UNIVERSITY OF CALIFORNIA

Santa Barbara

Reconceptualizing Physical Sex as Continuous:
Are there Sex Differences in Video Game Preference?

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Arts
in Communication

by

Chelsea Anne Lonergan

Committee in charge:
Professor René Weber, Chair
Professor Daniel Linz
Professor W. James Potter

December 2016

The thesis of Chelsea Anne Lonergan is approved.

W. James Potter

Daniel Linz

René Weber, Committee Chair

October 2016

Acknowledgments

I would like to express my sincere gratitude to my advisor and committee chair, Professor René Weber for entertaining my pipe dreams aiming to dismantle embedded notions of sex and gender, and for pushing me past my breaking point in order to make my manuscript that much better. My thanks also go out to my committee members, Professors Dan Linz and Jim Potter, for helping contribute to what might be the most intriguing and enjoyable meetings to have ever existed.

To everyone that has listened to me argue against myself as I've developed my thoughts, thank you. This includes my endlessly intelligent and helpful lab members: Britney Craighead, Richard Huskey, Michael Mangus, Jacob Fisher, and Freddy Hopp. To everyone who has humored me without any academic background whatsoever (my parents and family), thank you. Finally, thank you to my friends who have kept me sane throughout an insane process. I could not have done this without any of you.

Abstract

Reconceptualizing Physical Sex as Continuous:
Are there Sex Differences in Video Game Preference?

By Chelsea Lonergan

With video games representing a rapidly growing media platform, current events have pushed conflicts surrounding gender into the spotlight of the video gaming world. Previous research has established sex-typed cognitive advantages as well as sex-typed genre preference of video games. Potential explanations of how these align have been inconclusive. Concurrently, research on sex development has suggested that binary categorizations of sex (physical sex; male or female) don't capture the full variation of individuals and as such have been recognized as inferior measures of sex. This thesis uses two continuous markers of physical sex (hormonal 2D:4D finger ratio, and continuous skill-based performance) to predict video game preference and playing. Using univariate effect and equivalence tests, we find that binary markers of sex predict game preference, while theoretically more valid continuous markers do not. Our results challenge the notion of binary sex differences in predicting video game preferences, and suggest a more complex relationship between physical sex and video game preferences. Implications for future research on video game playing and sex are discussed.

Table of Contents

Reconceptualizing Physical Sex as Continuous:
Are there Sex Differences in Video Game Preference?

Research on sex and gender has not only been a highly prolific area of study, but has been a relatively un-explicated one. Often, researchers study sex using a simple, binary distinction between male or female (later defined as 'physical sex'). Developing research, however, suggests that biological sex may be more continuous in nature, and therefore requires continuous markers that are more theoretically valid.

One area that has garnered interest in sex differences is video game research. Specifically, within the past couple of years, an Internet war centered on video game culture has exploded. Termed *Gamergate*, this convoluted battle that began as a fight for better journalism ethics has turned into a movement against the involvement of today's diverse range of players (Dewey, 2014). This movement has led to a small group of male gamers launching online threats towards prominent female figures in video game production, playing and culture. Sentiments against so-called "fake gamer girls" and the "fake games" they play escalated quickly. Milder forms of female-targeted online harassment turned into severe threats of violence or even death, ultimately forcing the FBI to get involved (Harper, 2015). With such serious consequences, it is evident that this cultural war bares scrutiny and requires a closer look at the relationship between sex and video game playing behaviors (preferences, playing, and perceptions).

The video game industry is colossal, with a total of $23.5 billion spent on gaming content, hardware and accessories in 2015 alone. Yet, despite the stereotype that gamers are overwhelmingly male, over 41% of video game players in America today are actually female (Entertainment Software Association, 2016). Simultaneously, video game research has grown with scholars studying everything from educational video game use to video game addiction. In particular, understanding what drives game preference has wide implications. For example, research uncovering preferences for various game genres, mechanics or narratives may help inform those who aim to produce educational games (e.g. spatial-skill training; Quaiser-Pohl, Geiser & Lehmann, 2006).

Findings demonstrating relationships between sex, cognitive skill, and video game preference are informative but are muddied by other findings which challenge underlying assumptions about sex development. Video game enjoyment is strongly connected to player skill (Grodal, 2000; Sherry, 2004), which has been linked to individual sex (Kimura, 2000). However, research about relevant skills' sex differences have been inconsistent for a number of years (Caplan, MacPherson, & Tobin, 1985; Voyer, Voyer, & Bryden, 1995). Such contradictory findings may be explained by theories surrounding the variability of sex differences, such as the greater male variability hypothesis (Baumeister, 2010). More fundamentally, research in the area of sex development has even begun to question what constitutes 'male' or 'female.' As we will discuss, the current evidence suggests that non-binary (or continuous) markers of sex might lead to more valid models of video game preference.

Using the initial argument behind *Gamergate* (that females don't prefer and play 'real' games), in this article we clarify the new concept of continuous biological sex as a reconceptualization of physical binary sex, theories surrounding its differences or lack thereof, and show how this new notion of sex predicts video game preference and playing behavior. We explore whether two continuous markers of physical sex—one hormonally driven and one derived from measures of cognitive performance—lead to different interpretations of sex differences in video game preference, which may have the potential to resolve current inconsistencies in the literature.

Theories of Sex Differences

Several prominent schools of thought have aimed to explain the existence (or lack) of sex differences. While the *Gamergate* perspective might echo the initially popular gender differences hypothesis, other theories have demonstrated often contradictory and inconsistent findings (Caplan, MacPherson & Tobin, 1985). In fact, the gender similarities hypothesis would directly contradict the *Gamergate* claim. The greater male variability hypothesis represents one avenue of research that has been relatively unexplored and which provides a clear explanation for why sex differences seem to fluctuate and rarely stay consistently large. This theory supports our suggestion that continuous markers of physical sex, which take variation into account, are thus more valid.

The Gender Differences Hypothesis

For decades, the gender differences hypothesis—which argues that males and females are psychologically and vastly different—has continued to be a popular idea (Hyde, 2005). In a concise review of cognitive sex differences, Miller and Halpern (2014) discuss the biological, psychological, and social aspects of male and female cognitive differences. Although variable at times, the male advantage in both spatial tasks and the mathematical and scientific fields of study that utilize them is among the more consistent developmental trends.

A potential explanation for this trend is the influence of hormones during development. Females who have been exposed to higher levels of prenatal androgens, whether that is due to having congenital adrenal hyperplasia (an overproduction of androgens) or having a fraternal male twin, have exhibited higher spatial performance, higher mental rotation performance, and other sex-differentiated traits such as sensation-seeking and tooth size (Miller & Halpern, 2014). Still, in a meta-analysis of sex differences across a variety of cognitive domains, Linn and Peterson (1985) found inconsistent sex differences which were larger only for specific tasks (e.g. Vandenberg & Kuse's 1978 test of 3D mental rotation) and specific age ranges (e.g. above 18). They cite hormonal changes during puberty as a potential explanation for these inconsistencies, though there is still no definitive reason.

Going further, research has shown that adolescents with higher levels of prenatal androgen exposure have more typically male patterns of development in certain areas of the cerebral cortex. To add further complication, some researchers have found that males and females can achieve equal cognitive performance while using the same brain regions slightly differently. With this, findings indicate that there is greater brain connectivity within each hemisphere for males, and greater connectivity between the hemispheres for females, with equal performance on surface-level behavioral measures (Fine, 2014). Though studies have not yet established direct linkages between brain development and cognition, some researchers suggest that hormonally-driven development exists to some degree. Unfortunately, the exact degree is still being questioned, with a number of hypotheses and theories offered to explain the inconsistencies.

The Gender Similarities Hypothesis

Despite empirical backing of the so-called gender differences hypothesis, the gender similarities hypothesis has gained traction (Hyde, 2005). This hypothesis claims that males and females are in fact more alike than they are different. In a meta-meta-analysis of psychological sex differences, Hyde (2005) found that a striking 30% of the effect sizes found are in the close-to-zero range, with an additional 48% in the small range. This evidence that 78% of sex differences are small or close to zero is indicative of a paradox within the area of sex research.

Namely, there is a contradictory approach to how researchers view the role of sex and gender. Rather than making theoretical claims for why there might be sex differences a priori, they instead treat the sexes as equally cognitively endowed while touting even the smallest of subsequently discovered differences (Canary & Hause. 1993). This method leads to a host of problems, including methodological inconsistency in attempting to understand a fundamental aspect of human nature.

The Greater Male Variability Hypothesis

The greater male variability hypothesis is one idea put forth to explain these contradictory findings. Though controversial due to its originally sexist assertions that resources spent on educating females are inherently wasted, there has been a small revival of this dated theory. This hypothesis was first proposed in the late 1800s to explain the overrepresentation of males in both the upper and lower tails of performance, power, and prestige (Ellis, 1894). For example, not only are more presidents and CEOs male, criminals and homeless people are also more likely to be male (Baumeister, 2010).

This hypothesis has been expanded since to explain the evolutionary mechanisms by which males must vary more to increase their chances of reproducing. As Baumeister (2010) argues, humanity has evolved from females who ‘play it safe’ and males who take risks, and either succeed (ending up in the upper tail) or fail (ending up in the lower tail). The greater male variability hypothesis, rather than used to deterministically explain why females are essentially inferior and doomed to mediocrity, could explain the relatively small effect sizes found, as distributions of behavioral outcomes (e.g. performance on a 3D mental rotation task) for both males and females mostly overlap with differences in variation.

