Influences of androgenisation and personality on female sexual orientation, behaviour, and arousal

Luke Holmes

A thesis submitted for the degree of Doctor of Philosophy

Department of Psychology

University of Essex

December 2020

Declaration

I declare that this thesis, *Influences of androgenisation and personality on female sexual orientation, behaviour, and arousal*, represents my own work, except where otherwise stated. None of the work referred to in this thesis has been accepted in any previous application for a higher degree at this or any other University or institution. All quotations have been distinguished by quotation marks and the sources of information specifically acknowledged.

Submitted by: Luke Holmes

Signature of Candidate: 

Date: 13th December 2020

Acknowledgements

This thesis represents the culmination of four years of hard work and effort, and it would not have been possible without the support of a number of people from both inside and outside of academia.

I would like to extend my thanks, first and foremost, to my mentor and friend Gerulf Rieger. I would never have considered postgraduate study had we not – by pure chance – met each other when I was a first-year undergraduate, and struck off what I am sure will be a lifelong friendship. Your constant encouragement, support and "gentle reminders" to stay focused on my goals are the reason I am able to write this acknowledgements section today. Words alone are inadequate to express my gratitude for everything you have done for me over the past seven years.

I would also like to extend a special thanks to Tuesday M. Watts-Overall, my predecessor in obtaining a doctorate from this laboratory. We learned to conduct this research together, and you set a shining example of how a doctoral student should behave, which I have sought to emulate throughout my own studies. I also wish to thank the others within the department who have helped me, through various means, along the way. Chief among these people are Paul Hibbard, the Head of Department, Helge Gillmeister, my co-supervisor, Lesley Monk, the postgraduate administrator, and Lucy Glover, the departmental manager. The help you have offered me has proven invaluable on many occasions.

I would further like to thank my parents, Andrew and Carole Holmes, and to dedicate this thesis to the memory of the latter. Your unwavering belief in me throughout my entire life is a huge part of why I am here today, and I wish you could both be here to see it. I would also like to give special mention to my aunt, Karen Holmes, whose invaluable help at a very difficult time allowed me to continue with this work. Similarly, I would like to acknowledge the patience and support of my partner, Valeria Azara, who I met through this research. Your love and support have been invaluable to me, perhaps more than you know, especially during the more stressful moments of my doctoral studies. I look forward to many more adventures with you. Finally, to Tom and Callum; whose presence has served as a distraction from my troubles and a place to vent when I needed it. Thank you to you all.

Table of Contents

List of Figures

List of Tables

General Abstract

The present work investigated potential hormonal influences on sex differences in sexual arousal, with most men being aroused to one sex, and most women being aroused to both sexes. Within women, I also investigated whether prenatal hormonal influences explain why homosexual women have more male-typical arousal, and are more masculine, than heterosexual women. Chapter 1 summarises the literature. In Chapter 2, I investigated whether sex differences in sexual arousal, reflected in genital response and pupil dilation to sexual stimuli, are related to the sex difference in prenatal androgen exposure, reflected in a putative biomarker, the second to fourth finger digit ratio (2D:4D). In Chapter 3, I examined whether homosexual women's male-typical sexual arousal is explained by their male-typical 2D:4D, as compared to heterosexual women. Chapter 4 investigated whether homosexual women's behavioural masculinity is explained by their male-typical 2D:4D. Finally, Chapter 5 explored an alternative explanation for the sex difference in arousal: That unlike men, women may empathise with actors depicted in explicit sexual stimuli, and thus mirror female actors' sexual arousal by becoming aroused themselves. Overall, we confirmed hypothesised sex differences in sexual arousal, empathy, and 2D:4D. However, there was no evidence that these sex differences were interlinked. In women, we confirmed sexual orientation differences in sexual arousal and masculinity-femininity, but failed to replicate the previously reported sexual orientation difference in 2D:4D. Moreover, there was no evidence that women's sexual arousal patterns or masculinity were related to their 2D:4D. In sum, studied sex and sexual orientation differences might be driven by other factors than prenatal androgen exposure. However, the limitations of 2D:4D need to be considered. Furthermore, sex differences in empathy does not serve as an alternative explanation (alternative to androgen exposure) for sex differences in sexual arousal patterns.

Author Note

Chapters 2 – 5 of this thesis were written as independent manuscripts, with the aim of being submitted for publication in peer-reviewed scientific journals. As such, there is some overlap in the information provided in each chapter. This is most notable in the introductions and the method sections of chapters where similar concepts are explored, or similar measures are used. At the time of writing, Chapter 2 was published in *Journal of Sex Research*, Chapter 3 is under consideration at *Archives of Sexual Behavior*, Chapter 4 is under consideration at *PLOS ONE*, and Chapter 5 is awaiting potential new data before being submitted. As the publication standards vary slightly between journals, some chapters of this thesis are in British English, and others are in American English, though all chapters are formatted according to APA guidelines.

The scholarship of the principle investigator of this thesis was funded by the Economic and Social Research Council, a member of the UK Research and Innovation Body, under their Doctoral Training Centre (DTC) scheme. The research in this thesis was funded by the American Institute of Bisexuality (RSC2519) and the University of Essex (DG00832).

Other Publications

Holmes, L., Watts-Overall, T. M., Slettevold, E., Gruia, D. C., & Rieger, G. (2021). Sex Differences in Sexual Arousal and Finger Length Ratio. *The Journal of Sex Research,* 1-9. doi: 10.1080/00224499.2021.1874262

Jabbour, J., Holmes, L., Sylva, D., Hsu, K. J., Semon, T. L., Rosenthal, A. M., Safron, A., Slettevold, E., Watts-Overall, T. M., Savin-Williams, R. C., Sylla, J., Rieger, G. & Bailey, J. M. (2020). Robust evidence for bisexual orientation among men. *Proceedings of the National Academy of Sciences*, *117 (31),* 18369 – 18377. doi:10.1073/pnas.2003631117

Raines, J., Holmes, L., Watts-Overall, T. M., Slettevold, E., Gruia, D. C., Orbell, S., et al. (2021). Patterns of Genital Sexual Arousal in Transgender Men. Psychological Science, 0956797620971654. doi: 10.1177/0956797620971654

Rieger, G., Holmes, L., Watts-Overall, T. M., Gruia, D., Savin-Williams, R. C., & Bailey, J. M. (2020). Gender nonconformity of bisexual men and women. *Archives of Sexual Behavior*. doi:10.1007/s10508-020-01766-z

Rieger, G., Watts-Overall, T. M., Holmes, L., & Gruia, D. C. (2020). Gender nonconformity of identical twins with discordant sexual orientations: Evidence from video recordings. *Archives of Sexual Behavior*. doi:10.1007/s10508-020-01709-8

Slettevold, E., Holmes, L., Gruia, D., Nyssen, C. P., Watts-Overall, T. M., & Rieger, G. (2019). Bisexual men with bisexual and monosexual genital arousal patterns. *Biological Psychology, 148*, 107763. doi:10.1016/j.biopsycho.2019.107763

Watts, T. M., Holmes, L., Raines, J., Orbell, S., & Rieger, G. (2018). Gender nonconformity of identical twins with discordant sexual orientations: Evidence from childhood photographs. *Developmental Psychology, 54*, 788-801. doi:10.1037/dev0000461

Watts, T. M., Holmes, L., Raines, J., Orbell, S., & Rieger, G. (2018). Finger length ratios of identical twins with discordant sexual orientations. *Archives of Sexual Behavior, 47*, 2435–2444. doi:10.1007/s10508-018-1262-z

Watts, T. M., Holmes, L., Raines, J., Orbell, S., & Rieger, G. (2018). Sexual arousal patterns of identical twins with discordant sexual orientations. *Scientific Reports, 8*. doi:10.1038/s41598-018-33188-2

Watts, T. M., Holmes, L., Savin-Williams, R. C., & Rieger, G. (2017). Pupil dilation to explicit and non-explicit sexual stimuli. Archives of Sexual Behavior, 46, 155–165. doi:10.1007/s10508-016-0801-8

Chapter 1 General Introduction

Research into sexual arousal has uncovered a robust difference between the sexes: Men tend to respond to their preferred sex over the other sex, whereas women tend to respond to both sexes more equally than men (Chivers, Rieger, Latty, & Bailey, 2004). This sex difference has been found using several measures of sexual response, including genital arousal (Chivers et al., 2004; Rieger et al., 2015; Rieger, Savin-Williams, Chivers, & Bailey, 2016; Suschinsky, Lalumière, & Chivers, 2009), pupil dilation (Attard-Johnson, Bindemann, & Ó Ciardha, 2016; Rieger et al., 2015; Rieger & Savin-Williams, 2012), viewing time (Ebsworth & Lalumière, 2012; Israel & Strassberg, 2009) and neural responses (Safron, Sylva, Klimaj, Rosenthal, & Bailey, 2019; Sylva et al., 2013). Thus, men's sexual arousal patterns correspond strongly with their self-reported sexual orientation, whereas in women, their sexual arousal patterns correspond only weakly with their sexual orientation. Previously this has resulted in a debate over whether women have a sexual orientation at all: In men, their sexual orientation is informed by their sexual arousal, and cause them to seek out desirable sexual partners, whereas in women, their sexual arousal may have less of a purpose to orient towards a potential partner, but to respond to a sexual situation (Bailey, 2009).

There are several theories on why this particular sexual arousal pattern occurs in women, which are not necessarily mutually exclusive (see review by Chivers, 2017): For example, that women's sexuality may depend more on contextual and social factors than men's (Baumeister, 2000), or that their sexual responses change according to their levels of fertility, which does not occur in men (Gangestad & Simpson, 2000). The most prominent of these proposed explanations is the "preparation hypothesis" (Suschinsky & Lalumière, 2011), which holds that since forced copulation is common in many animal species, including humans (Galdikas, 1985; Palmer, 1989), and because forced copulation can lead to genital injury (Slaughter, Brown, Crowley, & Peck, 1997), women may have evolved to respond to any sexual situation, even if unwanted or unpleasant, with genital arousal. Their arousal could result in vaginal lubrication, and thereby mitigate this risk of forced copulation, and over time, this may have resulted in females evolving to respond to any type of sexual stimulation with some degree of sexual response. This theory gained traction after a 2011 study in which male and female participants were shown a range of sexual stimuli: While men showed the strongest genital response to stimuli featuring consensual and non-violent sex, women showed similar responses to all stimuli, even those featuring violence (Suschinsky & Lalumière, 2011). However, more recent findings challenge this theory. Although women show an increase in vaginal pulse amplitude to both sexes, they still only lubricate to their preferred sex (Sawatsky, Dawson, & Lalumière, 2018). Despite this challenge to the preparation hypothesis, it is still possible that there is some truth to it. That is, perhaps increased vaginal pulse amplitude is a necessary pre-cursor for lubrication, but it only leads to lubrication when sexual activity is seemingly imminent, due to strong subjective arousal or the presence of a sexual threat. Thus, the preparation hypothesis remains the favored theory for explaining the non-specific sexual arousal patterns seen in women (see review by Lalumière, Sawatsky, Dawson, & Suschinsky, 2020).

The matter is complicated further by the presence of a weak, but reliable, sexual orientation difference within women: In general, all women respond to stimuli featuring either sex, regardless of their own sexual orientation, but homosexual women are an exception to this pattern: Although homosexual women do still respond to stimuli featuring both sexes, they show, on average, a slight preference for stimuli featuring their preferred sex (females) over their non-preferred sex (males). As with the sex difference, this pattern has been detected repeatedly using different measures, including genital arousal (Chivers et al., 2004; Chivers, Seto, & Blanchard, 2007; Rieger et al., 2016), pupil dilation (Rieger et al., 2015; Rieger & Savin-Williams, 2012), and viewing time (Ebsworth & Lalumière, 2012; Lippa, 2012). Thus, homosexual women's responses can be considered somewhat "male-typical", in that they show a preference for one sex other the other, and this pattern is not seen in heterosexual or bisexual women, who exhibit the more "female-typical" pattern of more equal, bisexual arousal.

In addition to their male-typical sexual arousal patterns, homosexual women tend to be more masculine and less feminine than heterosexual women, on average, in their nonsexual behaviors, interests, activities, and self-concepts (Bailey & Zucker, 1995; Lippa, 2020; Rieger et al., 2016). This extensively-studied sexual orientation difference originates in childhood, before individuals are aware of their adult sexual orientation, with longitudinal studies determining that early childhood masculinity robustly predicts homosexual attractions in females later in life (Li, Kung, & Hines, 2017; Xu, Norton, & Rahman, 2019). Furthermore, based on evaluations by others of their childhood photographs and videos, girls who identified as homosexual in adulthood were rated as more masculine and less feminine than girls who later identified as heterosexual (Rieger, Linsenmeier, Gygax, & Bailey, 2008; Watts, Holmes, Raines, Orbell, & Rieger, 2018b). Correspondingly, homosexual women go on to report more male-typical and less female-typical interests than heterosexual women in adulthood (Lippa, 1991, 2005, 2008).

In addition to this gender difference by sexual orientation in masculinity-femininity, there is greater variability in the masculinity-femininity of homosexual women than heterosexual women. Some homosexual women are particularly masculine, both in comparison with heterosexual women, and in comparison with other homosexual women, and this pattern has been found in self-reports about male-typical versus female-typical interests and activities, as well as observer ratings of women's male-versus female-typical behaviors (Lippa, 2005, 2008, 2015; Rieger et al., 2008; Rieger, Linsenmeier, Gygax, Garcia, & Bailey, 2010; Rieger et al., 2016). This variability is well-known in homosexual women's culture, with those who self-identify as masculine calling themselves "butch" and those who self-identify as feminine calling themselves "femme" (Brown, Finn, Cooke, & Breedlove, 2002a).

The present research program focused on several possible connections of the aforementioned distinctions between women of different sexual orientations. First, there may be a relationship between homosexual women's male-typical arousal and male-typical behavior, possibly because they have been shaped by a common factor. That is, perhaps there is a core influence such as prenatal androgenization that explains why homosexual women are, in general, more male-typical in both their sexual arousal and other, nonsexual behaviors and expressions. Previous research has examined a relevant hypothesis, assuming that masculinity-femininity is an indicator of early androgenization (Rieger et al., 2016). The prediction was that homosexual women's male-typical arousal (as compared to heterosexual women's) would be statistically explained (mediated) by their degree of nonsexual masculinity-femininity (measured via self-report and observer ratings). However, although both of these findings were separately confirmed (with homosexual women having male-typical arousal patterns and more masculine non-sexual behaviors), there was no indication that these findings were related in any way (Rieger et al., 2016).

A further focus of the present research was on the aforementioned finding that there is greater variability in homosexual women's degree of masculinity-femininity compared to heterosexual women. There is also extensive variability in women's sexual arousal patterns in general (Chivers et al., 2007; Rieger et al., 2015), and some homosexual women are particularly male-typical in their arousal compared to others (Rieger et al., 2016). Hence, those homosexual women who display the most male-typical sexual arousal patterns could also be those who show the most male-typical behavior as compared to other women. This hypothesis has also been tested, but no such pattern could be confirmed. A conclusion of that study was that even though female homosexuality links to male-typical sexual arousal, and, separately, to male-typical nonsexual behavior, these sexual and nonsexual traits did not appear to be influenced by a common factor (Rieger et al., 2016).

Yet, perhaps non-sexual masculinity-femininity is a poor measure of prenatal androgenization, and the previous research by Rieger et al. (2016) may have found a spurious null result for this reason. Perhaps a more appropriate measure of androgenization would better explain the male-typical sexual arousal patterns (if not the male-typical nonsexual behaviors) of homosexual women. A large body of literature suggests that most differences between the sexes, including sexual orientation and gender behavior, may be the result of exposure to elevated levels of androgens (testosterone in particular) during prenatal development (Breedlove, 2017; Morris, Jordan, & Breedlove, 2004). Since true causal evidence for this link cannot be ethically gathered from humans, it is informative to look at animal work on the effects of androgen exposure during fetal development. For example, female guinea pigs exposed to testosterone during foetal development displayed more male-typical mounting behaviours and less female-typical lordosis behaviours in adulthood than control females. These behavioural changes were permanent, and persisted after the foetal treatment had ceased (Arnold, 2009). Research on primates has found similar results: Pregnant female rhesus macaques were exposed to testosterone either at an early stage of gestation (day 40 – day 64, roughly equivalent to the first trimester in humans) or a late stage of gestation (day 115 – day 139, roughly equivalent to the third trimester in humans). Early-androgenized females showed somewhat more male-typical genitalia in addition to male-typical sexual behaviors, such as different kinds of mounts which are normally only displayed by males. Additionally, they groomed their mothers at a reduced rate – a male-typical non-sexual pattern of behavior. Late-androgenized females tended to show male-typical non-sexual behaviors such as rough play, play initiation, and being the recipient of play initiation (Goy, Bercovitch, & McBrair, 1988). It therefore appears that sexual and nonsexual behaviors may be masculinized by androgen exposure during different periods of prenatal development, but that there is some degree of overlap between periods which masculinize sexual and non-sexual behaviors. However, due to biological differences, animal studies cannot be taken as authoritative with regards to human populations.

