://ghr.nlm.nih.gov/primer/inheritance/heritability

What Is Preregistration, Anyway? (2016, August 18). Retrieved from https://www.psychologicalscience.org/publications/observer/obsonline/what-is-preregistration-anyway.html

Wray, N. & Visscher, P. (2008) Estimating trait heritability. *Nature Education* 1(1):29

Yadav, R. (2018, June). Freud and penis envy – a failure of courage? Retrieved from https://thepsychologist.bps.org.uk/volume-31/june-2018/freud-and-penis-envy-failure-courage