Feingold (1992) analyzed this idea using nationally representative datasets to establish whether there was larger variability in male performance. He studied four differing tests of cognitive ability: the Differential Aptitude Test, the Preliminary Scholastic Aptitude Test/Scholastic Aptitude Test (PSAT/SAT), the Wechsler Adult Intelligence Scale/Wechsler Adult Intelligence Scale-Revised (WAIS/WAIS-R), and the California achievement test. Overall, Feingold (1992) found that males had consistently more variation than females in tests of quantitative ability, spatial visualization, spelling and general knowledge. This was coupled with significant differences in central tendency—the traditional method of assessing sex differences in this area of research. With this in mind, Feingold (1992) argues that both variance and averages must be assessed to form correct judgments about sex differences. Resting on such empirical support, greater male variability could be the underlying mechanism explaining a significant number of social aspects, perhaps even behavioral outcomes such as learned aggression (Archer & Mehdikhani, 2003).

Defining Physical Sex, Biological Sex, and Gender

To evaluate *Gamergate's* claims, we must first explore the concepts of sex and gender. As we navigate this conceptual warzone, we first distinguish between genotypes and phenotypes in explaining conceptual distinctions. Genotype refers to the set of genes an organism carries (grouped in chromosomes or long chains of genes), with one's entire genomic makeup termed their genome. Phenotype instead refers to all observable characteristics of an individual (Purvis et al., 2001).

Physical sex can be understood as phenotypical, anatomically-dictated binary distinctions into either male or female. For many, this distinction is chromosomal. When embryos inherit 23 pairs of chromosomes, one pair contains the “sex chromosomes,” either XX for female or XY for male. The Y chromosome houses the gene for ‘maleness’, which when set off, begins a cascade of hormones that allow the fetus to develop male sex organs. When this gene is not turned on, the fetus develops female sex organs. In the case that it is turned on, prenatal androgens (mostly testosterone) are secreted (Halpern, 2013). Often, physical sex is determined by considering one’s genitals and physical features like body composition and facial structure.

Biological sex encompasses all sex-relevant baseline genotypical variation which predicts phenotypical variation. As Ainsworth (2015) explains, “almost everyone is, to varying degrees, a patchwork of genetically distinct cells, some with a sex that might not match that of the rest of their body” (288). In utero, sex development in a fetus is a highly complex process, and depending on how it plays out, one’s genotype (biological sex) may predict more ambiguous sex phenotypes.

It is evident that binary notions of physical sex, then, may not accurately represent variable (continuous) biological sex. For example, an individual who lacks testosterone receptors—a crucial step for a fetus to develop fully into a male—may initially begin to develop as male, but the lack of receptors causes the body to form traditionally female genitalia, even though the individual has XY chromosomes. Individuals in this situation would be neither fully male nor female, and instead would be biologically “all over the place” (Dreger, 2010). These so-called disorders of sex development (DSD) range from extreme variations (e.g. possessing ambiguous genitalia) to much more subtle ones (e.g.

possessing male genitals but having low sperm production caused by a variation in sex-development genes). Current research has suggested that as many as 1 in 100 people have some form of a DSD (Ainsworth, 2015).

Separately, gender is the phenotypical expression, predicted somewhat by genotype, of one's sex through their behavior. This has been used to refer to the social aspects of what it means to be a man or a woman. This includes stereotypical appearances (e.g. women have long hair, feminine features, wider hips, etc.) and social roles (e.g. women are nurturing caretakers, emotional romantic partners, etc.). An individual's level of masculinity and femininity exists on a spectrum, dependent on the magnitude of their stereotypical characteristics.

Because gender and sex are often intertwined, it is important we make it clear that going forward, when we refer to sex we will refer strictly to biologically based distinctions and not social ones. Still, the vast majority of research looking at gender or sex differences does not actually define sex or gender. Instead, researchers often include one item asking if participants are either male or female. Canary and Hause (1993) criticize this exact issue, arguing that because measuring either physical sex or gender using one dichotomous code is so simple, "many researchers probably include sex as an independent variable because it is so easy to do and it presents another legitimate way to probe for significant differences" (p. 137).

These blurred biological distinctions between physical males and females complicates *Gamergate* arguments. Taking into account previously discussed research, it is evident that continuous notions of physical sex cannot be accurately measured with binary measures of physical sex. We therefore propose two continuous markers of biological sex in our study; both are genotypically derived.

Continuous Conceptualizations of Sex

Taking into account current findings, a *Gamergate* proponent would argue that all sex-differentiated characteristics would predict game preference. This includes binary physical sex, as well as continuous genotypically-derived hormonal and cognitive markers. We challenge this gendered assumption using several continuous markers of biological sex.

Continuous Hormonal Marker

Sex development relies heavily on hormonal triggers in order to shape one's genes and physical attributes such as brain development and sex organs. These hormonal triggers, such as the release of testosterone in utero, can fluctuate based on a variety of complex factors (e.g. lack of testosterone receptors), and therefore exist on a continuum within the body. Researchers have utilized a number of methods to measure hormone levels relevant to sex development. Finger ratio represents one method which is valid, non-invasive and simple to execute.

The ratio between the second and fourth digits (2D:4D) "has been suggested as a sexually dimorphic biomarker for prenatal androgen exposure" (Tapp, Mayberry & Whitehouse, 2011, p. 719). Essentially, this is a continuous measurement of testosterone level, and by extension, the fluctuating hormonal dimension of the genome which predicts physical sex. 2D:4D ratios are sexually dimorphic in utero, and are stable such that "the dimorphism in infancy is strongly correlated with that in adults…and there is a strong longitudinal correlation between early and later 2D:4D" (Manning, Churchill, & Peters, 2007, p. 224). Although research has implicated 2D:4D ratio as a "sexually dimorphic" biomarker (meaning physical males and females have differing patterns of finger length), this study will focus on the continuous nature of the ratio itself.

Research has used the 2D:4D ratio to predict performance in sex-specific cognitive tasks. For instance, regardless of being physically male or female, individuals with smaller digit ratios (longer ring fingers compared to index fingers) scored higher on a mental rotation task (Peters, Manning, & Reimers, 2007), as well as on a visuospatial task (Collær, Reimers & Manning, 2007). This supports the notion of a complex underlying genome (biological sex) which predicts physical sex, hormonal markers and cognitive markers.

Continuous Skill-Based Performance (CSBP)

In line with the greater male variability hypothesis, research has shown consistent increased variance for male performance on specific sex-differing tasks (Kimura, 2000). Still, performance for both physical males and females mostly overlaps. However, as Feingold (1992) and Hyde (2014) explain, the tendency to only evaluate mean differences in performance leaves out an important piece of information. Variability must also be included to gain a full understanding of cognitive sex differences. We therefore propose that a continuous conceptualization of performance will more validly predict game preference than traditional dichotomous measures. We call this continuous skill-based performance (CSBP).

CSBP is measured by taking individuals' performance on several game-relevant tests, comparing them to norm distributions of performance, and creating composite scores for each. In-depth explanations of the tests and how CSBP scores are calculated will be later discussed.

Sex Differences and Video Games

From the perspective of a *Gamergate* activist, one would argue that physical sex predicts game preference. While the majority of young Americans play video games overall, males have been found to more readily label themselves as "gamers" based on frequency of playing (Duggan, 2015). Still, despite this commonality, research has indicated that there is a very small gender gap in the initial interest of video game playing (Selwyn, 2007). This can be seen in the high percentage (41%) of female players in the consumer market (Entertainment Software Association, 2016).

Game-Relevant Cognitive Sex Differences

These consumer differences in gameplay may be driven, at least in part, by differences in cognitive abilities (Sherry, 2004). On average, physical males tend to do slightly better than physical females on tasks which test mental rotation, navigation, and target-directed motor skills (Kimura, 2000). Many of the most popular, male-preferred, video games require exactly these sorts of skills (Carr, 2005; Sherry, 2004). For example, male-dominated action games often require the player to navigate a complex 3D environment (3D mental rotation, route navigation) while also killing opponents (targeting). Conversely, females outperform males in cognitive tasks measuring color memory, object location memory, and verbal fluency/memory. Female-preferred puzzle, card, and classic board games often test these exact skills (Sherry, 2004). Though associations exist between cognitive skill and game preference, few studies have directly tested sex-differing cognitive skill and game preference.

Cognitive Skill and Game Preference

Gamergate's assertions about individual game preferences and playing requires an understanding of the conceptual linkages behind preference. Tamborini and colleagues

(2010) discuss the details of defining enjoyment, arguing that enjoyment, appeal, liking, joy, and pleasure from gaming are terms often used interchangeably to indicate “a preference for and response to media exposure” (p. 759). Csikszentmihalyi’s famous flow theory can also help explain the common and highly enjoyable experience of being in the zone while engaging in certain activities (Nakamura & Csikszentmihalyi, 2002). The theory posits that flow is experienced when one perceives that the challenges involved in the task matches his or her skill, so that there is neither frustration (too much challenge) nor boredom (too little challenge).

Similarly, Sherry (2004) argues that individual skill predicts game preference. If an individual does not feel immersed in the experience due to a lack of skill (either because they are frustrated or bored), then they will not enjoy the game (Tamborini et al., 2010). Trepte and Reinecke (2011) found that game-related efficacy is a mediator between player performance and game enjoyment. Namely, they found that players who performed better in a game, perceived more efficacy and more enjoyment. Additionally, Klimmt, Hartmann, and Frey (2007) conducted a study looking at predictors of game enjoyment, finding that individuals’ perceptions of control and effectance (perceived influence on the game world) determined whether a game was perceived as enjoyable or not. All these studies suggest that at the minimum, individual skill predicts levels of competence, efficacy, control, and effectance within that game, leading to experiences of flow and enjoyment.