Although animal models suggest a causal link between prenatal androgen exposure and adulthood behavior, in humans, we must rely on the use of biomarkers – lasting indications on the body which are hypothesized to vary according to an individual's degree of exposure to androgens during fetal development. Of these biomarkers, the easiest to measure is the ratio of the length of the second digit to that of the fourth digit (2D:4D). This ratio shows a reliable sex difference, with men having lower (more masculine) ratios than women on average, suggesting they are exposed to higher levels of androgens even during the early stages of fetal development (Galis, Ten Broek, Van Dongen, & Wijnaendts, 2010; Grimbos, Dawood, Burriss, Zucker, & Puts, 2010; Hönekopp, Bartholdt, Beier, & Liebert, 2007; Xu & Zheng, 2015). Since, as previously mentioned, prenatal androgen exposure may be the underlying factor explaining the majority of sex differences in brain and behavior, it follows that this reliable sex difference in 2D:4D should be linked to the sex difference in sexual response patterns. In Chapter 2, we examined whether this was the case.

Furthermore, in women (but not in men), 2D:4D has repeatedly been linked to sexual orientation, with homosexual women having lower (more male-typical) 2D:4D than heterosexual women on average, which suggests that homosexual women are exposed to higher levels of androgens during fetal development than heterosexual women (Kraemer et al., 2006; Putz, Gaulin, Sporter, & McBurney, 2004; Rahman, 2005; Rahman & Wilson, 2003; Wallien, Zucker, Steensma, & Cohen-Kettenis, 2008; Watts, Holmes, Raines, Orbell, & Rieger, 2018a; Williams et al., 2000). Furthermore, one study suggests a link between 2D:4D and masculinity-femininity within homosexual women, with homosexual women who self-identify as “butch” (i.e. masculine) having a lower (more male-typical) 2D:4D than homosexual women who identify as “femme” (i.e. feminine) (Brown et al., 2002a). Thus, it is possible that prenatal androgen exposure – measured via 2D:4D – may prove to be the factor which explains why homosexual women are, on average, more male-typical in both (or either) their sexual arousal patterns and their nonsexual behaviors and self-concepts. Moreover, variation in 2D:4D could potentially explain increased variation of sex-typed traits within homosexual women, with some being more masculinized for a given trait than others.

Since the exact relationship between sexual arousal, sexual orientation, 2D:4D and masculinity-femininity is unclear, we conducted two separate empirical studies centered around 2D:4D as putative indicator of androgenization. In Chapter 3, we examined whether the relationship of male-typical sexual arousal and female sexual orientation can be explained by 2D:4D ratio. We also examined whether we could find evidence for the previously-stated hypothesis that the link between male-typical sexual arousal and sexual orientation can be explained by another proxy of early testosterone exposure, which was behavioral masculinity-femininity (Rieger et al., 2016).

In Chapter 4, we attempted to explain the link between behavioral female masculinity-femininity and sexual orientation using our putative measure of prenatal androgenization, 2D:4D. Additionally, because of the evidence that homosexual women may be more variable in masculinity-femininity than other women, we further examined whether variation in 2D:4D explained why some homosexual women are particularly masculine, and others are not. That is, we investigated whether the most masculine homosexual women had the most male-typical digit rations, both compared to heterosexual women but also to other homosexual women who are not very masculine.

A further goal of the present research program was unrelated to the studies that focused on female sexual orientation: Instead, we aimed to explain overall sex differences in sexual response. This particular study was based on the feedback from our participants. We sometimes present participants in the laboratory with their own results, and women are often surprised to find that they responded equally to both males and females despite having a strong preference for one over the other. Although many participants are amused and puzzled by this trend – and can offer no explanation for it – several female participants offered the explanation of empathy. Specifically, some suggested that women automatically put themselves “in the shoes” of the pornographic actors (particularly the females) and become aroused on her behalf. Explaining the sex difference in sexual arousal via this personality trait is, to our knowledge, not something previously attempted in any study. However, these suggestions by participants do fit with the literature on empathy: Several studies have pointed out that women have a greater degree of empathy, on average, than men. This difference first emerges in childhood, widens in adolescence, and carries on into adulthood (Hoffman, 1977; Lam, Solmeyer, & McHale, 2012; Rose & Rudolph, 2006). A subscale of these measures of empathy is “perspective-taking”, which deals specifically with putting oneself into the shoes of an observed individual, such as in the manner described by participants. The idea of a link between empathy and sexual arousal is further bolstered by plausible joint neurological and developmental mechanisms. Mirror neurons fire both when an action is performed and when the same action is observed: For example, a mirror neuron which responds when an individual grasps an object would respond identically when watching the hand of another individual grasping an object (Rizzolatti & Craighero, 2004), and previous research has shown that a person’s degree of self-reported empathy can predict the strength of this mirror neuron response (Jabbi, Swart, & Keysers, 2007). Furthermore, mirror neurons were reported to activate in response to sexual stimuli in a sample of male participants (Mouras et al., 2008). As such, activation of mirror neurons provides a potential explanation for how an observer may experience genital arousal from witnessing an actor in a stimulus experience genital arousal, regardless of the observer’s own sexual orientation, and the sex difference in empathy could explain why this occurs in women more strongly than in men.

In addition to mirror neurons as a potential explanation for the link of empathy to sexual response, there appears to be a link between empathy levels and prenatal androgen exposure. One study measured androgen levels in the amniotic fluid of 193 pregnant women, and later found that those children exposed to higher levels of androgens in the womb scored lower on self-reported empathy at ages 6-8 (Chapman et al., 2006). Another study measured androgens in children’s saliva and found similar results, with higher androgen levels predicting lower empathy scores (Pascual-Sagastizabal et al., 2013). We therefore consider the possibility of a link between empathy and sexual arousal patterns to be plausible, and we explore this in Chapter 5.

Thus, in broad terms, this PhD thesis attempted to explain sex and sexual orientation differences in sexual arousal and masculinity-femininity, with a focus on women. However, the structure of this thesis is not simply three studies which relate to each other in a linear fashion. Chapter 2 was concerned with sex differences in sexual response patterns between men and women, and whether these can be explained by a corresponding sex difference in 2D:4D. Chapter 3 examined whether differences in 2D:4D or masculinity-femininity can explain sexual orientation differences in sexual arousal among women. Chapter 4 examined whether differences in 2D:4D can explain sexual orientation differences in masculinity-femininity among women, or whether 2D:4D can differentiate between subgroups of homosexual women, distinguished from one another by their masculinity-femininity. Chapter 5 examined whether differences in empathy can explain the general difference between men and women in their patterns of sexual arousal. Although these are discrete studies intended to form self-contained manuscripts, in combination, they were designed to inform on the relationships between sexual characteristics, prenatal androgen exposure, and aspects of personality in women.

Chapter 2

Sex Differences in Sexual Arousal and Finger Length Ratios

2.1 Abstract

Most men show sexual arousal to one, preferred sex, whereas most women respond to both sexes, regardless of their sexual orientation. A different research program indicates that men have lower second-to-fourth finger length ratios (2D:4D) than women, possibly because men are exposed to higher levels of androgens during prenatal development. We hypothesized that sex differences in sexual arousal patterns are influenced by prenatal androgen exposure and would thus be explained by sex differences in 2D:4D. We measured the sexual response patterns of 139 men and 179 women via genital arousal and pupil dilation to erotic videos, in addition to their 2D:4D. Compared to women, men showed stronger responses to one sex over the other, although this pattern was clearer in genital arousal than pupil dilation. Men also had lower 2D:4D than women. However, there was no evidence that sex differences in sexual arousal related to sex differences in 2D:4D. Thus, whichever factor explains sex differences in sexual arousal patterns may not be reflected in 2D:4D.

2.2 Introduction

Research measuring sexual responses to explicit stimuli has uncovered a robust difference between the sexes: Most men show a strong sexual response to stimuli featuring their preferred sex, and little to no response to stimuli featuring their non-preferred sex. In contrast, women tend to respond equally to stimuli featuring either their preferred or non-preferred sex regardless of their sexual orientation (Bailey et al., 2016). This sex difference in sexual response is highly robust, and has been detected using measures of genital arousal (Rieger et al., 2015; Suschinsky et al., 2009), pupil dilation (Attard-Johnson et al., 2016; Rieger et al., 2015; Rieger & Savin-Williams, 2012), viewing time (Ebsworth & Lalumière, 2012; Israel & Strassberg, 2009), and neural responses (Safron et al., 2019; Sylva et al., 2013) to sexual stimuli. For this sex difference, a "specific" pattern of sexual arousal to one sex over the other has sometimes been described as "male-typical," and a "non-specific" pattern of arousal to both sexes as "female-typical" (Chivers et al., 2007; Rieger et al., 2016). There are exceptions to this general sex difference. For instance, in men, sexual responses to one, preferred gender is more pronounced in heterosexual and homosexual men than bisexual men (Jabbour et al., 2020). In women, non-specific responses to both sexes are more common in heterosexual women than bisexual or homosexual women (Rieger et al., 2016). In general, however, non-specific sexual arousal to both genders characterizes women more than men (Bailey, 2009).

Several (not necessarily mutually exclusive) theories seek to explain this non-specific pattern of sexual arousal in women (see review by see review by Chivers, 2017), including that women’s sexuality may depend more on contextual and social factors than men’s (Baumeister, 2000; Diamond, 2003), or that their sexual responses may change according to their levels of fertility, which does not occur in men

(Gangestad & Simpson, 2000). The most prominent proposal is the “preparation hypothesis” (Suschinsky & Lalumière, 2011). This hypothesis is based on the observation that forced copulation is common in many species, including humans (Galdikas, 1985; Palmer, 1989). Because forced copulation can lead to genital injury (Slaughter et al., 1997), women may have evolved to respond to any sexual situation with sexual arousal, as it leads to protective lubrication that minimizes the risks of genital trauma. The need for such a protective mechanism may have been so strong that females have evolved to respond to any sexual situation, perhaps especially if it contains cues to possible vaginal penetration, even if unwanted or unpleasant (Bossio, Suschinsky, Puts, & Chivers, 2014). Consistent with this hypothesis, heterosexual men, on average, show the strongest genital response to stimuli depicting consensual, non-violent sexual intercourse, while heterosexual women show similar responses to all sexual stimuli, including those featuring violent, nonconsensual sex (Suschinsky & Lalumière, 2011). However, more recent findings challenged this hypothesis: Although women show an increase in vaginal pulse amplitude to both sexes, they only lubricate to their preferred sex (Sawatsky et al., 2018). Yet, it is possible that increased vaginal pulse amplitude is a necessary pre-cursor for lubrication, but it only leads to lubrication when sexual activity is seemingly imminent, due to strong subjective arousal or the presence of a sexual threat. Thus, the preparation hypothesis remains the favored theory for explaining the non-specific sexual arousal patterns in women (see review by see review by Lalumière et al., 2020).

Unlike the aforementioned work, the present study did not focus on an ultimate explanation for the sex difference in sexual response. Instead, we examined potential developmental differences between men and women, to determine whether this could explain the difference in their sexual responses. In mammals, exposure to prenatal androgens – specifically testosterone – is thought to account for the majority of sex differences in brain and behavior (Breedlove, 2017; Swift-Gallant, Johnson, Di Rita, & Breedlove, 2020). For instance, individuals with an XY karyotype and complete androgen insensitivity syndrome (CAIS), which results in insensitivity to androgens throughout the lifespan, typically report sexual orientations towards males with the same frequency as genetic women (Wisniewski et al., 2000), and do not differ from genetic females in a wide range of psychosexual measures, such as gender identity and gender role behavior in both childhood and adulthood (Hines, Ahmed, & Hughes, 2003). Additionally, individuals with XY karyotype and CAIS show neural responses to both male and female sexual stimuli, and therefore have female-typical (non-specific) sexual responses (Hamann et al., 2014). Thus, if levels of early androgen exposure influence the development of sex differences, in general (Bailey et al., 2016; Breedlove, 2010; Motta-Mena & Puts, 2017; Puts & Motta-Mena, 2018), then they could possibly also affect sex differences in the specificity of their sexual arousal.

Since the direct measurement of androgen exposure *in utero* is difficult to achieve in humans, the majority of research relies on biomarkers – lasting indications on the body that are thought to reflect degree of androgen exposure during fetal development. Of these, the easiest to measure is the ratio of the length of the second digit to that of the fourth digit (2D:4D). Reviews and meta-analyses indicate that men have, on average, lower (more masculine) ratios than women, with the proposal that men's exposure to higher levels of androgens during fetal development drives this sex difference in digit ratio (Grimbos et al., 2010; Hönekopp et al., 2007; Swift-Gallant et al., 2020; Xu & Zheng, 2015).

It is worth noting that 2D:4D is a controversial measure, with ongoing debates about causation (McCormick & Carré, 2020; Swift-Gallant et al., 2020) and validity of any associated findings due to noise in the data (Bailey et al., 2016). However, much of this controversy pertains to the relationship between 2D:4D and sexual orientation, rather than sex differences in 2D:4D. Whereas a previous meta-analysis suggested publication bias with respect to the link of 2D:4D with sexual orientation – resulting in a potential overestimation of the strength of the relationship between the two – the same meta-analysis found no evidence of publication bias with regards to a sex difference in 2D:4D (Grimbos et al., 2010). Additionally, in individuals with XY karyotype and CAIS, finger length ratios are feminized, similar to those of unaffected females (with XX karyotype) (Berenbaum, Bryk, Nowak, Quigley, & Moffat, 2009). Thus, in the complete absence of androgens, it appears that individuals with XY karyotype develop female-typical 2D:4D along with female-typical psychosexual traits (Hines et al., 2003). In addition, in individuals with an XX karyotype and Congenital Adrenal Hyperplasia, which results in excessive androgen exposure throughout the lifespan, have more male-typical finger length ratios than unaffected females (Brown, Hines, Fane, & Breedlove, 2002b). That being said, there has been debate in the literature about whether any differences in 2D:4D, including sex differences, are confounded by overall digit length: As humans are a sexually dimorphic species with males being larger than females (Kurki, 2011) and having longer finger digits than females, on average (Kratochvíl & Flegr, 2009), the impact of finger length on 2D:4D could be substantial. As such, in the present research we made an effort to control for finger length, in an attempt to ensure that any detected differences in 2D:4D represent are not confounded by it.

In sum, the literature suggests a sex difference in sexual arousal patterns, with men generally showing strong responses to one sex over the other, whereas women show, compared to men, more equal responses to both sexes, regardless of their own sexual orientation. This sex difference in sexual arousal patterns may be driven by sex differences in prenatal androgen exposure, which may themselves be reflected by 2D:4D.

We therefore tested the following hypotheses:

Hypothesis 1. Men's sexual arousal patterns will be more sex-specific, on average, whereas women's sexual responses will be less sex-specific.

Hypothesis 2. On average, men will have a lower 2D:4D than women.

Hypothesis 3. Sex differences in the specificity of sexual arousal will be mediated by a sex difference in 2D:4D.

2.3 Method

2.3.1 Participants

Based on estimates from pooled previous data produced by our research group, we predicted that the effect size (Cohen's *d*) of the sex difference in employed variables of genital arousal to be *d* = 1.18. For the corresponding sex difference in pupil dilation we expected *d* = 1.00 (Rieger & Savin-Williams, 2012). For the sex difference in 2D:4D, we based our estimates on those published in a meta-analysis (converted from Hedge's *g* to Cohen's *d*), and anticipated that the effect size would be approximately *d* = .55 for the right hand, and *d* = .44 for the left hand. Power analyses in G*Power 3.1.9.7 (Faul, Erdfelder, Lang, & Buchner, 2007) estimated, for the smallest of these effects (*d* = .44) a minimum of 220 participants to achieve significant results with a power of .90. With regards to the mediation, estimating the necessary sample size proved difficult, as no other study has conducted a mediation in the same manner as the present study. We therefore erred on the side of caution with participant numbers: Our power analysis for the main effect was based on the more conservative power value of .90 rather than the commonly-used .80, resulting in a sample size requirement of 220 instead of 166 for the smallest expected main effect (*d* = .44). Additionally, we continued recruiting past this figure as participants were visiting our laboratory for other studies, resulting in a final sample size of 318 –substantially larger than that recommended by the power analysis.

We recruited participants via pride festivals, online magazines, and university fairs and mailing lists, and participants reported their sexual orientation using a 7-point scale (Kinsey, Pomeroy, Martin, & Gebhard, 1953). The 139 recruited men self-identified as "exclusively straight" (n = 33), "mostly straight" (n = 10), "bisexual leaning straight" (n = 4), "bisexual" (n = 10), "bisexual leaning gay" (n = 12), "mostly gay" (n =

11), or “exclusively gay” (n = 59). The 179 women self-identified as “exclusively straight” (n = 37), “mostly straight” (n = 25), “bisexual leaning straight” (n = 9), “bisexual” (n = 15), “bisexual leaning lesbian” (n = 11), “mostly lesbian” (n = 41), or “exclusively lesbian” (n = 41). The mean (SD) age was 24.45 (8.85) for men and 24.18 (7.21) for women. For men, 83% were White, followed by 3% Chinese, 3% Indian, and other ethnicities. For women, 79% were White, 5% Chinese, 4% Black, and other ethnicities.