‘Fake’ Vs. ‘Real’ Games

The bulk of *Gamergate’s* assertions lay behind what they term ‘fake’ as opposed to ‘real’ games. Unfortunately for researchers, there is no definitive list of what *Gamergate* means by this. In order to evaluate these sentiments, we must derive such a list from alternative empirical research. In a study on gamer identity and binary sex, Vermeulen and colleagues (2011) make a distinction between what they call “core” and “non-core” genres. They found that so-called core games are male-preferred and are defined as “skill-based games which are time-consuming and generally feature high-quality three-dimensional graphics” (p. 2). Such games also include high levels of violence, sexuality, and action (Lucas & Sherry, 2004). We argue that these are synonymous to what *Gamergate* proponents deem ‘real’ games, while those that play other types of games are not considered true gamers (Car, 2005). For example, females tend to prefer casual games such as puzzle and card games (Carr, 2005; Lucas & Sherry, 2004) which could be seen as ‘fake’ games. Vermeulen and colleagues’ (2011) list of core and non-core genres is tantamount to *Gamergate’s* as both are closely tied to gamer identity and genre authenticity. We therefore use “core” and “non-core” to mean ‘real’ and ‘fake,’ respectively.

Hypotheses

In summary, we challenge the notion that physical sex predicts video game preference and playing, as we introduce more valid markers of the genotypical components of sex. *Gamergate* activists would argue that binary and even ‘more valid’ continuous markers of physical sex all predict game preference and playing behavior. Specifically, they would argue that preference and playing of core or ‘real’ games is predicted by: 1) being physically male, 2) having lower 2D:4D ratios, and 3) demonstrating higher ability on male-advantaged cognitive tasks. Conversely, preference and playing of non-core or ‘fake’ games would be predicted by: 1) being physically female, 2) having higher 2D:4D ratios, and 3) demonstrating higher ability on female-advantaged cognitive tasks (and lower ability on male-advantaged ones). The present study tests these supposed relationships.

Method

Design

The methodology employed by this study was a survey combined with online cognitive skill tests. The between-subjects independent variables were physical sex and biological sex (ranging from highly male to highly female in both continuous conceptualizations). The main dependent variables were video game (genre) preferences, including preference measured via genre ranking and preference weighted by genre-specific playing time. Peripheral variables were also measured and included in the additional analyses section within results.

Participants

A total of 119 (50 females, 69 males) undergraduate students participated in the study to fulfill a research participation requirement for introductory Communication courses at a large university in the Western United States. Participants were recruited by convenience sampling through the SONA participant recruitment system and were required to be relatively frequent video game players. On average, participants had been playing video games for 10.94 years (SD = 4.41), with no participants indicating that they had been playing games for less than one year (3 people, or 2.5%, indicated they had played games for only one year).

Procedure

All proceedings took place online, through a private Qualtrics link that was sent to qualified individuals who signed up for the study through SONA. The researcher first presented the basic study procedures and demands in a short video, followed by online informed consent. The video was nearly a minute, and explained the need for participants to be ready with a printer, a camera, and an hour of time to spare before they continued onto the consent form. Those who chose not to consent after watching the video and reading the form were redirected out of the study.

Participants first completed questionnaires on individual demographics (e.g. age, ethnicity, year in school, etc.) as well as general video game playing experience (e.g. years, frequency and duration spent playing games). Participants were also asked about their overall preferred game genres as well as their current top three favorite games (and which of the given genres they belonged in). This was to constrain individuals as little as possible while also keeping with the manner of the Pew Research Center's study items (Lenhart et al., 2008). Participants then indicated how many hours on average they played each of their three preferred games in an average week.

A number of questionnaires regarding their most preferred game were then administered. As a check, they were asked to once again indicate the name of their top preferred game, to ensure they had a specific game in mind while completing the game-specific items. When assessing these items, 7 participants were excluded for not indicating a game for the game check item. Throughout the questionnaires, a variety of unrelated items were interspersed to ensure that participants were paying attention as they completed the items (e.g. "The sun rotates around the Earth" should be answered with "Strongly disagree"). Any participant who failed the attention checks was not included in the analyses. Five participants were excluded for failing all four attention checks. Participants who missed less than all four were still included in the analyses, in order to separate those who "never pay attention" while keeping with the idea that "removing all inattentive respondents may skew the sample and induce bias" (Berinsky, Margolis & Sances, 2012, p. 6).

Individuals first completed a questionnaire measuring enjoyment of their favorite game, followed by non-central items measuring subdimensions of perceived interactivity, and lastly by items measuring various uses and gratifications or motivations to play said favorite game. These items will be later discussed. Participants were then instructed to download and print out one sheet of given graph paper, to take photos of their hands against the paper, and to upload them into Qualtrics. Instructions on how to position their hands as well as photographic examples were given.

Lastly, participants were told they would complete a number of final tests. The order of these five tests were randomized within Qualtrics to account for possible effects of cognitive exhaustion. The five tests measured ability in a wide variety of game-relevant cognitive areas, including mental rotation, object location memory, navigation, targeting and disembedding.

Materials

A number of questionnaires (measuring enjoyment, interactivity, and uses & gratifications) were included in this study but were not central in our paper. These include a four-item scale measuring enjoyment and adapted from Skalski, Tamborini, Shelton, Buncher and Lindmark (2010); scales measuring six subdimensions of perceived interactivity (Customization & co-creation, controller responsiveness, feature-based interactivity, artificial intelligence, perceptual persuasivenesss, and exploration) with a total of 30 items (Weber, Behr, & DeMartino, 2014); and a 20-item scale measuring uses and gratifications of playing (Sherry, Lucas, Greenberg, & Lachlan, 2006). Results are discussed among the additional analyses section.

2D:4D Ratio: This hormonal marker of continuous biological sex is the first conceptualization put forth. 2D:4D ratio was measured using an adapted version of Manning and colleagues' (1998) self-report method. Their method asks participants to measure in millimeters the distance between the tip of the finger and the crease where the finger meets the palm.

In this study, ratios were calculated using individuals' uploaded photos. Only left hands were included in this study, as most individuals are right-handed, and might therefore experience increased difficulty when taking photos (with their left hand) of their right hand. Also, a meta-analysis by Putz and colleagues (2004) found that among studies measuring either hand or both hands, left ratios led to more significant correlations than right ratios. Current research has no established practice for measuring either left, right or both 2D:4D ratios.

Measuring finger ratios from these photos required several steps before analyses. Photoshop CS6's (Weinmann & Lourekas, 2012) unwarping tool was used in conjunction with the basic measuring tool to ensure that each photo was as un-skewed as possible. Overall, the first step was to measure each of the four corners of the graph paper (which was 5 mm) using the measuring tool, also in mm. The unwarping tool was used until the four corners were as close to each other as possible. Refer to Figure 1 for an illustration of the full measuring protocol.

Game Preference: Game preference was first measured using Pew Research Center's study item "What are your current top three favorite games?" (Lenhart et al., 2008). Participants then indicated which genre (of 16 provided) their top three favorite games belonged to. The game genres were adapted from Lucas and Sherry's (2004) list, to which

Multiplayer Online Battle Arena (MOBA) was added. This was due to the fact that participants who indicated “Other” as a genre almost exclusively named MOBA.

Next, game ranks were recoded for each genre. To calculate these ranks, top preferred rankings of 1 were recoded into 4, second preferred rankings of 2 were recoded into 3, third preferred rankings of 3 were recoded into 2, and any genre not nominated as a preferred genre was recoded into 1. Finally, ranks were added up for all nominated genre to calculate a rank sum.

For example, let us assume that an individual nominated fighting as their most preferred (#1) genre, sports as their second preferred (#2) genre, and fighting as their third preferred (#3) genre. This would initially give fighting both a number 1 (first preferred) and 3 (third preferred), while sports would get a number 2 (for second preferred). After recoding according to our procedure, fighting would instead get both a number 4 (for first preferred) and 2 (for third preferred), while sports would get a number 3 (second preferred). All other genres which were not nominated as the top three preferred would receive a number 1. Each genre would then have all numbers added up in order to calculate average rank sums across all genres. In this example, this would lead to the fighting genre having a total rank sum of 6 (4 from being first preferred plus 2 from being third preferred), the sports genre would receive a rank sum of 3 (for being second preferred), and all other genres would receive a rank sum of 1 (not nominated).

The chosen procedure allowed participants to hypothetically have all three favorite games be part of the same genre. This coding scheme was useful for a number of reasons. First, it allowed us to logically infer that higher numbers meant increased preference, rather than the other way around. Second, it allowed us to differentiate between different possible ranking orders. For example, if we recoded ranks into 3 (first preferred), 2 (second preferred), and 1 (third preferred), then a participant who indicated that their second and third preferred games were action games (2 plus 1, or 3 total) would have the same rank sum as someone who indicated only their first preferred game was an action game (3 total). Refer to Table 1 for a complete list of all possible rank sums for a given genre.

Weighted game preference was also calculated. This was done by multiplying genre rank sum by the amount of time participants reported playing the genre during an average playing session.