Only participants for whom valid 2D:4D data were available were entered into the current sample. As such, 2D:4D data were available for all 318 participants, apart from one female participant whose data were lost for the right hand. Some participants opted out of the genital arousal component, and independent of this, we experienced pupil data loss because of problems with the apparatus. Consequently, genital arousal data were available for 305 of the 318 participants, and pupil data for 273 of the 318 participants. Thus, the number of participants varies across analyses, and the specific number of participants included in each stage of analysis is listed in the caption of the corresponding table.

2.3.2 Measures and Materials

Self-reported sexual orientation. Participants reported both their sexual orientation and sexual attraction to men and women on 7-point scales (Kinsey et al., 1953). These scales were highly correlated in both men $p < .0001$, $r = .98$, 95% CI [.98, .99] and women, $p < .0001$, $r = .97$ [.96, .98], and therefore averaged within participants. For this average, a score of 0 represented exclusive heterosexuality, and 6 represented exclusive homosexuality. This composite score was used for all analyses.

2D:4D. Digit measurements were taken from either high-resolution photographs or scans of participants' hands. For the photographs, participants placed their hands on a flat surface in a supinated (palms facing up) position, with their fingers slightly spread apart, and images were taken from approximately 30 cm above this surface. For the scans, participants placed their hands flat in a pronated (palms facing down) position on the surface of the scanner. In both cases, the palmar surfaces of the hands were visible in the resultant images. Different methods of capturing images (photograph or scanner) did not moderate the relationship between sex and 2D:4D.

Images were measured in a double-blind procedure (with regards to participant sex and sexual orientation) by two independent raters. Measurements were performed with the vector graphics package Inkscape 0.92, as computer-assisted techniques produce the most reliable measurements (Allaway, Bloski, Pierson, & Lujan, 2009). Each rater drew a line as wide as the finger along the proximal skin crease at the base of the finger, between the metacarpal and proximal phalanx. A second line was drawn downwards from the tip of the finger, where it automatically snapped to the center of the base line. Raters then zoomed in on the tip of the finger for fine adjustments, to ensure that this line matched the tip as closely as possible. Measurements for each digit were averaged between raters, as inter-rater reliability (Cronbach's alpha) exceeded .99 for each digit. For each hand, 2D:4D was calculated by dividing the averaged length of the index finger by the averaged length of the ring finger.

It is possible that sex differences in 2D:4D may be confounded by overall digit length (Kratochvíl & Flegr, 2009). To examine this in the present data, we also kept the raw length variables for each digit for use as covariates.

Stimuli. The sexual stimuli consisted of 3-minute videos, three featuring a female model and three featuring a male model, each of them masturbating in a bedroom. During masturbation females were penetrating themselves with toys or their fingers. These stimuli were selected in a previous study in which 200 videos were rated on their sexual appeal by men and women of different sexual orientations (Rieger et al., 2015), and the top three female and male videos were used in the present study. Neutral stimuli to assess baseline genital responses were 2-minute clips taken from a nature documentary. Their engaging but nonsexual content facilitates participants' return to an unaroused level, and this has worked in previous projects (Rieger et al., 2015; Watts, Holmes, Raines, Orbell, & Rieger, 2018c). However, these nature videos were not used for pupil dilation baseline, as their engaging content might elicit dilation for reasons other than sexual arousal. Thus, two 1-minute animations of clouds were used to obtain a pupillary baseline. All videos were edited using MPEG Streamclip and Final Cut Pro to be of similar luminance.

Genital arousal. For both sexes, genital response was captured using a BIOPAC MP150 data acquisition unit and the AcqKnowledge software. For men, genital arousal was measured as change in penile circumference with a penile strain gauge. Prior to each participant, the gauge was calibrated on a cone at 80mm and 110mm. The signal was sampled at 200 Hz, low-pass filtered to 10 Hz and digitized with 16- bit resolution. For women, genital arousal was measured as changes in peak-to-trough vaginal pulse amplitude (VPA) with a vaginal photoplethysmograph. The signal was sampled at 200 Hz, and high-pass filtered at 0.5 Hz with 16-bit resolution. The VPA exhibits both convergent and discriminant validity for the measurement of female sexual response (Suschinsky et al., 2009).

Pupil dilation. Pupil dilation was measured with a SR Research EyeLink 1000 eye tracking unit. A 35 mm lens focused on the participant's right eye, positioned approximately 60 cm from the participant's head, and sampling at a rate of 500 Hz. The infrared light emitted by the eye tracker is reflected by the pupil, and the number of pixels reflected were recorded. Because raw pupil area data included "0's" for missing values, for instance from blinks or head movements, these values were removed prior to analyses.

2.3.3 Procedure

Participant session. The University of Essex's Ethics Committee approved this study (GR1702). All participants were over the age of 18 and provided written informed consent. After giving consent, participants completed a survey on their demographics and sexual orientation, and had their hands photographed or scanned. Participants were then seated in a sealed booth, with dim lighting conditions. Eyes were calibrated by participants fixating on dots outlining the screen. Participants were instructed on how to use the genital probe, and then were left to apply it in privacy. The experimenter was contactable via an intercom throughout the experiment. The signal from both the genital device and the eye tracker were checked before the experiment commenced. Participants were instructed to watch the screen throughout the experiment, regardless of whether they enjoyed the content. Participants first viewed an animation of clouds, followed by alternating sexual and nature videos. These were displayed in a random order, but a sexual video was always followed by a nature video. Following the sixth nature video, a final animation of clouds was displayed. Participants were compensated for their time. The entire procedure took approximately 90 minutes.

Data treatment. Following previous procedures (Watts et al., 2018c), genital data and pupil data were averaged across the duration of each stimulus and for each participant. These averages were then standardized within participants, producing a z-score for each participant and stimulus. For genital data, standardized responses to the 5 seconds preceding each sexual stimulus (following the display of a neutral stimulus, and after the participant had returned to baseline) were subtracted from the standardized response to the sexual stimulus. For pupil data, standardized responses to neutral stimuli (the animated clouds) were subtracted from standardized responses to all sexual stimuli. We then computed, for each participant, average responses across all sexual stimuli of a given type (female or male), which reflected their responses to each sex as compared to baseline.

These standardized response scores were used to calculate the two experimental variables. The first was the absolute difference between each participant's responses to males and females, calculated by deducting one mean from the other, such that zero indicates equal responses to males and females, and deviation from zero means a stronger response to one sex over the other. We expected men to have a large difference in their sexual arousal to one sex and the other. For example, a heterosexual man should show strong arousal to females, and little to males. Conversely, a homosexual man should show strong arousal to males and little to females. In each case, there would be a notable absolute difference in their responses to males or females (but see our below comment on bisexual men). In contrast to heterosexual and homosexual men, heterosexual and homosexual women (and bisexual women) were expected to show smaller absolute differences, because they would respond similarly to both males and females. Thus, regardless of their sexual orientations, we expected men's absolute differences in their arousal to males or females to be larger than women's (Rieger et al., 2015).

The second variable used in the present study was participants' responses to their less-arousing sex, or their minimum arousal. To calculate this, we selected the mean response to whichever stimulus category (male or female) each participant had a lower response to. For men (relative to women), we expected their response to their less-arousing sex to be low, because they often respond strongly to one sex and weakly to the other, regardless of their sexual orientation. For women, more than men, we expected their response to their less-arousing sex to be higher, because women of all sexual orientations are more likely to respond to both sexes, including their less-arousing sex (Rieger et al., 2015).

There was the possibility of an interaction between sex and sexual orientation affecting sexual response, because bisexual men's responses can be more bisexual than the responses of heterosexual or homosexual men, whereas in women, bisexual arousal is observed across all sexual orientations (Rieger et al., 2016; Slettevold et al., 2019). In fact, such a pattern was found in the present data. However, the present research had no main focus on sexual orientation, but rather on sex differences. Therefore, for the sake of simplicity, we did not concentrate on complex analyses that differentiated bisexual from heterosexual and homosexual men and women, and we decided to investigate sex differences across all participants, regardless of their sexual orientations. Moreover, the inclusion or exclusion of bisexual individuals (and in fact, of homosexual individuals) did not change the direction of main findings with respect to sex differences in sexual arousal, although, overall, effect sizes (*d's*) increased if only heterosexual men and women were compared, with the largest change being d = 1.96 to d = 2.69 for genital arousal to the less-arousing sex, and the smallest change being *d* = -.31 to *d* = -.75 for absolute difference in pupil dilation. However, the exclusion of homosexual and bisexual individuals left the sex difference in 2D:4D only marginally significant in the left hand (*p* = .06), although the right hand remained significant (*p* = .03), and the sex differences slightly increased, from .30 to .31, and from .30 to .36, respectively.

2.4 Results

2.4.1 Hypothesis 1.

We hypothesized that men's sexual responses will be more sex-specific, on average, whereas women's responses will be less sex-specific. We first examined the absolute difference in sexual arousal (measured via genital arousal and pupil dilation) to one sex or the other, expecting that it would be greater in men than in women. For genital arousal, an independent-samples t-test indicated that the absolute difference score was significantly higher in men (M = 1.41, SD = .63) than in women (M = .73, SD = .55), t(292) = -9.93, p < .0001, d = 1.17, 95% CI [1.11, 1.24] (Figure 2.1A). Similarly, for pupil dilation, the absolute difference score in men (M = .45, SD = .37) was significantly greater than the absolute difference score in women (M = .35, SD = .30), t(271) = -2.56, p = .01, d = .31 [.27, .36] (Figure 2.1B).

We then repeated these analyses for sexual response to the less-arousing sex, expecting it to be lower in men than in women. For genital arousal, men's responses to the less-arousing sex (M = .24, SD = .41) were significantly lower than women's responses to the less-arousing sex (M = 1.28, SD = .61), t(292) = 16.60, p < .0001, d = -1.96 [-2.02, -1.90] (Figure 2.2A).

For pupil dilation, men's responses to the less-arousing sex were not, as hypothesized, significantly lower, but rather greater (M = .62, SD = .64) than women's (M = .46, SD = .62) – this sex difference was significant, t(271) = -2.10, p = .04, d = .26 [.18, .33] (Figure 2.2B). As men and women differed as expected in their absolute difference in pupil dilation (Figure 2.1B), but men unexpectedly showed stronger dilation than women to their less-arousing sex (Figure 2.2B), we reasoned that this must be driven by an unpredicted difference between men and women in their pupil dilation to their more-arousing sex, whichever sex this may be. That is, we did not hypothesize any difference between men and women in their responses to the more-arousing sex, but investigated this possibility as a potential explanation for the unexpected findings thus far. In fact, men's pupil dilation to their more-arousing sex (M = 1.08, SD = .60) was significantly greater than women's (M = .81, SD = .67) $t(271)$ = -3.41, p = .0007, d = .42 [.34, .50]. We revisit this finding in the Discussion.

2.4.2 Hypothesis 2.

We hypothesized that, on average, men will have a lower 2D:4D than women. In their left hand, men (M = .96, SD = .04) had significantly lower 2D:4D than women, (M = .98, SD = .04) $t(316)$ = -2.68, p = .008, d = -.304 [-.309, -.300] (Figure 2.3A). Similarly, in their right hand, men (M = .96, SD = .04) had significantly lower 2D:4D than women, (M = .98, SD = .04) $t(315)$ = -2.62, p = .009, d = -.298 [-.302, -.294] (Figure 2.3B).

We further examined whether overall digit length was a confounding variable in the relationship between sex and 2D:4D. To do this, we computed a series of regression analyses predicting either left-hand or right-hand 2D:4D by participant sex and a single digit length variable from the same hand (as digit lengths were highly correlated with each other, r >.98, controlling for both digit lengths simultaneously produced collinearity issues). Sex differences in 2D:4D remained significant, and similar in magnitude, regardless of whether digit length as a potential confound variable was included in the regression analysis or not. For instance, the sex difference in left-hand 2D:4D was significant and identical in effect before controlling for left-hand fourth digit length, p = .008, β = .15 [.04, .26], and afterwards, p = .008, β = .14 [.04, .26]. Thus, digit length did not appear to be a confounding factor for observed sex differences in digit ratios.

2.4.3 Hypothesis 3.

We hypothesized that sex differences in the specificity of sexual arousal will be mediated by a sex difference in 2D:4D. To investigate this, we computed a total of 12 regression analyses predicting absolute difference or response to the less-arousing sex for both genital arousal and pupil dilation. In each analysis, in Step 1, sex was the only predictor of sexual response. In Step 2, one 2D:4D variable – either left-hand (Table 2.1) or right-hand (Table 2.2) was included alongside sex as a predictor. If 2D:4D explained differences between men and women in their sexual response, then the inclusion of either 2D:4D variable should weaken the relationship between sex and sexual response.

Both Table 2.1 and Table 2.2 show that the inclusion of either 2D:4D variable had almost no effect on the relationship between sex and any of the four measures of sexual response. In all Step 2 analyses, the relationship between sex and measure of sexual response remained as strong in effect, and as significant, as in Step 1. However, to systematically test for potential mediation effects of 2D:4D variables, we followed these regression analyses with mediation analyses on the basis of 10,000 bootstrapped samples (Preacher & Hayes, 2008). Neither measure of 2D:4D significantly mediated the relationship between sex and any of the four measures of sexual response, as the confidence intervals of the indirect effects included zero in all eight computed analyses. Betas of indirect effects ranged from -.002 to .030, and their CI's ranged from -.041 to .073.

2.5 Discussion

The present research confirmed, in general, two sex differences that have been previously reported in separate research programs: men and women differ in the specificity of their sexual response patterns (Chivers et al., 2004; Rieger et al., 2015; Rieger et al., 2016; Suschinsky et al., 2009), and men have lower 2D:4D than women in both hands (Grimbos et al., 2010). Yet in the present data there was no evidence that these two patterns were related. If one assumes that 2D:4D is a valid measure of prenatal androgen exposure, then this would suggest that sex differences in sexual arousal patterns and digit ratios develop independently of each other. A sex difference in 2D:4D appears to show in fetuses as young as 9 weeks of gestation (Malas, Dogan, Evcil, & Desdicioglu, 2006), and perhaps the development of sexual arousal is unaffected by androgens during this period. In a similar vein, primate research points to the possibility that exposure to prenatal androgens during different developmental time frames masculinizes sexual and non-sexual behaviors independently from each other (Goy et al., 1988). It is therefore possible that there are several "critical windows" of prenatal androgen exposure, with one masculinizing sexual response patterns and the other 2D:4D, and their expressions do not relate to each other later in life.

We further stress that the effectiveness of 2D:4D as a measure of prenatal androgenization in general has previously been called into question. However, a large part of this debate is specifically with regards to using 2D:4D as a predictor of sexual orientation, whereas the sex difference is generally considered to be highly robust and reliable (Bailey et al., 2016; Swift-Gallant et al., 2020). A meta-analysis found no evidence of publication bias in the literature on sex differences in 2D:4D (Grimbos et al., 2010). Moreover, our participant numbers substantially exceeded the sample size determined by the power analysis. We therefore believe that if there had been an interlink between sex differences in sexual arousal and in digit ratios, we should have detected it.

Additionally, we note that the sexual response patterns found in pupil dilation did not fully reflect those found in genital arousal, and that for one variable (pupil dilation to the less-arousing sex), the sex difference was in the opposite direction to the predicted (Figure 2.2B). While genital arousal is a well-evidenced measure of sexual arousal (Janssen, 2002; Suschinsky et al., 2009), pupil dilation is comparatively newer and has a smaller body of associated evidence. In some studies it has produced results which match with those found through genital arousal (Rieger et al., 2015), but in others it has produced results which are not fully identical to genital arousal patterns (Watts et al., 2018c). Because pupil dilation indicates not only sexual arousal, but also emotion, cognition, or non-sexual interest in stimuli (Bradley, Miccoli, Escrig, & Lang, 2008; Goldinger & Papesh, 2012), it is likely that for some participants, pupillary responses were driven by factors other than sexual interest. Despite this limitation, the present unexpected finding that men showed greater pupil dilation to their more-arousing sex than women is still in line with the general observation that men, unlike women, respond more strongly to one sex than the other.

In sum, the present study found evidence that men have more specific patterns of sexual response than women, and that men have more male-typical 2D:4D. However, we found no evidence that these two patterns were linked to one another. It may be the case that whichever factor contributes to sex differences in sexual arousal patterns is not reflected in 2D:4D.

Table 2.1 - *Multiple Regression Analyses for Sex and Left Hand 2D:4D Predicting Absolute Difference in Genital Arousal and Response to the Less-Arousing Sex (N = 294) and Pupil Dilation (N = 273).*

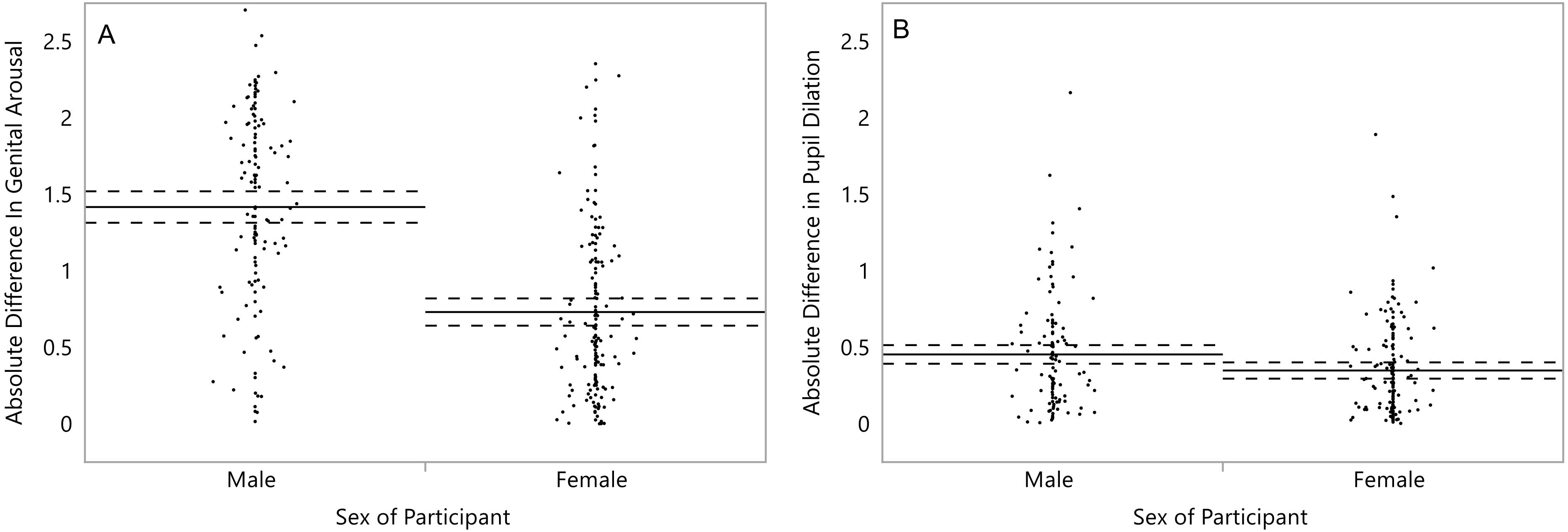
Step 1	**Absolute Difference in Genital Arousal**	**Genital Response to Less-Arousing Sex**	**Absolute Difference in Pupil Dilation**	**Pupil Dilation to Less-Arousing Sex**
Variables	β	β	β	β
Sex[1]	-.50 [-.60, -.40]**	.70 [.61, .78]**	-.15 [-.27, -.04]*	-.13 [-.25, -.01]*
Step 2	**Absolute Difference in Genital Arousal**	**Genital Response to Less-Arousing Sex**	**Absolute Difference in Pupil Dilation**	**Pupil Dilation to Less-Arousing Sex**
Variables	β	β	β	β
Sex[1]	-.51 [-.61, -.41]**	.69 [.61, .78]**	-.16 [-.28, -.04]*	-.13 [-.25, -.01]*
Left-Hand 2D:4D[2]	.07 [-.03, .17]	.02 [-.06, .11]	.02 [-.10, .14]	.03 [-.09, .15]

Note. R^2*'s* for the four models are .25, .49, .02 and .02 in Step 1, and .26, .49, .02 and .02 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Males were coded as 0, females as 1. [2]Lower scores indicate a more male-typical 2D:4D. * $p < .05$, ** $p < .01$.

Table 2.2 - *Multiple Regression Analyses for Sex and Right Hand 2D:4D Predicting Absolute Difference in Genital Arousal and Response to the Less-Arousing Sex (N = 293) and Pupil Dilation (N = 273).*

Step 1	**Absolute Difference in Genital Arousal**	**Genital Response to Less-Arousing Sex**	**Absolute Difference in Pupil Dilation**	**Pupil Dilation to Less-Arousing Sex**
Variables	β	β	β	β
Sex[1]	-.50 [-.60, -.40]**	.70 [.61, .78]**	-.15 [-.27, -.04]*	-.13 [-.25, -.01]*
Step 2	**Absolute Difference in Genital Arousal**	**Genital Response to Less-Arousing Sex**	**Absolute Difference in Pupil Dilation**	**Pupil Dilation to Less-Arousing Sex**
Variables	β	β	β	β
Sex[1]	-.52 [-.62, -.42]**	.70 [.61, .78]**	-.15 [-.27, -.03]*	-.13 [-.25, -.01]*
Right-Hand 2D:4D[2]	.10 [.00, .20] †	.01 [-.07, .09]	-.01 [-.13, .11]	.03 [-.09, .15]

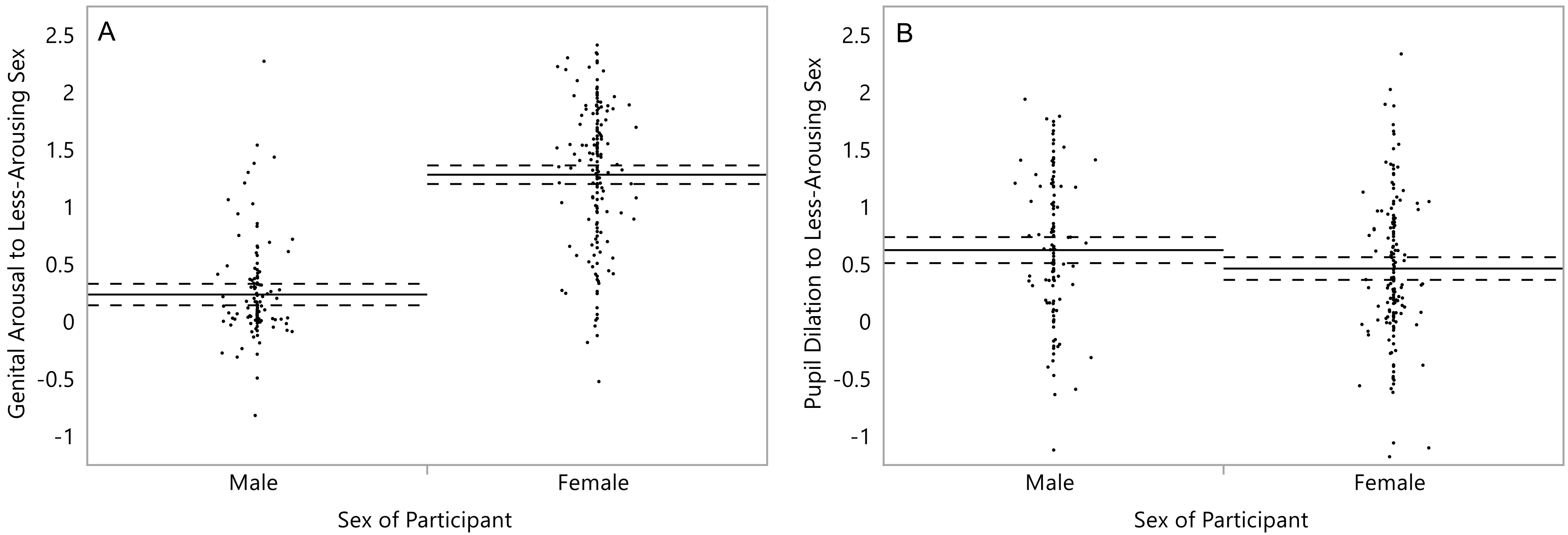
Note. R^2's for the four models are .25, .49, .02 and .02 in Step 1, and .26, .49, .02 and .02 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Males were coded as 0, females as 1. [2]Lower scores indicate a more male-typical 2D:4D. † $p < .10$, * $p < .05$, ** $p < .01$.



t = -9.93, *p* < .001, *d* = 1.17 [1.11, 1.24]

t = -2.56, *p* < .001, *d* = .31 [.27, .36]

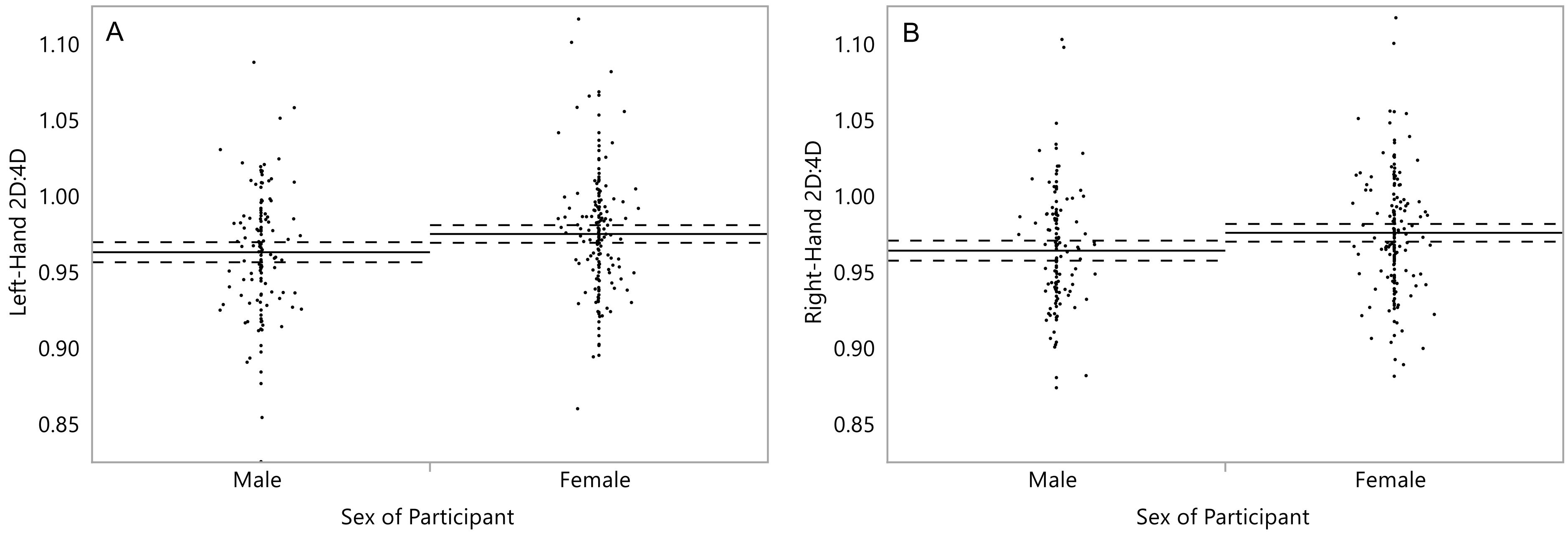
Figure 2.1 - Absolute difference in genital arousal and pupil dilation between stimuli featuring males and stimuli featuring females in (A) genital responses of 126 men and 168 women and (B) pupil dilation of 118 men and 155 women. On the Y axis, scores reflect the absolute difference between sexual arousal to males and females, standardized within participants. Solid lines represent group means, and dashed lines their 95% confidence intervals. Dots represent participants' average scores. Captions are independent-samples t-tests, with effect sizes and their 95% confidence intervals.



t = 16.60, *p* < .001, *d* = -1.96 [-2.02, -1.90]

t = -2.10, *p* = .03, *d* = .26 [.18, .33]

Figure 2.2 - Minimum genital arousal and pupil dilation to stimuli featuring males and stimuli featuring females (whichever is lowest) in (A) genital responses of 126 men and 168 women and (B) pupil dilation of 118 men and 155 women. On the Y axis, scores reflect minimum arousal values, standardized within participants. Solid lines represent group means, and dashed lines their 95% confidence intervals. Dots represent participants' average scores. Captions are independent-samples t-tests, with effect sizes and their 95% confidence intervals.



t = -2.68, *p* = .008, *d* = -.304 [-.309, -.300]

t = -2.62, *p* = .009, *d* = -.298 [-.302, -.294]

Figure 2.3 - Left-hand 2D:4D (A) of 139 men and 179 women, and right-hand 2D:4D (B) of 139 men and 178 women. On the Y axis, 2D:4D is the length of the index finger divided by the length of the ring finger. Solid lines represent group means, and dashed lines their 95% confidence intervals. Dots represent participants' scores. Captions are independent-samples t-tests, with effect sizes and their 95% confidence intervals.

Chapter 3

Sexual Orientation, Sexual Arousal, and Finger Length Ratios in Women

3.1 Abstract

In general, women show physiological sexual arousal to both sexes. However, compared with heterosexual women, homosexual women are more aroused to their preferred sex, a pattern typically found in men. We hypothesized that homosexual women's male-typical arousal is due to their sex-atypical masculinization during prenatal development. We measured the sexual responses of 199 women (including 67 homosexual women) via their genital arousal and pupil dilation to female and male sexual stimuli. Our main marker of masculinization was the ratio of the index to ring finger, which we expected to be lower (a masculine pattern) in homosexual women due to increased levels of prenatal androgens. We further measured observer- and self-ratings of psychological masculinity-femininity as possible proxies of prenatal androgenization. Homosexual women responded more strongly to female stimuli than male stimuli, and therefore had more male-typical sexual responses than heterosexual women. However, they did not have more male-typical digit ratios, even though this difference became stronger if analyses were restricted to white participants. Still, variation in women's digit ratios did not account for the link between their sexual orientation and their male-typical sexual responses. Furthermore, homosexual women reported and displayed more masculinity than heterosexual women, but their masculinity was not associated with their male-typical sexual arousal. Thus, women's sexual and behavioral traits, and potential anatomical traits, are possibly masculinized at different stages of gestation.

3.2 Introduction

The present research examined whether women's sexual arousal patterns, dependent on their sexual orientation, could be explained by differences in a putative marker of prenatal androgen exposure, which is the ratio of the second to fourth digit.

Women show, on average, substantial physiological sexual arousal to sexual stimuli featuring either females or males, regardless of their self-reported sexual orientation. Conversely, most men show substantial sexual arousal to their preferred sex but not to their non-preferred sex (Bailey, 2009). This sex difference appears to be robust, as it has been reported with several measures of sexual response including genital arousal (Chivers et al., 2004; Rieger et al., 2015; Rieger et al., 2016; Suschinsky et al., 2009), pupil dilation (Attard-Johnson et al., 2016; Rieger et al., 2015; Rieger & Savin-Williams, 2012), viewing time (Ebsworth & Lalumière, 2012; Israel & Strassberg, 2009), and neural responses (Safron et al., 2019; Sylva et al., 2013).

Substantial sexual arousal to both female and male sexual stimuli can therefore be described as female-typical, and substantial sexual arousal only to the preferred sex as male-typical (Chivers et al., 2007). Yet, this sex difference only applies on average and there are exceptions. One of these exceptions involves homosexual women. Although, like heterosexual women, homosexual women show arousal to both sexes, they also have greater sexual responses to their preferred sexual stimuli (females) compared to their non-preferred sex (males); this is the case when measuring their genital arousal (Chivers et al., 2004; Chivers et al., 2007; Rieger et al., 2016), pupil dilation (Rieger et al., 2015; Rieger & Savin-Williams, 2012), viewing time (Ebsworth & Lalumière, 2012; Lippa, 2012), and neural responses (Safron et al., 2017). Thus, the arousal patterns of homosexual women can be described as “male-typical,” as stronger responses to one’s preferred sex are usually found in men (Bailey, 2009). We stress, however, that although homosexual women show a preference for one sex, they still respond, on average, to stimuli featuring either sex. As such, their response patterns can only be considered male-typical in comparison to those of heterosexual women, and we predicted such a pattern in the present research (Hypothesis 1). Bisexual women are, on average, between heterosexual and homosexual women in their sexual responses, showing more of a preference for female stimuli than heterosexual women, but less of a preference than homosexual women (Rieger et al., 2016; Timmers, Bouchard, & Chivers, 2015).

It is possible that the arousal patterns of homosexual women are influenced by factors that typically influence sexual arousal in men. A prominent candidate is exposure to androgens during gestation. In mammals, exposure to prenatal androgens–specifically testosterone - accounts for the majority of sex differences in brain and behavior (Breedlove, 2017; Morris et al., 2004). In humans, our knowledge of the effects of prenatal androgen exposure is informed by genetic conditions which impact the availability of androgens or the individual’s sensitivity to them (Hines, 2009). One such condition is Congenital Adrenal Hyperplasia (CAH), which causes excessive production of androgens from gestation onwards. During their childhood, these genetic females are more likely to engage in male-typical play (Pasterski et al., 2005), and in adulthood they are more likely to express bisexual or homosexual attractions (Hines, Brook, & Conway, 2004; Meyer-Bahlburg, Dolezal, Baker, & New, 2008; Zucker, Bradley, Oliver, & Blake, 1996). Another relevant condition is Complete Androgen Insensitivity Syndrome (CAIS), which is an immunity to the effects of androgens at all stages of development. Genetic males with CAIS typically report sexual orientations towards males (Wisniewski et al., 2000), and show neural responses to both male and female sexual stimuli, and therefore have female-typical sexual responses (Hamann et al., 2014). Thus, levels of early androgen exposure may affect not only the development of general sex differences, but also the formation of sexual orientation and sexual arousal patterns within each sex (Bailey et al., 2016; Breedlove, 2010).

It is difficult to measure the level of androgens a human fetus is exposed to in utero, and research on the subject is often informed by postnatal biomarkers assumed to reflect prenatal androgen exposure. Of these, the most researched is the ratio of the length of the second digit to that of the fourth digit (2D:4D). There is a robust sex difference in this ratio; with men having a lower 2D:4D than women (Grimbos et al., 2010; Xu & Zheng, 2015). This sex difference emerges early in fetal development (Galis et al., 2010). Furthermore, females with CAH have lower (more male-typical) 2D:4D in the right hand than females without CAH (Brown et al., 2002b; Ciumas, Hirschberg, & Savic, 2008; Ökten, Kalyoncu, & Yariş, 2002), whereas genetic males with CAIS have 2D:4D comparable to those of typically-developed females (Berenbaum et al., 2009). Thus, these digit ratios may indeed reflect exposure to androgens during prenatal development.