Cognitive Skill Tests: Five tests were presented in a randomized order. This included a redrawn version of the widely-used Vandenberg and Kuse's (1978) Mental Rotation Task (MRT; Peters et al., 1995), which assesses one's ability to discern between different rotated objects. In it, participants are shown a reference object followed by four alternative objects, and have to indicate which two match the first one, but are rotated. They have six minutes total to complete both parts of the MRT.

The second cognitive test measures individual object location memory, and is based on the well-known paradigm from Silverman and Eals (1992). In this test, three arrays of drawn everyday objects (e.g. a stool or a stapler) are presented to the participant. Participants have one minute to study the first array before being shown the second array, which has new objects added. They then have one minute to indicate which objects were not in the original array. Finally, they are shown a third array which has the original objects, but in different locations. They have unlimited time to indicate which objects changed location and which did not.

The third test, the Money Road Map Test (RMT; Money, Alexander, & Walker, 1965) measures route navigation. In this test, participants are shown a rudimentary path drawn through a city and are asked to indicate, at each turn, whether they would turn right or left. They have one minute to get through as many of the 32 turns as possible.

The fourth test measures field independence, or disembedding ability, and is adapted from the widely-used Group Embedded Figures Test (GEFT; Witkin, 1971). This test includes twelve items, with three minutes to complete them. For each item, participants are shown a complex figure and asked to choose which of seven simple figures are embedded within the complex one. There is only one correct answer for each item, and drawings are proportional, as explained in the initial instructions.

The final test measures targeting ability, and is derived from an online task demonstrating Fitt's Law (Psytoolkit.org; Stoet, 2010). In this test, participants complete 20 trials where they move their mouse between two boxes. They start by clicking on an origin box in the top left corner of the screen. Then, a second box of randomly-generated size and distance from the origin appears on-screen. Reaction times (RTs) are recorded to measure how long it takes participants to move their cursor from the origin box and click on the second box. Targeting skill is measured using RTs—standardized scores are calculated by subtracting actual RTs from the predicted RTs using Fitt's Law. Lower scores indicated faster times and thus higher ability.

We calculated CSBP scores using these test's 'norm' data, taken from previous studies. It is crucial to note that norm data did not exist in the traditional sense. Therefore, distributions were hypothesized from reports of means and standard deviations from either the creators of the tasks, or those who implemented them, such as Burte (2014) who adapted and implemented Money and colleague's (1965) test in the same manner as we did. In some cases, authors did not report whether the distributions were normal or not, and therefore it was assumed for our purposes. Although this is not an ideal methodology, it represents an important first step in compiling future norms and a way to test the calculation procedures. In every instance, it was attempted that norms be taken from those who implemented the test in as similar a manner as possible, and that had as big of a sample as possible.

To calculate one's CSBP score, z-scores were calculated from each individual's raw score, for both the male distribution and the female distribution. Refer to Figure 2 for an example graph of these norms. Once z-scores were calculated for the male and female distributions, area under the curve (AUC) was calculated for each score for both male and female distributions. Finally, a ratio of the male AUC divided by the female AUC was calculated. This was done for each of the CSBP scores so that for all tasks, the higher an AUC ratio was, the more 'masculine' that individual was (regardless of physical sex). For an in-depth example of this calculation method, refer to Figure 3.

Results

Gaming Behavior

On average, participants indicated that they had played on 14.18 days within the previous month (*SD* = 7.95), and on 3.34 days within the previous week (*SD* = 1.86). Overall, participants reported playing for an average of 137.79 minutes per playing session (*SD* = 102.55). This translates to an average of about 2 hours and 18 minutes per session, with a standard deviation of about an hour and 43 minutes. All distributions of playing time and frequency were normally distributed, except for distribution of total playtime per session which was skewed slightly right.

Participants ranked gaming devices (with 1 being most often used), resulting in the PlayStation 4 as the most frequently used device ($M_{\text{rank}} = 2.11$, $SD = 2.09$), followed by a smartphone ($M_{\text{rank}} = 2.34$, $SD = 1.26$), then by a computer ($M_{\text{rank}} = 2.38$, $SD = 1.66$). Those ranked as the least frequently used devices included the PlayStation ($M_{\text{rank}} = 9.75$, $SD = 2.75$) and Xbox ($M_{\text{rank}} = 6.50$, $SD = 4.43$) consoles. Among all 15 named game genres presented, most participants indicated they played racing (55%), action (55%), first-person shooters (51.7%), sports (51.7%), fighting (48.3%), adventure (47.5%) and puzzle (40%) games. The least popular genres were Massively Multiplayer Online Games (MMOG; 15.8%), survival horror (14.2%), virtual worlds (10%), and finally Multiplayer Online Battle Arena (2.5%) games. 0.8% of participants named Other genres.

Self-reported video game skill data was also gathered on a Likert scale from 1 (Low Skill) to 7 (High Skill). Participants' self-reported skill was normally distributed, with an average skill level of 4.80 ($SD = 1.19$). An independent samples t-test was conducted to compare self-reported skill between males and females. There was a significant difference in scores, showing that males ($M = 5.29$, $SD = 0.96$) report higher skill than females ($M = 4.12$, $SD = 1.15$); $t(117) = -6.03$, $p < .001$ (refer to Figure 4).

2D:4D Ratios

To assess inter and intra-reliability of photo measuring, 10 of the 87 submitted photos (11.5%) were randomly chosen and re-measured by the initial coder as well as the researcher. All correlations were very highly significant ($r > 0.90$). Inter-coder reliability was established for 2D, 4D, and 2D:4D, $r = .997$, $p <.001$. Intra-coder reliability was also established, $r = .998$, $p < .001$.

The average difference between all four graph paper corners ranged between 0 and 1.90 mm (*M* = 0.50, *SD* = 0.43). An independent samples t-test revealed that 2D and 4D lengths differed significantly between males and females. Males had longer 2Ds (*M* = 78.98, *SD* = 10.05) and 4Ds (*M* = 81.76, *SD* = 9.72) than females (M_{2D} = 67.10, SD_{2D} = 7.46; M_{4D} = 69.58, SD_{4D} = 8.55). Overall, the average Left 2D:4D ratio was 0.96 (*SD* = 0.05). An independent samples t-test was conducted to compare 2D:4D ratios between males and females. There was no significant difference in ratios for males (*M* = 0.96, *SD* = 0.04) and females (*M* = 0.97, *SD* = 0.06); *t*(85) = .441, *p* = .159 (refer to Figure 5).

Relationship Between Physical Sex and Game Preferences

A number of correlations were conducted in order to assess the relationship between physical sex and rank sum. To do this, physical sex was dummy coded with females as 0 and males as 1. Thus, positive correlations were associated with male preference while negative correlations were associated with female preference. Overall, average core genre rank sums were significantly positively correlated with physical sex, *r* = .351, *p* < .001, while non-core genre rank sums were significantly negatively correlated with physical sex, *r* = -.406, *p* < .001. Males (dummy coded as 1) were more likely to have higher rank sums for core games while females were more likely to have higher rank sum for non-core games. Specifically, significant positive correlations also existed between physical sex and sports rank sums (*r* = .426, *p* < .001). Significant negative correlations were found between physical sex and racing rank sums (*r* = -.298, *p* < .001), puzzle rank sums (*r* = -.325, *p* < .001), and simulation rank sums (*r* = -.230, *p* < .05).

A number of correlations were also conducted in order to assess the relationship between physical sex and weighted preference (by playing time). Overall, weighted core genre preference was significantly positively correlated with physical sex, $r = .265$, $p < .01$, while weighted non-core genre preference was significantly negatively correlated with physical sex, $r = -.246$, $p < .01$. Males were more likely to have higher preference (weighted by playing time) for core genres, while females were more likely to have higher preference (weighted by playing time) for non-core genres. For specific weighted genre preferences, we found significant positive correlations between physical sex and sports weighted preference ($r = .276$, $p < .01$), and first-person shooter weighted preference ($r = .188$, $p < .05$). Finally, there was a significant negative correlation between physical sex and puzzle weighted preference ($r = -.226$, $p < .05$). Generally speaking, the first *Gamergate*-sanctioned hypothesis was supported.

Relationship between 2D:4D and Game Preferences

Another series of correlations were conducted to assess the relationship between a marker of biological sex (2D:4D) and game preference. Overall, there was no significant correlation between 2D:4D ratio and core genre rank sum ($r = -.070$, $p = .522$), or between 2D:4D ratio and non-core genre rank sum. In fact, in a test of bivariate correlation equivalence (Weber & Popova, 2012) assuming a small to medium effect size, 2D:4D ratio was significantly equivalent for core rank sum at $\Delta = 0.24$ ($p = .048$), and for non-core rank sum at $\Delta = 0.19$ ($p = .046$). In all tests of bivariate equivalence, effect sizes were drawn from both previously-established average effect sizes within media effects research in communication (Weber & Popova, 2012) as well as from effect sizes within sex differences research (Hyde, 2005).

Correlations were also conducted to assess the relationship between 2D:4D ratio and weighted preference by playing time. There was no significant relationship for either core genres ($r = .037$, $p = .732$) or non-core genres ($r = .020$, $p = .856$). A test of bivariate correlation equivalence, again assuming a small to medium effect size, revealed 2D:4D ratio was significantly equivalent for core weighted preference at $\Delta = 0.21$ ($p = .048$), and for non-core weighted preference at $\Delta = 0.20$ ($p = .043$). Overall, the continuous marker of biological sex did not predict game preferences or playing., which does not support the *Gamergate*-sanctioned hypothesis.