Sexual orientation in women has repeatedly been linked to 2D:4D, with homosexual women having lower (more male-typical) 2D:4D than heterosexual women on average, (Kraemer et al., 2006; Putz et al., 2004; Rahman, 2005; Wallien et al., 2008; Watts et al., 2018a), and we predicted the same pattern in present research (Hypothesis 2). A meta-analysis of 34 independent samples totaling 5,828 participants confirmed that women with a nonheterosexual orientation had lower 2D:4D in the left and right hand, compared with heterosexual women, Hedge's *g's* = .23 and .29, .04 < 95% CIs < .51 (Grimbos et al., 2010). Unlike in women, variations in 2D:4D are not consistently linked to differences in sexual orientation in men, and digit ratios may only serve as a reliable biomarker of early androgen exposure with respect to the development of female sexual orientation (Swift-Gallant et al., 2020).

If homosexual women were indeed exposed to elevated levels of prenatal androgens compared with heterosexual women, as reflected in their male-typical digit ratios, it could mean that prenatal androgenization influenced both their sexual orientation towards women and their male-typical arousal patterns. For this reason, the putative marker of prenatal androgen exposure, their 2D:4D, could account for the relationship between women's sexual orientation and their sexual arousal patterns. Statistically this implies that 2D:4D would mediate the relationship of women's sexual orientation with their sexual arousal patterns (Hypothesis 3).

A different line of research links sexual orientation to masculinity and femininity, which can be defined as opposite poles of an encompassing psychological and behavioral trait (Lippa, 1991, 2005, 2008). Homosexual women are more masculine (and less feminine) than heterosexual women in their self-reported behaviors and interests (Lippa, 2005). Furthermore, others perceive homosexual women as more masculine than heterosexual women, based on the way they appear, sound, or move; and this observable difference emerges in their early childhood (Johnson, Gill, Reichman, & Tassinary, 2007; Rieger et al., 2010; Watts et al., 2018b). We predicted such differences in masculinity-femininity, dependent on women's sexual orientation, in the present work (Hypothesis 4).

These nonsexual psychological and behavioral traits of homosexual women could also be indicators of their masculinization, possibly by hormone exposure during the prenatal period (Bailey et al., 2016; Bailey & Zucker, 1995; Brown et al., 2002a). Measures of masculinity-femininity could therefore be used as a proxy of prenatal androgenization to explain their male-typical sexual arousal patterns. This hypothesis has been tested by assessing women's nonsexual traits via self-reported and observer-rated masculinity-femininity, and their sexual arousal via the measure of genital response and pupil dilation to female and male sexual stimuli. Homosexual women were more masculine in their nonsexual behaviors and self-reports and had more male-typical sexual responses than heterosexual women. Yet, these two findings were not interlinked; women's nonsexual masculinity did not mediate the relationship of sexual orientation with sexual arousal (Rieger et al., 2016). Perhaps male-typical behaviors and sexual arousal patterns in women develop independently of each other. Yet, there is also the possibility that a null finding was obtained by chance. Thus, in addition to testing whether women's sexual responses were related to their digit ratios, we re-examined whether behavioral masculinity of homosexual women could explain their male-typical sexual arousal patterns (Hypothesis 5).

Similar to their genital arousal patterns, bisexual women are intermediate between heterosexual and homosexual women in their masculinity-femininity (Lippa, 2005, 2008). Perhaps they are also intermediate with respect to their 2D:4D, although to our knowledge, this has not been explicitly studied previously. Bisexual women were part of the present research, but to ease interpretations, we focus in the following on comparisons between heterosexual and homosexual women, and we revisit bisexual women's responses in the Discussion.

We therefore tested the following hypotheses:

Hypothesis 1. Homosexual women are more male-typical than heterosexual women in their sexual arousal, with stronger arousal to their preferred sex than their non-preferred sex. In comparison, heterosexual women show similar arousal to both sexes.

Hypothesis 2. Homosexual women have, on average, more male-typical (lower) 2D:4D than heterosexual women.

Hypothesis 3. The relationship between sexual orientation and male-typical sexual arousal in women is mediated by a putative marker of androgen exposure, 2D:4D.

Hypothesis 4. Homosexual women are, on average, more masculine than heterosexual women in their nonsexual self-concepts and behaviors.

Hypothesis 5. The relationship between sexual orientation and male-typical sexual arousal in women is mediated by their masculine self-concepts and behaviors.

3.3 Method

3.3.1 Participants

Target participants. Our sample size was planned based on previous studies that used methodologies identical to ours and which computed the correlations of female sexual orientation with genital arousal or pupil dilation to male or female sexual stimuli, or of female sexual orientation with 2D:4D measures. The reported correlations were .20, .27 and .30, respectively (Rieger et al., 2016; Watts et al., 2018a). A power analysis conducted in G*Power determined that a sample size of 195 would be necessary for the smallest estimated effect (*r* = .20) to achieve significant results with a power of .80. A total of 199 women were recruited via Pride festivals, online news sites for lesbian women, and university mailing lists. Using a 7-point scale (Kinsey et al., 1953), women self-identified as “exclusively straight” (n = 44), “mostly straight” (n = 42), “bisexual leaning straight” (n = 15), “bisexual” (n = 18), “bisexual leaning lesbian” (n = 13), “mostly lesbian” (n = 26), or “exclusively lesbian” (n = 41). The mean (SD) age of the sample was 24.22 (6.98), and most were Caucasian (78%), followed by Chinese (5%), Black (4%), and other ethnicities. Eighteen of the 199 participants in the present study were identical twins whose data have been previously published (Watts et al., 2018a, 2018c). Since identical twins are comparable to non-twins in their sexual arousal patterns, and in the interest of maximizing statistical power in the present sample, we decided to include these individuals in the analyses.

Due to some participants opting out of the genital arousal component, others not responding to our messages to provide additional data (some participants were invited to provide 2D:4D data only after their original visit to the lab), and pupil data loss because of problems with the apparatus, genital arousal data were available for 184 women, pupil dilation data for 175 women, and 2D:4D data for 182 women.

Computations of multiple imputations (5 total) were conducted across all examined variables using linear regression analyses as the model type. The imputed data suggested that if below analyses were repeated with missing data being imputed, in order to have an entire set of 199 data points across all variables, it changed neither the direction, nor magnitude, nor significance of effects. Thus, we decided to focus on analyses with actual, observed data, even if that meant that participant numbers varied across analyses. The exact number of participants included in each stage of analysis can be found in the captions of the corresponding tables and figures, and a full listing is given in Table 3.1.

3.3.2 Measures and Materials

Self-reported sexual orientation. Participants reported both their sexual orientation and sexual attraction to men and women on 7-point scales (Kinsey et al., 1953). These scales were highly correlated, $p < .0001$, $r = .97$, 95% CI [.95, .97], and averaged within participants. For this average, a score of 0 represented exclusive heterosexuality, a score of 3 bisexuality with equal attractions to women and men, and 6 represented exclusive homosexuality. This composite score was used for all analyses. Note that, although we focused on differences between heterosexual (Kinsey 0-1) and homosexual (Kinsey 5-6) women, data from bisexual (Kinsey 2-4) women were available, and the sexual orientation variable was always treated as continuous except for descriptive purposes in Table 3.1.

Self-reported masculinity-femininity. Childhood masculinity was assessed using six items from the Childhood Gender Nonconformity Scale, and adulthood behavioral masculinity was measured using six items from the Continuous Gender Identity Scale. These scales produce sexual orientation differences in masculinity-femininity in the predicted directions (Rieger et al., 2008; Rieger et al., 2010).

However, some of the items used outdated wording or were targeted towards an US (rather than UK) sample. We therefore removed one item of the original childhood scale and completely reworded items of the adulthood scale (see Appendix for a full list of items). Participants responded to statements on 7-point scales ranging from 1 (strongly disagree) to 7 (strongly agree). Answers were scored such that higher numbers represented greater masculinity. Item reliability (Cronbach's alpha) was .89 for the Childhood Gender Nonconformity Scale and .93 for the Continuous Gender Identity Scale. Because the adulthood scale was entirely reworded, we conducted further checks on its psychometric properties. It correlated with sexual orientation (r = .36), self-report from childhood (r = .67), and observer ratings from adulthood (r = .48) in the same manner, if not slightly better, than the corresponding correlations across heterosexual and homosexual women in research using the previous version (Rieger et al, 2008, 2010), lending support to its convergent validity.

2D:4D. Digit measurements were taken from either high-resolution photographs or scans of participants' hands, similar to past work (Allaway et al., 2009; Watts et al., 2018a). For the photographs, participants placed their hands on a flat surface with their fingers slightly spread apart, and images were taken from approximately 30cm above this surface. For the scans, participants placed their hands flat on the surface of the scanner. Different methods of capturing images (photograph or scanner) did not moderate the relationship between sexual orientation and 2D:4D.

From the resulting images of hands, digit ratios were measured by two independent raters who were blind to the sex and sexual orientation of the participants. Raters used the open-source vector graphics package Inkscape 0.92, as computer-assisted measurement techniques produce highly reliable measurements (Watts et al., 2018a). Each rater drew a line as wide as the finger along the lowest crease at the base of the finger, between the metacarpal and proximal phalanx. A second line was then drawn from the tip of the finger down towards the base, where it automatically snapped to the center of the base line. Fine adjustments were then made at a higher level of zoom, to ensure that this line matched the tip of the finger as closely as possible. Inter-rater reliability (Cronbach's alpha) exceeded .99 for each digit. Therefore, the measurements for each digit were averaged between raters. Finally, 2D:4D was calculated by dividing the averaged length of the index finger by the averaged length of the ring finger for each hand of each participant.

Stimuli. The sexual stimuli consisted of 3-minute videos, three featuring a female model and three featuring a male model, each of them masturbating in a bedroom. These stimuli were selected in a previous study in which 200 videos were rated on their sexual appeal by men and women of different sexual orientations (Rieger et al., 2015), and the top three female and male videos were used in the present study. Neutral stimuli to assess baseline genital responses were 2-minute clips taken from a nature documentary. Their engaging but nonsexual content facilitated participants' return to an unaroused level. However, these nature videos were not used for pupil dilation baseline, as their engaging content might elicit dilation for reasons other than sexual arousal. Thus, two 1-minute animations of clouds were used to obtain a pupillary baseline. All videos were edited using MPEG Streamclip and Final Cut Pro to be of similar luminance.

Genital arousal. Genital arousal was measured as changes in peak-to-trough vaginal pulse amplitude (VPA) using a vaginal photoplethysmograph. The signal was recorded using a BIOPAC MP150 data acquisition unit, sampled at 200 Hz, and high-pass filtered at 0.5 Hz with 16-bit resolution. The VPA exhibits both convergent and discriminant validity for the measurement of female sexual response (Suschinsky et al., 2009).

Pupil dilation. Pupil dilation data was measured with a SR Research EyeLink 1000 infrared eye tracking unit. A 35 mm lens focused on the participants' right eye, positioned approximately 60 cm from the participants' head, and sampling at a rate of 500 Hz. The infrared light emitted by the eye tracker is reflected by the pupil, and the number of pixels reflected were recorded. Because raw pupil area data included "0's" for missing values, for instance from blinks or head movements, these values were removed prior to analyses. Pupil dilation data gathered and processed in this manner has previously produced similar patterns to those found in genital arousal in both men and women (Rieger et al., 2015).

3.3.3 Procedure

Participant session. After giving written informed consent, participants completed a survey on their demographics, sexual orientation, and masculinity-femininity, and had photographs or scans of their hands taken. They were then seated in a chair and had their entire body video-recorded for 5-10 minutes to capture their gestures and movements. Participants answered questions about the weather and their interests and were not interrupted while answering, nor were participants informed that these videos would be used to assess their masculinity-femininity, but rather were told that they would be rated for "measures of psychological interest." For our observer ratings of masculinity-femininity we used their answer to a neutral question: "How would you describe the weather at this time of year?"

Participants were then seated in a sealed booth, with dim lighting conditions. Eyes were calibrated by participants fixating on dots outlining the screen. They were then instructed on how to use the genital measure, which they inserted in privacy after the experimenter left the booth (and at which point both their eye data and genital data were checked remotely). Pupil dilation and genital arousal were measured simultaneously throughout the experiment. Participants were instructed to watch the screen throughout the experiment, regardless of the content. They first viewed an animation of clouds, followed by alternating sexual and nature videos. These were displayed in a random order, but a sexual video was always followed by a nature video. After each nature video, the experiment displayed a grey screen of similar luminance to the videos until the experimenter verified that participants had returned to baseline arousal for a minimum of 5 seconds. Following this, the next erotic video was shown. After the sixth nature video, a final animation of clouds was displayed. Participants were paid £50. The entire procedure took approximately 90 minutes.

For each participant, genital data and pupil data were averaged across the duration of each stimulus. These averages were then standardized within participants, producing a z-score for each participant and stimulus. For genital data, standardized responses to the 5 seconds preceding each sexual stimulus (following the display of a neutral stimulus, and after the participant had returned to baseline) were subtracted from the standardized response to the sexual stimulus. For pupil data, standardized responses to neutral stimuli (the animated clouds) were subtracted from standardized responses to all sexual stimuli. We then computed, for each participant, average responses across all sexual stimuli of a given type (female or male), which reflected their responses to each sex as compared to baseline. These scores were then used to calculate a contrast score representing their response to females over males, such that a positive score indicated a preference for females, a negative score indicated a preference for males, and a score of zero indicated equal preferences. Caution must be taken when producing such a contrast score from genital arousal scores of heterosexual women, because this can lead to the averaging of responses which are individually very different, which produces the illusion of a non-specific response on a group level (Lalumière et al., 2020). We therefore checked whether this was the case in the present sample. The distribution of genital scores for heterosexual women were normally distributed and centered on zero, and their mean was not significantly different from zero. As such, we are confident that averaging the genital arousal scores of heterosexual women does not distort the pattern of the data in the present sample.

Editing of participant videos. Recordings of participants' answers to our question about the weather were edited in Shotcut and used for analyses. We selected the first sentence that the participants articulated within the first 20 seconds of their answer. If responses were less than 6 seconds, we took a combination of their first and second sentence. The majority of selected videos were approximately 10 seconds long, and all clips included audio. Raters can reliably judge behavioral traits associated with sexual orientation from brief video clips such as these (Tskhay & Rule, 2013).

Ratings of masculinity-femininity. Psychology students participated as raters of masculinity-femininity for course credit, and each video-recorded target was evaluated by a minimum of 21 and a maximum of 46 raters. In total we had 48 heterosexual male raters, 21 non-heterosexual male raters, 71 heterosexual female raters, and 29 non-heterosexual female raters. Videos were rated in batches of 20-30 to avoid rater fatigue, raters from each rater group were randomly assigned to a batch,

Raters were blind to the participants' sexual orientation. They were not trained in how to rate, but instructed to indicate their impression of each woman's appearance and demeanor, in comparison with other women of the same age. For example, they were told to "rate whether this woman appeared or behaved in a more feminine or masculine way." Ratings were completed on 7-point scales, where a score of 1 was "more feminine," 4 "average," and 7 "more masculine." Heterosexual raters tended to give higher scores than non-heterosexual raters [Mean (SD): heterosexual males: 3.05 (1.04); heterosexual females: 3.29 (1.17); non-heterosexual males: 2.59 (1.09); non-heterosexual females: 2.89 (1.26)], but correlations between rater groups (heterosexual and non-heterosexual men and women) ranged from r = .68 to r = .81, and all relationships were positive. Additionally, ratings were highly reliable within each rater group and across all raters (all Cronbach's α's > .95). Evaluations were therefore averaged across all raters, producing an average observer-rated masculinity-femininity score for each video-recorded participant.

3.4 Results

3.4.1 Initial Analyses.

Although we treat sexual orientation as a continuous variable in main analyses, we first present a summary of our key variables with participants grouped according to their scores on the Kinsey scale, with Kinsey 0-1 considered heterosexual, 2-4 considered bisexual, and 5-6 considered homosexual (Table 3.1). Significance values are also given, using heterosexuals as the comparison group. Unexpectedly, bisexual women as a group had more feminine 2D:4D ratios than both heterosexual and homosexual women. This was not predicted, and we revisit it in the Discussion. Also note that on average, heterosexual and homosexual women did not differ in their digit ratios.

3.4.2 Hypothesis 1.

We hypothesized that homosexual women would be more sexually aroused to stimuli featuring females, whereas heterosexual women would show similar arousal to both sexes. We regressed women's responses to sexual stimuli onto their sexual orientation. For each measure of sexual arousal (genital arousal or pupil dilation) we had three dependent variables: their responses to females, their responses to males, and their responses to females over males. We originally tested for both a linear and curvilinear effect of sexual orientation on women's sexual responses, to account for the possibility that differences between heterosexual, bisexual, and homosexual women may not always follow a simple linear trend (Rieger et al., 2016). However, in the present data, the curvilinear effects did not explain arousal over and above linear effects. For the sake of simplicity, in the following we focus on analyses in which only linear effects of sexual orientation were tested.