Relationship Between CSBP Scores and Game Preferences

Table 2 outlines all five cognitive tests, the corresponding norm data which scores were compared to, and overall score results for physical males and females. It is important to note that the targeting test does not have norm data because it was not based on any previously-implemented test. This is reflected in the table, though targeting results are still shown.

A series of correlations were conducted between core and non-core rank sums, the AUC ratios for four tasks and the targeting RTs. Core rank sum was not significantly related to any CSBP. Non-core rank sum was significantly correlated with MRT AUC ratio, $r= -.185$, $p < .05$, indicating that lower non-core preference was associated with higher MRT maleness (the probability that an individual is male increases with the AUC ratio). Additionally, non-core rank sum was significantly related to targeting score, $r = -.195$, $p < .05$, indicating that lower non-core preference was associated with higher targeting skills (slower, worse scores).

Correlations between core and non-core weighted preference by playing time, AUC ratios, and targeting RTs were also conducted. Core weighted preference was significantly positively correlated with MRT AUC ratio, $r = .313$, $p < .01$, as well as significantly positively related to GEFT AUC ratio, $r = .208$, $p < .05$. This signifies that increased core genre preference (weighted by playing time) is related to both increased MRT maleness and GEFT maleness. Non-core weighted preference was significantly negatively correlated with targeting score, $r = -.197$, $p < .05$, indicating that decreased non-core preference is related to higher targeting skill.

Additional Analyses

Differences in motivations to play games and perceptions of playing were also investigated. Correlations were conducted to assess the relationship between physical sex, self-reported video game skill, enjoyment of their favorite game, perceptions of interactivity within their favorite game, and uses and gratifications behind their favorite game. There was a significant correlation between physical sex and self-reported skill, $r = .487$, $p < .001$, with increased maleness associated with increased self-reported skill. Enjoyment was marginally significantly related to physical sex, $r = .190$, $p = .068$, with more maleness associated with more enjoyment. Additionally, increased self-reported skill was significantly positively related to increased enjoyment, $r = .285$, $p < .01$.

Correlations revealed some significant relationships between the six sub-dimensions of interactivity and CSBP. Overall, binary physical sex did not significantly correlate with any subdimensions of interactivity. 2D:4D also did not significantly correlate with interactivity subdimensions. However, increased exploration interactivity was associated with decreased GEFT AUC ratio, $r = -.267$, $p < .001$, as well as with decreased targeting ability, $r = .252$, $p < .001$. Customization/co-creation perception was also associated with decreased GEFT AUC ratio, $r = -.189$, $p < .05$. Artificial intelligence was also associated with decreased targeting ability, $r = .234$, $p < .05$. Finally, Controller responsiveness was significantly correlated with decreased object memory AUC ratio, $r = -.193$, $p < .05$.

Uses and Gratifications motivations were also related to several variables. Physical sex was significantly positively correlated with the drive for competition ($r = .190$, $p < .05$), challenge ($r = .216$, $p < .05$), social interaction ($r = .255$, $p < .01$), and arousal ($r = .355$, $p < .001$). 2D:4D was not significantly correlated with any uses and gratifications. Still, competition motivation was significantly positively correlated with location memory AUC ratio, $r =.213$, $p < .05$, while RMT AUC ratio was significantly correlated with diversion, $r = .217$, $p < .05$.

Discussion

We tested the *Gamergate* perspective that core preference could be predicted by being male, having lower 2D:4D ratios, and by performing well on male-advantaged tests, while non-core preference could be predicted by being female, having higher 2D:4D ratios, and by performing well on female-advantaged tests. We found that physical sex differences in core and non-core preferences and playing times was consistent with *Gamergate* presumptions. However, we did not find these relationships using biological sex. 2D:4D ratio, a hormonal marker of biological sex, was not related to gaming behavior as bivariate correlation equivalence tests revealed.

There were some, albeit sometimes contradictory, findings regarding CSBP scores. While core rank sum was not at all related to CSBP scores, higher non-core rank sum was predicted by lower MRT performance (consistent with Lucas & Sherry, 2004), as well as lower targeting performance (inconsistent with previous literature). For playing time weighted preference, increased core genre playing was predicted by increased MRT performance, and increased GEFT performance (consistent with literature on game-relevant mental rotation and disembedding skills). Conversely, increased non-core playing was predicted by increased targeting school (inconsistent with previous literature). Here, the strongest evidence for the *Gamergate* perspective is really only mental rotation ability, and perhaps somewhat disembedding ability.

Thus, we conclude that *Gamergate's* central assertions were not supported by our findings. Continuous—and according to the current literature more valid—measures of physical sex (as a proxy for continuous biological sex) did not consistently predict video game preference and playing. These findings have major implications for not only the future of the gaming industry, but for the theoretical basis of movements like *Gamergate* as well. Whereas *Gamergate* argues that one's physical sex influences their choice of 'real' versus 'fake' games, we suggest that variance within these groups might in fact eradicate these results. With this, we can only speculate how many so-called sex or gender differences are couched in reductive binary logic. Such findings might not only defeat the logic underlying *Gamergate's* assumptions, but they could change the way researchers think about and study one of the most fundamental—and surprisingly misunderstood—aspects of human nature.

Additional analyses looking at perceptions of video game experience and motivations reveal a number of further intriguing research avenues. The finding that females view themselves as less skilled in gaming scenarios than males do is consistent with previous work (Selwyn, 2007) and is important to note in a study where game-relevant cognitive tasks followed this self-report item. It was unexpected that physical males and females would report differing levels of enjoyment for their favorite game. This measure was included as a check, with the expectation that all individuals, by virtue of naming a 'favorite' game, would essentially demonstrate a ceiling effect. This is possibly tied to more social, gendered effects of video game culture, with females perceiving themselves not only as less skilled but as enjoying a stereotypically-male activity less (Selwyn, 2007).

The specific finding that physical sex predicts certain motivations to play may have implications for which general game mechanics increase enjoyment and preference. For example, males tended to cite competition, challenge, social interaction and arousal as motivations to play significantly more than women. These findings may help to increase in-game flow by allowing producers and creators to create games which capitalize on these motivations. These findings might inform those who aim to produce more enjoyable educational games (e.g. spatial-skill training; Quaiser-Pohl, Geiser & Lehmann, 2006).

While we believe this study sheds light on a complex and politically-charged issue, with inconsistencies in the literature, we would like to make our audience aware of its limitations. First, data from a subject pool of undergraduate students is, admittedly, limited in the populations to which it can be generalized. The average university student is Caucasian, young, and comes from a higher SES background. These factors may have played a significant role in their prior experience with certain video games and tasks, as well as individual preferences. However, distributions of cognitive task performance have been shown to be relatively consistent between groups. Further, 34.7% of this sample was White, 28.9% were Asian and 23.1% were Hispanic or Latino. This indicates a relatively diverse sample, especially when comparing it to mostly White university populations.

A second limitation of this study surrounds its use of 2D:4D ratio. Although this attribute has been gaining popularity as a cheap, easy way to get a non-invasive glimpse inside one's hormonal makeup, it is crude at best. Not only is there no established method for actually measuring 2D:4D ratio, there are some inconsistencies regarding which hand should be used, and how precisely it actually measures the underlying bone. For example, a common protocol (and the one we used) directs the researcher to measure from the center of the fingertip, to the middle of the bottom-most knuckle fold (theoretically, where the finger meets the palm). Looking at even just a few hands, you will notice that people's hands greatly vary, and that the bottommost knuckle fold is oftentimes not very close to the palm. In those cases, we followed the protocol, but could not help but wonder how accurately such a marker was measuring the bone length underneath. Still, in our measurements we conducted several reliability tests and assured the highest possible reliability of the 2D:4D measure. These tests included manual un-warping of photos by repeatedly measuring, adjusting, and measuring again the grid squares until differences were minimized. Additionally, inter- and intra-coder reliability was attained by randomly re-measuring 10 of the 87 photos. Reliabilities were all above $r = 0.90$. Thus, the lack of substantial correlations between 2D:4D as a marker of biological sex and game preferences cannot be explained with the lack of the measure's reliability and potentially increased error variance as a consequence of it. In fact, we employed an equivalence test procedure according to the protocol in Weber and Popova (2012) and established significant equivalence for the relationship between biological sex and game preferences. It should be noted that for this equivalence test we assumed a small to medium effect size between sex and game preferences according to the recommendations in Weber and Popova (2012). This is a conservative equivalence test strategy. Had we assumed a strong effect size, demonstrating equivalence would have been easier (i.e. would have led to even lower p values).

A third, important limitation of this study is the dire lack of accessible norm data. As we stated, we were forced to extrapolate norms from other published research, where distributions were not normally shown. In these cases, we were made to hypothesize from only means and standard deviations, assuming that the distributions were normal. Still, this study represents the first step in the right direction. Perhaps others will use this research as reason to compile norms, or to release norms which they lock away for profit. This limitation of our study is directly a consequence of non-cumulative research efforts and proprietary, non-academic research activities in this area. Given the importance of this research, we can only hope that this reality will change in the future.

References

Ainsworth, C. (2015). Sex redefined. Nature, 518(7539), 288-291. doi: 10.1038/518288a.

Archer, J., & Mehdikhani, M. (2003). Variability among males in sexually selected attributes. Review of General Psychology, 7(3), 219.

Baumeister, R. F. (2010). Is there anything good about men?: how cultures flourish by exploiting men. Oxford University Press.

BBC News (2013) "What Is the Appeal of Candy Crush Saga?" BBC News. BBC (British Broadcasting Network), 17 Dec. 2013. Web. 4 Oct. 2015.