We first regressed women's genital arousal to female stimuli and male stimuli onto their sexual orientation. Homosexual women (Kinsey Scores 5-6) did not respond significantly more to females as compared with heterosexual women (Kinsey Scores 0-1), p = .45, $β$ = .06, 95% CI [-.09, .20] (Figure 3.1A), but did respond significantly less to males, p = .03, $β$ = -.16 [-.31, -.02] (Figure 3.1B). We then regressed women's genital arousal contrast of females over males onto their sexual orientation. Homosexual women responded significantly more to their preferred sex (females) than their non-preferred sex (males) as compared with heterosexual women, p = .02, $β$ = .17 [.03, .31] (Figure 3.1C).

Independent of these patterns, women, in general, responded significantly to both females and males as compared to baseline. That is, across sexual orientations, the confidence intervals of the regression coefficients were above 0 (Figures 3.1A & 3.1B).

We then repeated the above analyses with pupil dilation to sexual stimuli as the dependent variable. Results were similar to those for genital arousal. Homosexual women did not respond significantly more to females as compared with heterosexual women, p = .37, $β$ = -.07 [-.22, .08] (Figure 3.2A), but responded significantly less to males as compared with heterosexual women, p = .03, $β$ = -.17 [-.31, -.02] (Figure 3.2B). We then regressed women's pupil dilation to females over males onto their sexual orientation. Homosexual women responded significantly more to females than males as compared with heterosexual women, p = .046, $β$ = .15 [.00, .30] (Figure 3.2C). As with genital arousal, women of all sexual orientations responded significantly to both female and male sexual stimuli as compared to baseline (i.e. the confidence intervals of the effects were above 0; Figures 3.2A & 3.2B).

3.4.3 Hypothesis 2.

We hypothesized that homosexual women have more male-typical (lower) 2D:4D than heterosexual women. We regressed women's left hand and right hand 2D:4D onto their sexual orientation. Homosexual women did not have a significantly lower 2D:4D than heterosexual women in either their left hand, p = .29, β = -.08 [-.23, .07], or their right-hand, p = .65, β = -.03 [-.18, .12] (Figures 3.3A & 3.3B).

As aforementioned, Table 3.1 shows that bisexual women had, unexpectedly, higher 2D:4D than both heterosexual and homosexual women. However, even when excluding bisexual women, homosexual women did not have more male-typical ratios than heterosexual women (this can also be seen in Table 3.1), plus the pattern of subsequent mediation analyses remained identical, with or without bisexual women excluded. All further reported results are therefore from analyses in which bisexual women were included.

Furthermore, there is evidence from one study that ethnicity can influence the relationship of women's sexual orientation with their 2D:4D (Lippa, 2003), even though there is no evidence for this from a meta-analysis (Grimbos et al., 2010). Still, we repeated the analysis for only the 136 Caucasian participants, as they formed the majority of our sample. Despite the reduced sample size, the relationship between 2D:4D and sexual orientation was closer to significance in both the left hand, p = .06, β = -.16 [-.33, .01], and their right-hand, p = .16, β = -.12 [-.29, .05] (Figures 3.4A & 3.4B). However, the inclusion or exclusion of non-white ethnicities did not affect the patterns of subsequent analyses mediation analyses, and we therefore kept non-white ethnicities included.

3.4.4 Hypothesis 3.

We hypothesized that the relationship between sexual orientation and male-typical sexual arousal in women is mediated by their male-typical 2D:4D. Although the relationship between 2D:4D and sexual orientation was not significant, we conducted this analysis regardless, as it was planned in advance. We computed multiple regression analyses predicting the genital arousal or pupil dilation contrast (responses to females over males) by sexual orientation and digit ratios. We focused on the contrast score since it tended to show, across the two measures of sexual response, a somewhat stronger relation with sexual orientation than responses to females or responses to males (Figures 3.1 - 3.2).

In Step 1, sexual orientation was the only predictor of sexual response. In Step 2, we included a 2D:4D variable–either left or right hand - as a mediator. If our hypothesis were confirmed, then the inclusion of 2D:4D as a mediator should weaken the relationship between sexual orientation and either measure of sexual arousal. Tables 3.2 and 3.3 summarize the results of the analyses both for genital arousal and for pupil dilation, for Step 1 (without 2D:4D as mediator) and Step 2 (with 2D:4D as mediator). The effect of sexual orientation on sexual arousal remained similar (or slightly increased) after including 2D:4D as a predictor. Furthermore, a mediation analysis on the basis of 10,000 bootstrapped samples (Preacher & Hayes, 2008) did not indicate that left-hand 2D:4D mediated the relationship between sexual orientation and sexual responses, because the confidence intervals of the estimated mediation effects included zero. This was the case for left-hand 2D:4D predicting genital arousal, β = -.01 [-.04, .01], and pupil dilation β = .007 [-.01, .04]. It was also the case for right-hand 2D:4D predicting genital arousal β = -.005 [-.03, .02], and pupil dilation β = -.003 [-.02, .02].

3.4.5 Hypothesis 4.

We hypothesized that homosexual women were more masculine and less feminine in their self-report and behaviors than heterosexual women. We regressed women's self-reported adulthood and childhood masculinity-femininity, in addition to observer ratings of their masculinity-femininity, onto their sexual orientation. Homosexual women were significantly more masculine than heterosexual women in their self-reports of childhood, $p = .001$, $\beta = .23$ [.09, .37], and adulthood, $p < .001$, $\beta = .31$ [.17, .44], and when rated by others, $p < .001$, $\beta = .38$ [.25, .51] (Figures 3.5A - 3.5C).

3.4.6 Hypothesis 5.

We hypothesized that the effect of sexual orientation on male-typical sexual arousal was mediated by homosexual women's higher levels of masculinity. We built a total of 12 regression models. In each of these models, the dependent variable was either genital arousal or pupil dilation to females over males. The independent variables were sexual orientation and one measure of masculinity-femininity: self-report from childhood or adulthood, or observer ratings from adulthood. If male-typical sexual responses in homosexual women were mediated by their masculinity, then the inclusion of a measure of masculinity should weaken the relationship of sexual orientation with sexual arousal.

Tables 3.3-3.5 summarize the results of these analyses. In general, sexual orientation effects on sexual response did not decrease (and if anything, increased) with a measure of masculinity-femininity as a covariate. Likewise, mediation analyses on the basis of 10,000 bootstrapped samples (Preacher & Hayes, 2008) did not suggest mediation by any measure of masculinity on either measure of sexual response.

3.5 Discussion

The present data confirmed that homosexual women had more male-typical sexual arousal patterns than heterosexual women, as indicated by both their genital arousal and their pupil dilation. However, there was no evidence that they had more male-typical digit ratios, or that digit ratios mediated the relationship between women's sexual orientation and their male-typical sexual arousal patterns. Moreover, even though homosexual women were more masculine than heterosexual women in their self-reports or via observer ratings, this pattern, too, did not explain their male-typical arousal patterns.

The finding that homosexual women had stronger responses to their preferred sex than heterosexual women is consistent with previous research both for genital arousal and pupil dilation (Chivers et al., 2004; Chivers et al., 2007; Rieger et al., 2015; Rieger et al., 2016). However, the finding that 2D:4D was not significantly lower in homosexual women than heterosexual women is puzzling, as it was confirmed previously in a meta-analysis (Grimbos et al., 2010). This may have been due to methodological reasons: Although between-rater reliability was high, and computer-assisted measurement techniques, such as those employed in the current study, have been shown to have the highest reliability compared to other methods of measuring 2D:4D (Allaway et al., 2009), we cannot say with certainty that our measure was valid.

Indeed, there is an ongoing debate about the utility of 2D:4D: Although it is regarded as a valid measure with respect to sex differences and female sexual orientation differences, it is also the case that there is much variability in this measure across individuals, and findings only apply on aggregate and do not apply to single people (Swift-Gallant et al., 2020). Furthermore, the aforementioned meta-analysis suggested a publication bias in reported relationships of sexual orientation with 2D:4D

(Grimbos et al., 2010), and the true effect could therefore be smaller than usually published. In the present research, the strongest effect of sexual orientation on 2D:4D was r (or β) = -.12 in the right hand. With this effect, post-hoc power analyses indicated a minimum sample of 542 women of different sexual orientations for it to become significant. If our a-priori sample size estimate had returned such a large number, we would have considered it an unreasonable goal for a lab-based study like ours.

Another possible explanation for the present null finding with respect to 2D:4D is the ethnic makeup of the sample. We did not factor this into planning the present study because the meta-analysis pointed to an ethnicity effect only in men and not in women (Grimbos et al., 2010), although other research has found an influence of ethnicity on 2D:4D in women (Lippa, 2003). Indeed, excluding all non-Caucasian participants from the present sample made the association between 2D:4D and sexual orientation stronger (although still non-significant) in both hands (Figure 3.4). Thus, future research measuring the relationship between 2D:4D and sexual orientation may wish to either employ a racially homogenous participant sample, or recruit enough participants that per-race comparisons are feasible. Note that even within the white sample, 2D:4D did not appear to explain (mediate) any relationship of women's sexual orientation with their sexual response patterns.

Given this lack of sexual orientation difference in 2D:4D, it is unsurprising that the present study found no evidence of a mediating role of 2D:4D. Hence, it is impossible to draw any conclusions from the present data about whether the relationship between 2D:4D and sexual orientation mediates the relationship between sexual orientation and sexual responses, simply because 2D:4D in itself did not relate to sexual orientation. With regards to masculinity-femininity, if anything, statistically controlling for any of the three masculinity-femininity variables made the correspondence of women’s sexual orientation with their male-typical sexual arousal stronger. This pattern–a strengthening of the effect of sexual orientation on sexual response when measures of behavioral masculinity are statistically controlled for–has been previously noted (Rieger et al., 2016). In combination with present findings, it appears unlikely that it was previously a chance finding.

Another methodological limitation of the current study is missing data. Due to participants opting out of certain study components, not replying to follow-up messages, as well as data loss, we were left with incomplete data sets for a number of women from our sample, and the sample size for each analysis varied as a result. Thus, data were not missing at random and findings have look different for a sample of women who would have given data across all required variables.

If one assumed for a moment that the present findings are accurate, what could be their reasons? For females it is possible that there exist several “sensitive periods” of masculinization during prenatal development, and that these periods differ for different traits (McCarthy, Herold, & Stockman, 2018; Xu et al., 2019). At least in non-human primates, exposure to testosterone at different stages of gestation may masculinize sexual behaviors independently from nonsexual behaviors (Goy et al., 1988). Specifically, Goy et al. (1988) reported that female rhesus macaques exposed to testosterone during their prenatal development had different behavioral outcomes depending on the timing, with those exposed early in gestation displaying male-typical sexual behaviors (e.g., mounting other females) and those exposed late in gestation displaying male-typical nonsexual behaviors (e.g., rough play). It is possible that behavioral traits and sexual arousal patterns are masculinized at different stages of development in humans also, and thus, are not necessarily interlinked within individuals–for example, those who have male-typical arousal may not have male-typical gender-related behaviors and vice versa.

A final point concerns bisexual women, who appeared between heterosexual and homosexual women in their sexual arousal and masculinity-femininity. An hypothesis is that due to intermediate dosages of genetic or prenatal hormonal influences, bisexual individuals, who could be considered to have sexual orientations between heterosexual and homosexual, also fall intermediate with respect to correlates of sexual orientation (Rieger et al., 2020a). Thus, regarding bisexual women's 2D:4D, we assumed that they could also be intermediate between heterosexual and homosexual women on this measure. Contrary to this assumption, bisexual women had more feminine 2D:4D than both heterosexual and homosexual women (Table 3.1). It has been proposed that personality differences between homosexual and heterosexual women may be caused by exposure to androgens during prenatal development, whereas the distinct personality traits of bisexual individuals (e.g. higher sociosexuality compared to heterosexual and homosexual) may be a correlate of their higher levels of post-natal androgens (Lippa, 2020). If the present findings are valid, they would suggest that bisexual women also differ from heterosexual and homosexual women with respect to prenatal androgenization, but this would imply that they have been less masculinized than other groups, and we cannot offer an explanation for why this would be the case.

In sum, the findings of the present research suggest that there is no link between the male-typical sexual responses of homosexual women and putative markers of prenatal androgenization. Of course, other purported markers of androgen exposure may reveal a different pattern than the one reported here. Such markers include the distance between the anus and the genitalia (Barrett et al., 2018) and otoacoustic emissions, which are tiny sounds emitted by the inner ear (McFadden & Pasanen, 1998). Another avenue for future research would involve individuals with conditions affecting the availability of androgens, or their sensitivity to them. To our knowledge, no studies to date investigated the arousal patterns of women with CAH. If androgen exposure does indeed impact sexual responses–and given the apparent impact of excessive androgens on the sexual orientation of women with CAH (Meyer-Bahlburg et al., 2008; Zucker et al., 1996)–women with CAH may show male-typical specificity in their sexual arousal.

Table 3.1 - *Means, Confidence Intervals, Standard Deviations and Sample Sizes for Variables, Split by Sexual Orientation Groups.*

	Genital Arousal to Females over Males	Pupil Dilation to Females over Males	Right-Hand 2D:4D	Left-Hand 2D:4D	Self-Reported Childhood Masculinity	Self-Reported Adulthood Masculinity	Observer-Rated Adulthood Masculinity
Heterosexual (Kinsey 0-1)	.07 [-.09, .22] (.72, N = 82)	.15 [-.01, .30] (.61, N = 64)	.975 [.965, .985] (.044, N = 83)	.973 [.965, .981] (.037, N = 82)	2.99 [2.63, 3.34] (1.65, N = 85)	2.31 [2.01, 2.61] (1.41, N = 85)	2.78 [2.59, 2.96] (.86, N = 85)
Bisexual (Kinsey 2-4)	.40 [.16, .62]* (.73, N = 42)	.20 [.02, .38] (.60, N = 46)	.990 [.977, 1.00]* (.040, N = 38)	.990 [.974, 1.00]* (.047, N = 38)	3.63 [3.17, 4.10]* (1.81, N = 45)	2.73 [2.35, 3.11] (1.26, N = 45)	3.11 [2.86, 3.36]† (.81, N = 43)
Homosexual (Kinsey 5-6)	.33 [.11, .55]* (.84, N = 60)	.36 [.18, .53]† (.71, N = 65)	.967 [.959, .974] (.029, N = 62)	.970 [.961, .979] (.035, N = 62)	3.87 [3.42, 4.31]* (1.80, N = 66)	3.41 [2.98, 3.84]* (1.75, N = 66)	3.67 [3.34, 4.00]* (1.30, N = 63)

Note. Numbers in square brackets represent 95% confidence intervals of the mean. Numbers in round brackets represent standard deviations of the mean and sample sizes. Participants were grouped according to their scores on the Kinsey scale, with Kinsey 0-1 considered heterosexual, 2-4 considered bisexual and 5-6 considered homosexual. Asterixis indicate significant difference to heterosexual, † $p < .10$, * $p < .05$.

Table 3.2 - Multiple Regression Analyses for Sexual Orientation and Left Hand 2D:4D predicting Genital Arousal (Step 1 N = 184, Step 2 N = 174) and Pupil Dilation (Step 1 N = 175, Step 2 N = 160) to Females over Males.

Step 1	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.17 [.03, .31]*	.15 [.00, .30]*
Step 2	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.20 [.05, .35]**	.14 [-.02, .30]†
Left Hand 2D:4D[2]	.13 [-.02, .28]†	-.05 [-.21, .11]

Note. R^2*'s* for the two models are .03 and .02 in Step 1, and .05 and .02 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Higher scores indicate a more homosexual orientation. [2]Lower scores indicate more male-typical 2D:4D. † $p < .10$, * $p < .05$, ** $p < .01$.

Table 3.3 - Multiple Regression Analyses for Sexual Orientation and Right Hand 2D:4D predicting Genital Arousal (Step 1 N = 184, Step 2 N = 173) and Pupil Dilation (Step 1 N = 175, Step 2 N = 160) to Females over Males.

<u>Step 1</u>	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.17 [.03, .31]*	.15 [.00, .30]*
<u>Step 2</u>	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.19 [.04, .33]*	.15 [.00, .31]†
Right Hand 2D:4D[2]	.15 [.01, .30]*	.03 [-.13, .19]

Note. R^2*'s* for the two models are .03 and .02 in Step 1, and .06 and .02 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Higher scores indicate a more homosexual orientation. [2]Lower scores indicate more male-typical 2D:4D. † $p < .10$, * $p < .05$, ** $p < .01$.

Table 3.4 - *Multiple Regression Analyses for Sexual Orientation and Self-Reported Childhood Masculinity predicting Genital Arousal (Step 1 N = 184, Step 2 N = 181) and Pupil Dilation (Step 1 N = 175, Step 2 N = 172) to Females over Males.*

<u>Step 1</u>	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.17 [.03, .31]*	.15 [.00, .30]*
<u>Step 2</u>	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.18 [.03, .33]*	.18 [.02, .33]*
Self-Reported Childhood Masculinity[2]	.00 [-.15, .15]	-.11 [-.26, .05]

Note. R^2*'s* for the two models are .03 and .02 in Step 1, and .03 and .03 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Higher scores indicate a more homosexual orientation. [2]Higher scores indicate higher self-reported childhood behavioral masculinity. † $p < .10$, * $p < .05$, ** $p < .01$.

Table 3.5 - *Multiple Regression Analyses for Sexual Orientation and Self-Reported Adulthood Masculinity predicting Genital Arousal (Step 1 N = 184, Step 2 N = 181) and Pupil Dilation (Step 1 N = 175, Step 2 N = 172) to Females over Males.*

Step 1	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.17 [.03, .31]*	.15 [.00, .30]*

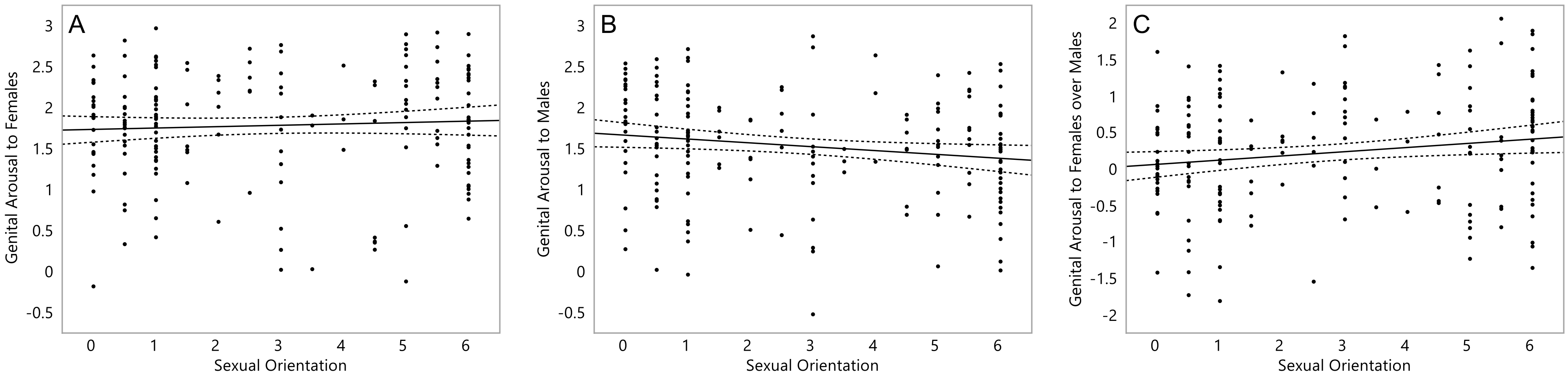
Step 2	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.19 [.04, .35]*	.19 [.03, .35]*
Self-Reported Adulthood Masculinity[2]	-.05 [-.20, .10]	-.12 [-.28, .04]

Note. R^2*'s* for the two models are .03 and .02 in Step 1, and .03 and .03 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Higher scores indicate a more homosexual orientation. [2]Higher scores indicate higher self-reported adulthood behavioral masculinity. $^\dagger p < .10$, * $p < .05$, ** $p < .01$.

Table 3.6 - *Multiple Regression Analyses for Sexual Orientation and Video Observer-Rated Adulthood Masculinity predicting Genital Arousal (Step 1 N = 184, Step 2 N = 180) and Pupil Dilation (Step 1 N = 175, Step 2 N = 167) to Females over Males.*

Step 1	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.17 [.03, .31]*	.15 [.00, .30]*
Step 2	**Genital Arousal to Females over Males**	**Pupil Dilation to Females over Males**
Variables	β	β
Sexual Orientation (SO)[1]	.17 [.01, .33]*	.15 [-.02, .31]
Observer-Rated Masculinity[2]	-.04 [-.19, .12]	-.01 [-.17, .16]

Note. R^2*'s* for the two models are .03 and .02 in Step 1, and .03 and .02 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Higher scores indicate a more homosexual orientation. [2]Higher scores indicate higher observer-rated masculinity. † $p < .1$, * $p < .05$, ** $p < .01$.

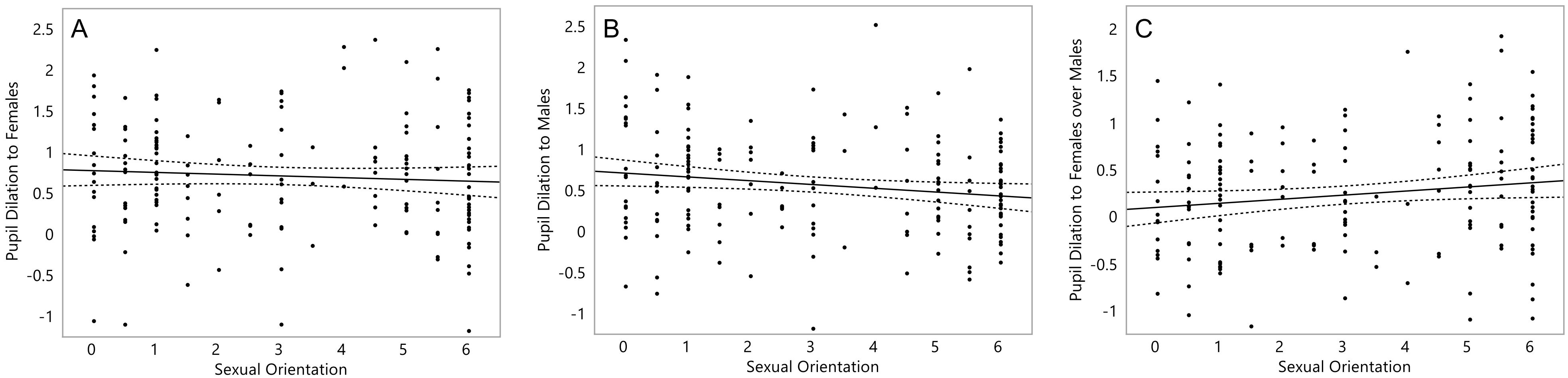


p = .45, β = .06 [-.09, .20]

p = .03, β = -.16 [-.31, -.02]

p = .02, β = .17 [.03, .31]

Figure 3.1 - Women's genital responses to sexual stimuli. Genital arousal of 184 women in response to stimuli featuring females (A), males (B), and females over males (C). On the Y axis, genital arousal scores reflect changes compared to the 5 seconds preceding a sexual stimulus, standardized within participants. On the X axis, 0 represents exclusive heterosexuality, 3 bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants' average scores.



p = .37, *β* = -.07 [-.22, .08]   *p* = .03, *β* = -.17 [-.31, -.02]   *p* = .046, *β* = .15 [.00, .30]

Figure 3.2 - Women's pupil dilation to sexual stimuli. Pupil dilation of 175 women in response to stimuli featuring females (A), males (B) and females over males (C). On the Y axis, pupil dilation scores reflect changes compared to the neutral stimuli, standardized within participants. On the X axis, 0 represents exclusive heterosexuality, 3 represents bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants' average scores.

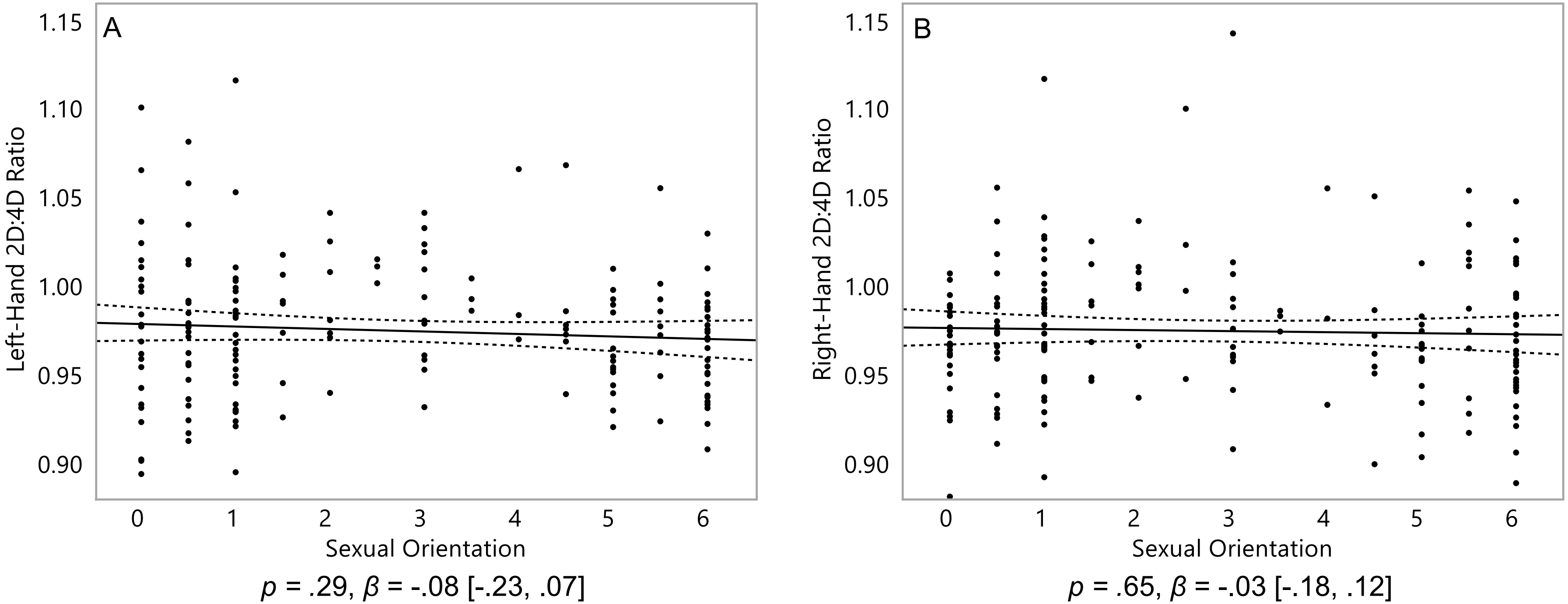


Figure 3.3 - The left hand 2D:4D of 183 women (A) and the right hand 2D:4D of 182 women (B). On the Y axis, 2D:4D is the length of the index finger divided by the length of the ring finger. On the X axis, 0 represents exclusive heterosexuality, 3 represents bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants' scores.

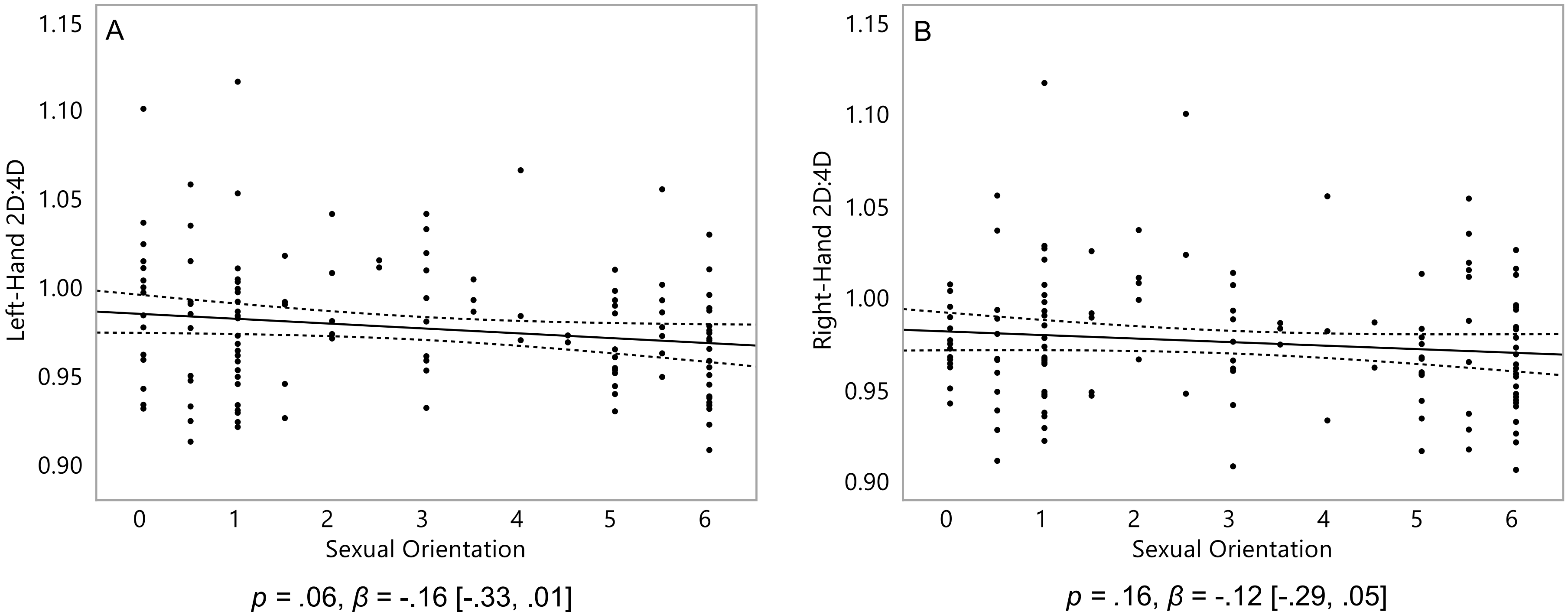
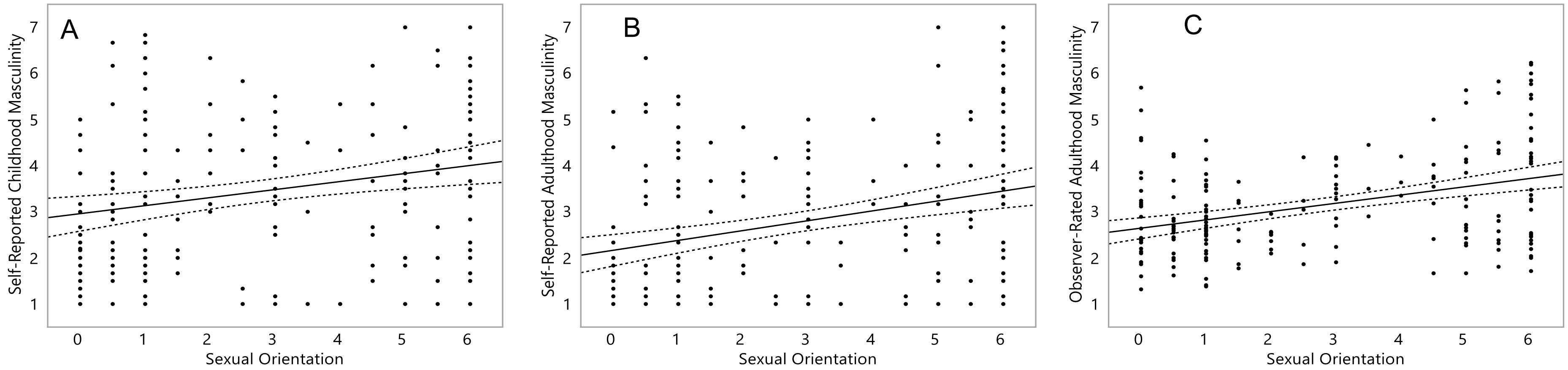


Figure 3.4 - The left hand 2D:4D of 136 Caucasian women (A) and the right hand 2D:4D of 135 Caucasian women (B). On the Y axis, 2D:4D is the length of the index finger divided by the length of the ring finger. On the X axis, 0 represents exclusive heterosexuality, 3 represents bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants' scores.



p = .001, *β* = .23 [.09, .37]   *p* < .001, *β* = .31 [.17, .44]   *p* < .001, *β* = .38 [.25, .51]

Figure 3.5 - The childhood and adulthood self-report and observer-rated behavioral masculinity data of 192 women (self-report) and 191 women (observer-ratings). On the Y axis, behavioral masculinity scores, with higher scores representing a greater degree of masculinity. On the X axis, 0 represents exclusive heterosexuality, 3 represents bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants’ scores.

Chapter 4

Digit Ratios and Masculinity - Femininity of Women

4.1 Abstract

Homosexual women are, on average, more masculine in their appearance and behaviour than heterosexual women. We hypothesized that their masculinity was influenced by exposure to elevated levels of androgens during prenatal development. We recruited 199 women (including 67 homosexual women) and measured their masculinity via self-report and observer ratings. Our putative measure of prenatal androgen exposure was the ratio of the index to ring finger (2D:4D), which is hypothesized to be lower in women exposed to elevated levels of androgens during prenatal development. Homosexual women were substantially more masculine than heterosexual women in both self-report and observer ratings. However, homosexual women neither had more male-typical finger length ratios, nor did their finger length ratios relate to their masculinity in any predicted direction. Thus, we found no evidence that increased prenatal androgen exposure influenced masculinity in homosexual women.

4.2 Introduction

Research on masculinity and femininity shows a consistent pattern with respect to women's sexual orientation: Homosexual women recall or report more masculine and less feminine behaviours and self-concepts in childhood and adulthood (Bailey & Zucker, 1995; Rieger et al., 2016), and report more male-typical and less female-typical interests than heterosexual women in adulthood (Lippa, 2005, 2008, 2020). Longitudinal work also suggests that early childhood masculinity is robustly associated with homosexual attractions in females later in life (Li et al., 2017; Xu et al., 2019). Furthermore, based on evaluations by others of their childhood photographs and videos, girls who identified as homosexual in adulthood were rated as more masculine and less feminine than girls who later identified as heterosexual (Rieger et al., 2008; Watts et al., 2018b). In adulthood, homosexual women are also perceived as more masculine than heterosexual women (Rieger et al., 2016; Rieger, Watts-Overall, Holmes, & Gruia, 2020b). These perceptions by others are particularly valuable, because self-reports of masculinity-femininity are possibly biased due to social desirability (Gottschalk, 2003). Thus, in the present study, we included evaluations of participants by observers, in addition to self-reports, to verify the link of female sexual orientation with masculinity-femininity with multiple measures.