Berinsky, A. J., Margolis, M., & Sances, M. W. (2012, March). Separating the shirkers from the workers? Making sure respondents pay attention on internet surveys. In NYU CESS 5th Annual Experimental Political Science Conference.

Burte, H. (2014). Individual and strategy differences in allocentric-heading recall performance and its neural basis (Dissertation). University of California, Santa Barbara.

Canary, D. J., & Hause, K. S. (1993). Is there any reason to research sex differences in communication?. Communication Quarterly, 41(2), 129-144.

Caplan, P. J., MacPherson, G. M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. American Psychologist, 40(7), 786.

Carr, D. (2005). Contexts, gaming pleasures, and gendered preferences. Simulation & Gaming, 36(4), 464-482.

Collaer, M. L. (2001). Judgment of Line Angle and Position Test-15 line version (JLAP-15). Unpublished test. Middlebury, VT: Middlebury College.

Collaer, M. L., Reimers, S., & Manning, J. T. (2007). Visuospatial performance on an internet line judgment task and potential hormonal markers: sex, sexual orientation, and 2D: 4D. Archives of Sexual Behavior, 36(2), 177-192.

Dewey, C. (2014). The only guide to Gamergate you will ever need to read. Washington Post.

Dreger, A. (2010). Is anatomy destiny? TED Talks (Video). Filmed December.

Duggan, M. (2015, December 15). Gaming and Gamers. Retrieved November 02, 2016, from http://www.pewinternet.org/2015/12/15/gaming-and-gamers/

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Kit of factor-referenced cognitive tests. Educational Testing Service (ETS).

Ellis, H. (1894). Man and woman. London: Walter Scott.

Entertainment Software Association. (2016). Essential facts about the computer and video game industry. Retrieved from http://essentialfacts.theesa.com/Essential-Facts-2016.pdf

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. Review of Educational Research, 62(1), 61-84.

Fine, C. (2014). His brain, her brain?. Science, 346(6212), 915-916.

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. Personality and Individual Differences, 14(1), 53-65.

Grodal, T. (2000). Video games and the pleasures of control. Media entertainment: The psychology of its appeal, 197-213.

Hall, J. A. Y., & Kimura, D. (1995). Sexual orientation and performance on sexually dimorphic motor tasks. Archives of Sexual Behavior, 24(4), 395-407.

Halpern, D. F. (2013). Sex differences in cognitive abilities. Psychology press.

Harper, S. (2015, April 24). GamerGate, 'Fake Gamer Girls' and Princess Elsa: How Important Are Female Gamers? Retrieved September 25, 2016, from http://www.dojit.com/gamergate-fake-gamer-girls-and-princess-elsa-how-important-are-female-gamers/

Hundhammer, T., & Mussweiler, T. (2012). How sex puts you in gendered shoes: sexuality-priming leads to gender-based self-perception and behavior. Journal of personality and social psychology, 103(1), 176.

Hyde, J. S. (2005). The gender similarities hypothesis. American psychologist, 60(6), 581.

Hyde, J. S. (2014). Gender similarities and differences. Annual review of psychology, 65, 373-398.

Kepner, M. D., & Neimark, E. D. (1984). Test–retest reliability and differential patterns of score change on the Group Embedded Figures Test. Journal of Personality and Social Psychology, 46(6), 1405.

Kimura, D. (2000). Sex and cognition. MIT press.

Klimmt, C., Hartmann, T., & Frey, A. (2007). Effectance and control as determinants of video game enjoyment. Cyberpsychology & Behavior, 10(6), 845-848.

Lenhart, A., Kahne, J., Middaugh, E., Macgill, A. R., Evans, C., & Vitak, J. (2008). Teens, Video Games, and Civics: Teens' Gaming Experiences Are Diverse and Include Significant Social Interaction and Civic Engagement. Pew Internet & American Life Project.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child development, 1479-1498.

Lucas, K., & Sherry, J. L. (2004). Sex differences in video game play: A communication based explanation. Communication Research, 31(5), 499-523.

Manning, J. T., Churchill, A. J., & Peters, M. (2007). The effects of sex, ethnicity, and sexual orientation on self-measured digit ratio (2D: 4D). Archives of sexual behavior, 36(2), 223-233.

Manning, J. T., Scutt, D., Wilson, J., & Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. Human Reproduction, 13(11), 3000-3004.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. Trends in Cognitive Sciences, 18(1), 37-45.

Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. Handbook of Positive Psychology, 89-105.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test-different versions and factors that affect performance. Brain and cognition, 28(1), 39-58.

Peters, M., Mackenzie, K., & Bryden, P. (2002). Finger length and distal finger extent patterns in humans. American Journal of Physical Anthropology,117(3), 209-217.

Peters, M., Manning, J. T., & Reimers, S. (2007). The effects of sex, sexual orientation, and digit ratio (2D: 4D) on mental rotation performance. Archives of Sexual Behavior, 36(2), 251-260.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A. S., McNamara, J. O., & Williams, S. M. (2001). What is Sex? In Neuroscience. 2nd edition. Sunderland

(MA): Sinauer Associates. Available from: https://www.ncbi.nlm.nih.gov/books/NBK10943/

Putz, D. A., Gaulin, S. J., Sporter, R. J., & McBurney, D. H. (2004). Sex hormones and finger length: What does 2D: 4D indicate?. Evolution and Human Behavior, 25(3), 182-199.

Quaiser-Pohl, C., Geiser, C., & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental-rotation ability. Personality and Individual differences, 40(3), 609-619.

Selwyn, N. (2007). Hi-tech= guy-tech?: An exploration of undergraduate students' gendered perceptions of information and communication technologies. Sex Roles, 56(7-8), 525-536.

Sherry, J. (2004). Flow and media enjoyment. Communication Theory, 14(4), 328–347. doi:10.1111/j.1468-2885.2004.tb00318.x

Sherry, J., Rosaen, S., Bowman, N. D., & Huh, S. (2006). Cognitive skill predicts video game ability. In Paper presented at the annual conference of the International Communication Association, Dresden, Germany.

Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. Barkow, I. Cosmides, and J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture. New York: Oxford University Press.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. Behavior Research Methods, 42(4), 1096-1104.

Tamborini, R., Bowman, N. D., Eden, A., Grizzard, M., & Organ, A. (2010). Defining media enjoyment as the satisfaction of intrinsic needs. Journal of communication, 60(4), 758-777.

Tapp, A. L., Maybery, M. T., & Whitehouse, A. J. (2011). Evaluating the twin testosterone transfer hypothesis: a review of the empirical evidence. Hormones and Behavior, 60(5), 713-722.

Trepte, S., & Reinecke, L. (2011). The pleasures of success: Game-related efficacy experiences as a mediator between player performance and game enjoyment. Cyberpsychology, Behavior, and Social Networking, 14(9), 555-557.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. Perceptual and Motor Skills, 47(2), 599-604.

Vermeulen, L., Van Looy, J., De Grove, F., & Courtois, C. (2011). You are what you play?: a quantitative study into game design preferences across gender and their interaction with gaming habits. In DiGRA 2011: Think, design, play. Digital Games Research Association (DiGRA).

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychological Bulletin, 117(2), 250.

Watson, N. V., & Kimura, D. (1991). Nontrivial sex differences in throwing and intercepting: Relation to psychometrically-defined spatial functions. Personality and Individual Differences, 12(5), 375-385.

Weber, R., & Popova, L. (2012). Testing equivalence in communication research: Theory and application. *Communication Methods and Measures*, 6(3), 190–213. doi:10.1080/19312458.2012.703834

Weinmann, E., & Lourekas, P. (2012). Photoshop CS6: For Windows and Macintosh. Peachpit Press.

Witkin, H. A. (1971). Group embedded figures test.

Appendix A

Table **1**. Example rank sums for a game genre (X)

Initial Ranking (1st, 2nd, 3rd)	Recoded Ranking	Rank Sum
X, X, X	4, 3, 2	9
X, X, __	4, 3, __	7
X, __, X	4, __, 2	6
__, X, X	__, 3, 2	5
X, __, __	4, __, __	4
__, X, __	__, 3, __	3
__, __, X	__, __, 2	2
__, __, __	__, __, __	1

Note: Each X represents one hypothetical genre (for example, action) while each blank (__) represents any other genre (in this example, any genre besides action). Initial ranking is the raw rank by the participant (as first, second, or third); recoded ranking is according to our outlined procedure; rank sum represents the added up recoded rank score for the hypothetical genre (i.e. action). Notice that at the bottom, where the genre is not nominated by the hypothetical participant, it receives a rank sum of 1.

Appendix B

Table 2 Cognitive test norm data and results.

Cognitive Test	Norm Data	Results
MRT	From *Peter et al., 1995*	
Scoring	Max. score 24	
Males	(*N* = 237, *M* = 13.45, *SD* = 4.80)	(*N* = 69, *M* = 10.18, *SD* = 4.76)**
Females	(*N* = 399, *M* = 9.30, *SD* = 4.00)	(*N* = 50, *M* = 7.23, *SD* = 4.73)**
Object Memory	From *Silverman & Eals, 1992*	
Scoring	Min. score -27 and max. score 20	
Males	(*N* = 63, *M* = 12.25, *SD* = 4.27)	(*N* = 69, *M* = 13.74, *SD* = 5.77)
Females	(*N* = 115, *M* = 14.15, *SD* = 3.90)	(*N* = 50, *M* = 11.70, *SD* = 6.94)
Location Memory	From *Silverman & Eals, 1992*	
Scoring	Max. score 27	
Males	(*N* = 83, *M* = 18.45, *SD* = 3.58)	(*N* = 69, *M* = 17.52, *SD* = 6.41)
Females	(*N* = 134, *M* = 20.14, *SD* = 4.11)	(*N* = 50, *M* = 15.06, *SD* = 8.34)
RMT	From *Burte, 2014*	
Scoring	Max. score 32	
Males	(*N* = 44, *M* = 23.00, *SD* = 7.80)	(*N* = 67, *M* = 16.76, *SD* = 9.79)
Females	(*N* = 58, *M* = 17.60, *SD* = 8.40)	(*N* = 47, *M* = 13.68, *SD* = 7.11)
GEFT	From *Kepner & Neimark, 1984*	
Scoring	Percentage correct	
Males	(*N* = 55, *M* = 0.38, *SD* = 0.22)	(*N* = 69, *M* = 0.33, *SD* = 0.11)
Females	(*N* = 129, *M* = 0.33, *SD* = 0.20)	(*N* = 50, *M* = 0.31, *SD* = 0.12)
Targeting	—	
Scoring	Predicted minus actual RT	
Males	—	(*N* = 66, *M* = -2,207.85, *SD* = 2,062.67)
Females	—	(*N* = 47, *M* = -3,342.12, *SD* = 4,357.78)

**. T-test (between physical male and female scores) is significant at the 0.01 level (2-tailed).

Appendix C

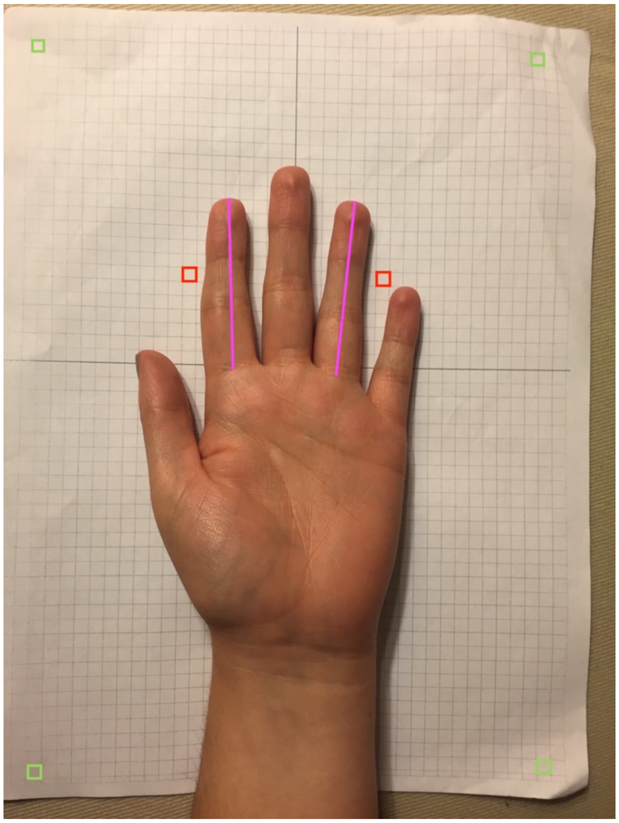

Figure 1: An example measuring protocol for an uploaded photo of a hand. In this example, the first step in measuring right 2D:4D would be to measure each of the four corners (one side of the box; highlighted by green boxes). Photoshop CS6's (Weinmann & Lourekas, 2012) unwarping tool would be used until the green boxes were as close (in mm) as possible. Then, two boxes on either side of the fingers (highlighted by red boxes) would be measured to act as a scaling tool for the finger lengths. Finally, the fingers (purple lines) would be measured using Manning and colleague's (1998) protocol. Because each box is 5 cubic millimeters, the finger ratios are calculated by first dividing 5 mm by the box's measurement, then dividing that ratio by the finger's measured distance. For example, to measure this example right 4D (ring finger), 5mm would be divided by the box's measured distance (18.3

mm, thus a ratio of .273), which would then be multiplied by the finger distance (241.4 mm) to find that the real-life finger length would be 65.87 mm..

Appendix D

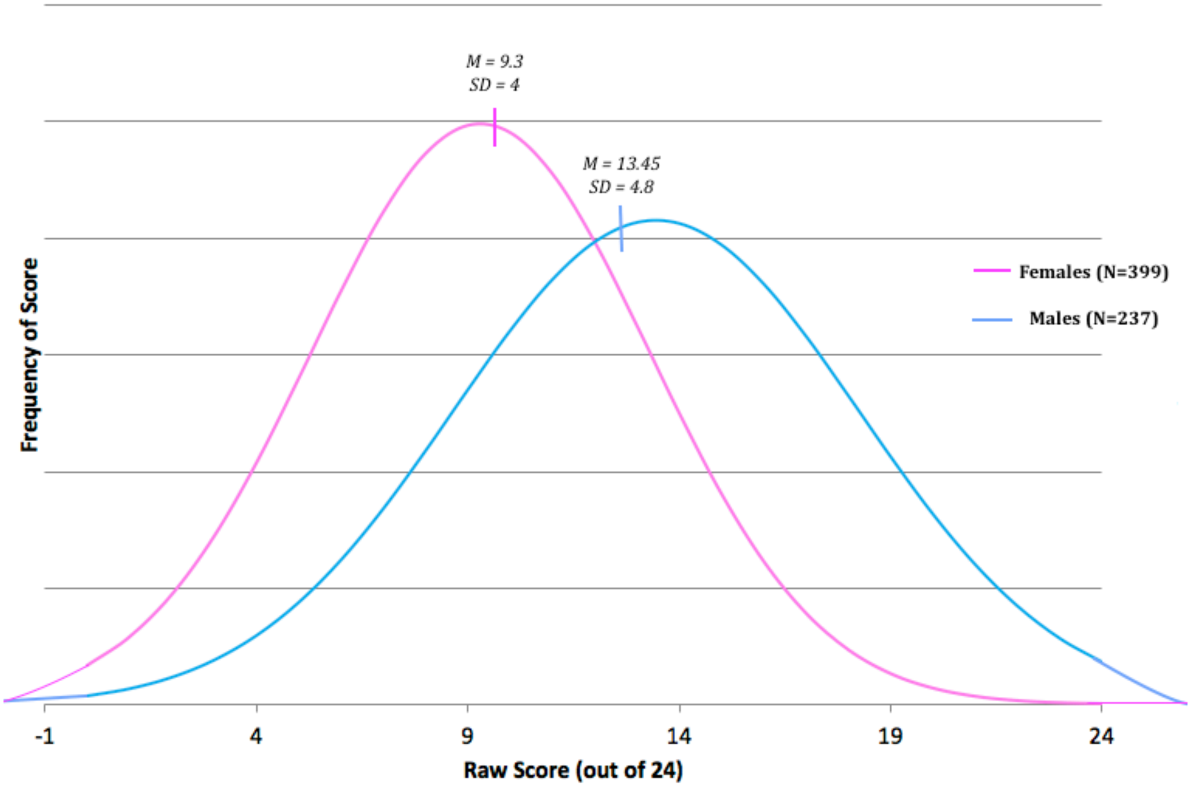

Figure 2: A hypothetical continuous distribution of performance on the redrawn 3D MRT from Peters and colleagues (1995). Physical males tend to have a slightly higher average (*M* = 13.45) performance than physical females (*M* = 9.3), with greater variability (*SD* = 4.8 for males; *SD* = 4 for females).

Appendix E

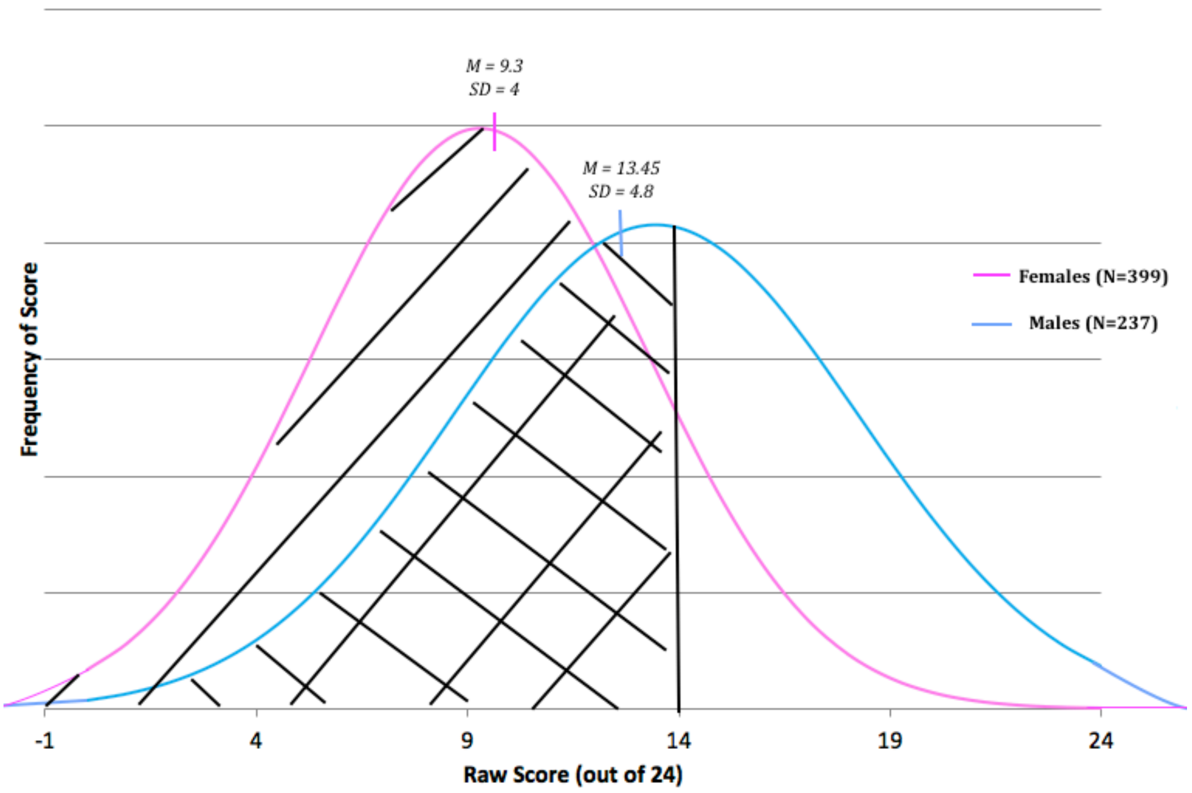

Figure 3: An example calculation procedure for MRT CSBP score. All scores which fell outside the 95% probability line, such as scores of 0, were recoded to be within the 95% interval. This was because the distribution of M/F AUC ratio, within that 95% interval, defined a monotonic increasing, almost linear function. The black line represents a hypothetical individual's score on the task, in this case 14 out of the maximum 24. Each CSBP score is found by first calculating the z-score for both the female distribution $\left(\frac{14-9.3}{4}\right)$ = 1.175 and the male distribution $\left(\frac{14-13.45}{4.8}\right)$ = 0.115. Areas under the curve (AUC) are then calculated for both z-scores (Female AUC = 0.88; Male AUC = 0.55. In this case, the male AUC (indicated by uni-directional diagonal lines underneath the left side of the male distribution) is divided by the female AUC (indicated by the bi-directional diagonal lines

underneath the left side of the female distribution), to get $\frac{0.115}{1.175} = 0.098$. This CSBP score indicates that with a given raw score of 14, the AUC for the male distribution is 0.098 times as large as the AUC for the female distribution. Thus, the bigger this SBP score, the more 'masculine' an individual is (meaning their score falls more in line with males' performance). This makes sense as 14 is a rather high score out of 24, and is above the male average of 13.45.

Appendix F

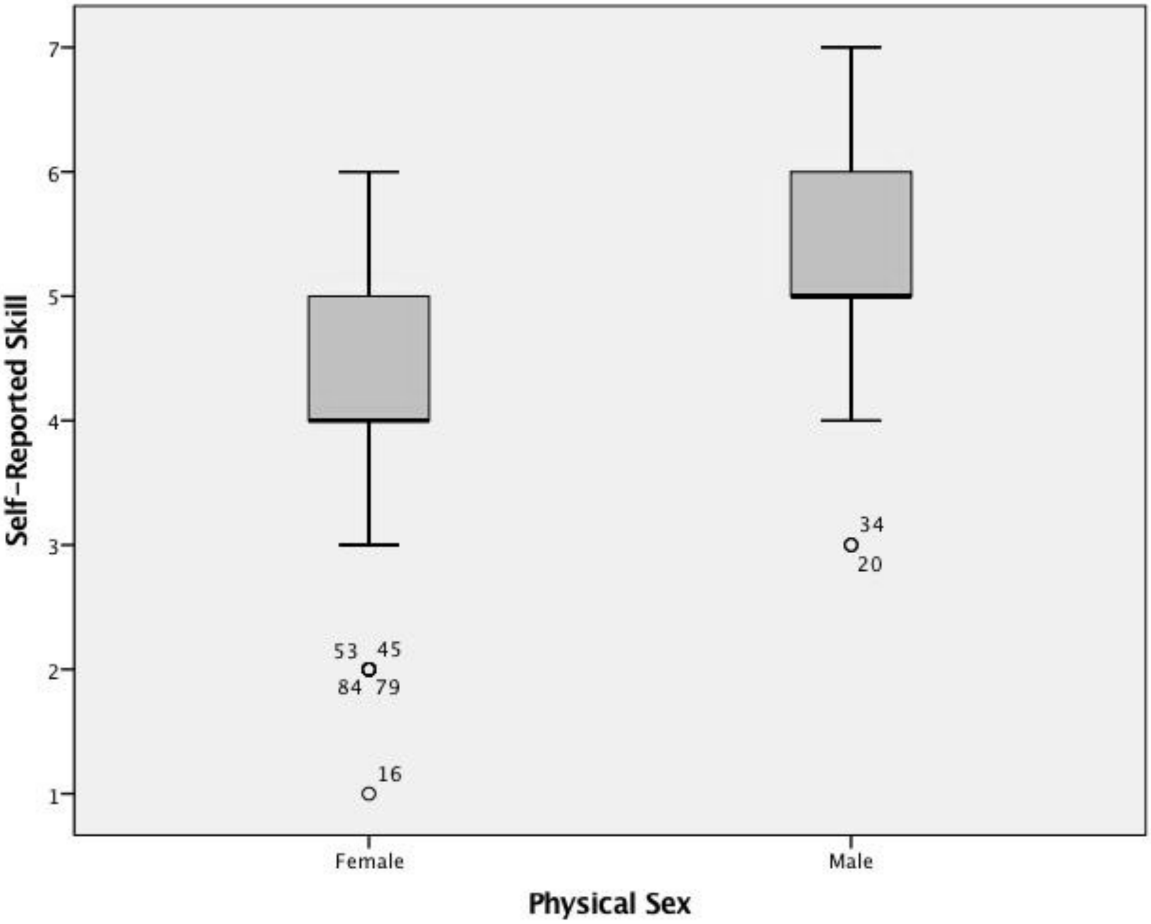

Figure 4: A boxplot representing self-reported video game skill for physical males and physical females. Males reported a higher mean perception of skill than females.

Appendix G

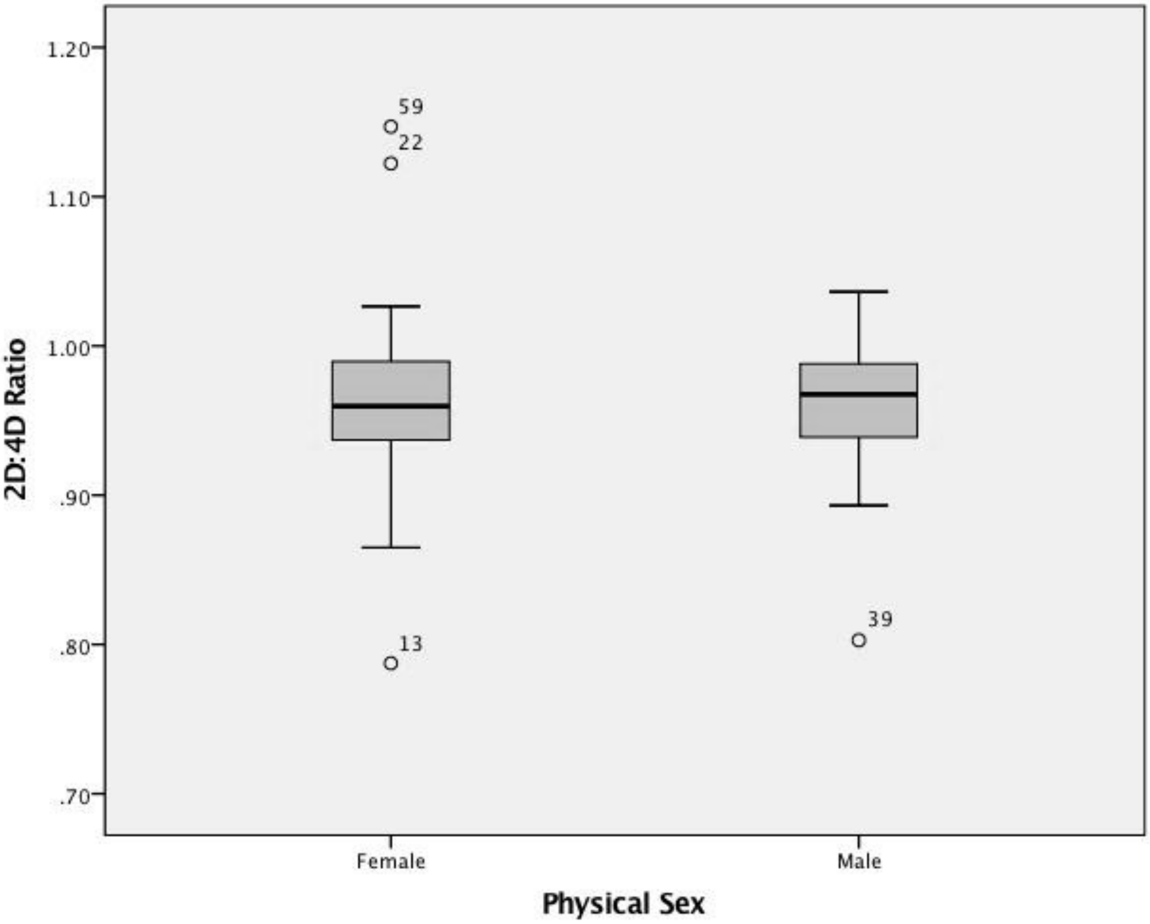

Figure 5: A boxplot representing Left 2D:4D ratio for physical females and males. Females had a slightly higher ratio than males, with higher ratios indicating lower levels of prenatal androgens such as testosterone.