Androgen exposure during prenatal development is a potential explanation for the link between sexual orientation and masculinity-femininity within each sex, in addition to explaining overall differences in masculinity-femininity between males and females (Breedlove, 2017; Morris et al., 2004). For example, females with Congenital Adrenal Hyperplasia (CAH), which results in elevated androgen exposure in early gestation, are more likely than their unaffected siblings to engage in male-typical behaviour during childhood (Pasterski et al., 2005), and report same-sex sexual attractions during adulthood (Meyer-Bahlburg et al., 2008).

However, most research on the subject of prenatal androgen influences in humans is informed by postnatal measures, which are assumed to reflect early exposure. The most prominent of these is the ratio of the length of the second to fourth finger digits (2D:4D). Men have lower 2D:4D than women (Grimbos et al., 2010), and this sex difference emerges in early gestation (Galis et al., 2010). Moreover, women with CAH also have lower (more male-typical) 2D:4D than other women, possibly due to the increased androgens exposure (Brown et al., 2002b). Homosexual women have more male-typical 2D:4D than heterosexual women, on average (Rahman & Wilson, 2003; Watts et al., 2018a). This effect was confirmed in both their left and right hands in a meta-analysis, Hedge's *g's* = .23 and .29, .04 < 95% CIs < .51. In contrast, homosexual men do not robustly differ in 2D:4D ratio compared to heterosexual men (Grimbos et al., 2010).

It should be noted that 2D:4D as a measure of prenatal androgen exposure remains a controversial topic because of ongoing debates about causation (McCormick & Carré, 2020; Swift-Gallant et al., 2020), validity due to small effects in the presence of noise in the data (Bailey et al., 2016), and the possibility of publication bias in the literature (Grimbos et al., 2010). That is, the meta-analysis by Grimbos et al. (2010) estimated a given amount of unpublished data with null findings. Once included in their main analysis, these estimated null findings reduced the link between women's sexual orientation and their digit ratios from Hedge's *g's* = .23 and .29, for the left and right hand, to .07 and .13, respectively. Thus, even though reasonable arguments can be made that 2D:4D reflects early androgen exposure related to sexual orientation (Breedlove, 2017), the exact strength of the relationships between these traits remains unclear. For this reason, studies using 2D:4D as a measure as prenatal androgen exposure must be interpreted with caution.

In sum, there may be a relationship between female sexual orientation and masculinity-femininity, and between female sexual orientation and 2D:4D. It is possible that these patterns are further associated. For example, prenatal androgen exposure, possibly reflected by 2D:4D, could be the common factor that influences both women's sexual orientation and their degree of masculinity-femininity. If this is the case, one could expect that homosexual women's increased masculinity, in comparison with heterosexual women, is explained by the finding that homosexual women have, on average, more male-typical 2D:4D than heterosexual women. Hence, the differences in 2D:4D across all women could mediate the relationship between their sexual orientation and their degree of masculinity.

Another possibility is that an interaction between sexual orientation and 2D:4D explains why certain women show a greater degree of masculinity in their behaviours and self-concepts. There is significantly more variability in measures of masculinity-femininity among homosexual women than heterosexual women, because some homosexual women are especially masculine, compared both with heterosexual women and other homosexual women. Homosexual women's greater degree of variability in their masculinity has been repeatedly reported in different studies, and with different measures of masculinity-femininity, including both self-reports and observer ratings (Lippa, 2005, 2008, 2015; Rieger et al., 2008; Rieger et al., 2010; Rieger et al., 2016). For instance, Lippa's (2005) meta-analysis showed that homosexual women were more variable in self-reported masculinity-femininity than heterosexual women, with a mean variance ratio of .67 between the groups. In other words, some homosexual women are especially masculine compared with both heterosexual women and other homosexual women. It is possible that the most masculine homosexual women, in particular, have been exposed to elevated levels of androgens during early development. Hence, homosexual women who have the most male-typical markers of androgen exposure, such as the most male-typical 2D:4D, may also be the women that are the most masculine, as compared to both heterosexual women and other homosexual women with less male-typical 2D:4D. This line of reasoning points to a potential interaction between 2D:4D and sexual orientation, predicting degree of masculinity.

One previous study provided support for the hypothesis that variability in homosexual women's self-reported masculinity-femininity is partly explained by their differing degrees of androgen exposure. Homosexual women who self-identified as "butch" (i.e., masculine) had significantly more male-typical 2D:4D than those who self-identified as "femme" (i.e., feminine; i.e., feminine; Brown et al., 2002a). This finding suggests that there may be different types of homosexual women, with prenatal androgen exposure possibly being the developmental factor which distinguishes between them. Another study measured homosexual women's reported roles during sex (classed as "butch/active" versus "femme/passive") and found no association between reported sex roles and their level of 2D:4D (Rahman & Wilson, 2003). However, because sex roles of homosexual women may simply not equate to their degree of masculinity-femininity (Bailey, Kim, Hills, & Linsenmeier, 1997), we considered the findings by Brown et al. (2002) as potentially more informative with respect to the hypothesis that variation in homosexual women's masculinity-femininity is explained by differences in their 2D:4D.

If it is the case that there is more variability in masculinity-femininity in homosexual women than in heterosexual women, and this is explained by differences in their digit ratios, then it could also imply that there is more variability in 2D:4D among homosexual women than heterosexual women. To our knowledge, no previous research has examined the degree of variability in 2D:4D across women with different sexual orientations. The present research examined this pattern.

Our discussion thus far has focused on a comparison of heterosexual and homosexual women. Bisexual women appear to be intermediate between heterosexual and homosexual women in their masculinity-femininity (Lippa, 2005, 2008).We are not aware of research that specifically compared 2D:4D of bisexual women to those of other women. Bisexual women were included in the present research. However, in order to ease interpretation, we mostly focus on comparisons between heterosexual and homosexual women, and bisexual women are revisited in the Discussion.

We therefore tested the following hypotheses:

Hypothesis 1. Homosexual women are, on average, more masculine than heterosexual women by both self-report and via observer ratings.

Hypothesis 2. Homosexual women have, on average, more male-typical (lower) 2D:4D than heterosexual women.

Hypothesis 3. The relationship between sexual orientation and masculinity in women is mediated by their male-typical 2D:4D.

Hypothesis 4. Homosexual women are, on average, more variable than heterosexual women in both their masculinity-femininity and their 2D:4D.

Hypothesis 5. Homosexual women with the most male-typical 2D:4D show the greatest degree of masculinity, as compared to heterosexual women and other homosexual women with less male-typical 2D:4D.

4.3 Method

4.3.1 Participants

Target participants. In planning our sample size, we drew upon previous studies which used identical measures, and which had computed the correlations of sexual orientation with either masculinity-femininity, or with 2D:4D. Correlations ranged from .30 (for 2D:4D) to .40 to .60 (for measures of masculinity-femininity; for measures of masculinity-femininity; Rieger et al., 2016; Watts et al., 2018a). A power analysis conducted in G*Power determined that a minimum of 112 women would be necessary for the smallest estimated main effect (r = .30) to achieve significant results with a power of .90.

With regards to the moderation and mediation, estimating the necessary sample size proved difficult, as no other study has conducted a moderation or mediation in the same manner as the present study. As such, we erred on the side of caution with participant numbers: Our power analysis for the main effect was based on the more conservative power value of .90 rather than the commonly-used .80, resulting in a sample size requirement of 112 instead of 82 for the smallest expected main effect (r = .30). Additionally, we continued recruiting past this figure as participants were visiting our laboratory for other studies regardless, resulting in a final sample size of 199 – almost double the requirement for the estimated main effects to reach significance. However, even though we did everything we could to ensure a lab-based study such as ours was sufficiently powered, the uncertainty regarding power of the moderation and mediation analyses in particular should be noted, and we revisit this limitation in the Discussion.

A total of 199 women were recruited via Pride festivals, online news sites for homosexual women, and university mailing lists. Using a 7-point scale (Kinsey et al., 1953), women self-identified as "exclusively straight" (n = 44), "mostly straight" (n = 42), "bisexual leaning straight" (n = 15), "bisexual" (n = 18), "bisexual leaning lesbian" (n = 13), "mostly lesbian" (n = 26), or "exclusively lesbian" (n = 41). The mean (SD) age of the sample was 24.22 (6.98), and most were Caucasian (78%), followed by Black (6%), Chinese (5%), and other ethnicities. Some participants opted not to self-report their masculinity-femininity, be video-recorded for observer ratings, or have their digits measured (see Procedure). Due to this, data were available for 196 women for self-reports, 191 women for observer ratings, and 182 women for 2D:4D measures, and numbers of participants varied across analyses.

Raters. Psychology students participated as raters of masculinity-femininity for course credit. In total we had 48 heterosexual male raters, 21 nonheterosexual male raters, 71 heterosexual female raters, and 29 nonheterosexual female raters. The higher proportion of female raters reflects the fact that in our department, the majority of students are female. Ratings of masculinity-femininity are minimally affected by the raters' sex and sexual orientation (Rieger et al., 2010), and this was also the case in the present research.

4.3.2 Measures and Materials

Self-reported sexual orientation. Participants reported their sexual orientation and sexual attraction to men and women on 7-point scales (Kinsey et al., 1953). These scales were highly correlated, $p < .0001$, $r = .97$, 95% CI [.93, 1.0], and averaged within participants. For this average, a score of 0 represented exclusive heterosexuality, a score of 3 bisexuality with equal attractions, and 6 represented exclusive homosexuality.

Self-reported masculinity-femininity. Childhood masculinity was assessed using six items from the Childhood Gender Nonconformity Scale (Rieger et al., 2008), and adulthood masculinity was measured using six items from the Continuous Gender Identity Scale (Rieger et al., 2008). Each scale consists of statements such as “As a child I was perceived as masculine by my peers.” for childhood, and “My mannerisms are not very feminine” for adulthood. Responses were given on 7-point scales ranging from 1 (strongly disagree) to 7 (strongly agree), with answers recoded such that higher numbers always represented greater masculinity. Item reliability (Cronbach’s alpha) was .89 for the Childhood Gender Nonconformity Scale and .92 for the Continuous Gender Identity Scale. These questions were slightly modified to be clearer to a British audience, and full questions are available in the appendix.

2D:4D. Digit measurements were taken from either high-resolution photographs or scans of participants’ hands. For the photographs, participants placed their hands on a flat surface in a supinated (palms facing up) position, with their fingers slightly spread apart, and images were taken from approximately 30 cm above this surface. For the scans, participants placed their hands flat in a pronated (palms facing down) position on the surface of the scanner. In both cases, the palmar surfaces of the hands were visible in the resultant images. Different methods of capturing images (photograph or scanner) did not moderate the relationship between sexual orientation and 2D:4D.

Using these images, digit ratios were measured by two independent raters who were blind to the participants’ sex and sexual orientation. Measurements were performed with the vector graphics package Inkscape 0.92, as computer-assisted techniques produce the most reliable measurements (Allaway et al., 2009). Each rater drew a line as wide as the finger along the proximal skin crease at the base of the finger, between the metacarpal and proximal phalanx. A second line was drawn downwards from the tip of the finger, where it automatically snapped to the centre of the base line. Raters then zoomed in on the tip of the finger for fine adjustments, to ensure that this line matched the tip as closely as possible. Measurements for each digit were averaged between raters, as inter-rater reliability (Cronbach's alpha) exceeded .99 for each digit. For each hand, 2D:4D was calculated by dividing the averaged length of the index finger by the averaged length of the ring finger.

4.3.3 Procedure

Participant session. The University of Essex's Ethics Committee approved this study (GR1303), and the experiment was performed in full accordance with their guidelines and regulations. All participants were over the age of 18 and provided written informed consent. After providing consent, participants completed a survey on their demographics, sexual orientation, and masculinity-femininity, and had photographs or scans of their hands taken. They were then seated in a chair in front of a white wall and had their entire body video-recorded for 5-10 minutes to capture their gestures and movements. Participants answered questions about the weather, their interests, and their childhood, and were not interrupted while answering. Analyses were based on their answer to a neutral question: "How would you describe the weather at this time of year?" A session took approximately 30 minutes and participants were compensated for their time.

Editing of participant videos. The first complete sentence participants spoke in response to the neutral question was extracted using Shotcut. Created clips generally lasted between 10 and 20 seconds. If responses were less than 6 seconds, we took a combination of their first and second sentence. Raters can reliably judge behaviours related to sexual orientation from brief video clips such as these (Rieger et al., 2010; Tskhay & Rule, 2013).

Ratings of masculinity-femininity. Raters, who were blind to the participants' sexual orientation, viewed the edited video clips of target participants. They were instructed to indicate their impression of each woman's appearance and demeanour, in comparison with other women of the same age. For example, they were told to "rate whether this woman appeared or behaved in a more feminine or masculine way". Ratings were completed on 7-point scales, where a score of 1 was "more feminine", 4 "average," and 7 "more masculine." These observer ratings were highly reliable within each rater group (heterosexual and non-heterosexual men and women) and across all raters (all Cronbach's α's > .95). Evaluations were therefore averaged across all raters, producing an average observer-rated masculinity-femininity score for each video-recorded target participant.

4.4 Results

4.4.1 Hypothesis 1.

We hypothesized that homosexual women would be more masculine than heterosexual women. Our measures of masculinity-femininity were self-reports from childhood and adulthood, and observer ratings of adulthood behaviour. For each of these measures, we regressed women's masculinity scores onto their sexual orientation, with sexual orientation treated as a continuous variable in all analyses. Thus, although we focus throughout on the differences between heterosexual women (Kinsey scores 0-1) and homosexual women (Kinsey Scores 5-6), bisexual women (Kinsey scores 2-4) were included in all analyses. We originally tested for both linear and curvilinear effects, to account, for example, for the possibility that bisexual women are closer to homosexual than heterosexual women in their masculinity. However, no such patterns were detected, and we focused exclusively on reporting linear effects.

Homosexual women were significantly more masculine than heterosexual women in their self-reports of both childhood, $p < .001$, $\beta = .24$, 95% CI [.11, .38] (Figure 4.1A), and adulthood, $p < .001$, $\beta = .31$ [.18, .45] (Figure 4.1B), and in observer ratings of their behaviour in adulthood, $p < .001$, $\beta = .38$ [.25, .51] (Figure 4.1C).

4.4.2 Hypothesis 2.

We hypothesized that homosexual women would have more male-typical (lower) 2D:4D than heterosexual women in both their left and right hand. We regressed women's left hand and right hand 2D:4D onto their sexual orientation. Although in the predicted directions, homosexual women did not have a significantly lower 2D:4D than heterosexual women in either their left hand, $p = .26$, $\beta = -.08$ [-.23, .06], or their right hand, $p = .67$, $\beta = -.03$ [-.18, .11] (Figures 4.2A & 4.2B). Thus, our hypothesis that homosexual women would show signs of exposure to elevated levels of prenatal androgens was not supported.

4.4.3 Hypothesis 3.

We hypothesized that the relationship between women's sexual orientation and their degree of masculinity was mediated by their male-typical 2D:4D. Although 2D:4D did not significantly link to sexual orientation, we still conducted this analysis as it was planned in advance. We computed multiple regression analyses. We predicted one of the three masculinity-femininity variables by sexual orientation in Step 1, and by sexual orientation plus left-hand 2D:4D as a mediator in Step 2. We chose to focus on left-hand 2D:4D as it was closer to significance than right-hand 2D:4D (Figure 4.3A). However, we did conduct similar analyses with right-hand 2D:4D, and as expected, it did not influence any effects of sexual orientation on masculinity.

Table 4.1 summarizes the results of the analyses for all three masculinity variables. The effect of sexual orientation on masculinity remained similar in magnitude and levels of significance before and after the inclusion of left-hand 2D:4D as a second predictor. Hence, thus far, there was no evidence that 2D:4D mediated the relationship between sexual orientation and masculinity. Yet, systematic mediation analyses were still necessary to confirm this. We therefore computed three mediation analyses (one for each measure of masculinity-femininity) on the basis of 10,000 bootstrapped samples (Preacher & Hayes, 2008). Left-hand 2D:4D did not significantly mediate the relationship between sexual orientation and masculinity, as the confidence intervals of the mediation effects included zero. This was true for self-recalled childhood masculinity, β = -.005 [-.02, .01], self-reported adulthood masculinity, β = .004 [-.001, .03], and observer-rated adulthood masculinity, β = .002 [-.01, .02].

4.4.4 Hypothesis 4.

We hypothesized that homosexual women would be more variable than heterosexual women in both their masculinity and 2D:4D. To test for their increased variability, we first calculated the residuals for the main effect of sexual orientation on each of the three measures of masculinity-femininity depicted in Hypothesis 1 (Figures 4.1A – 4.1C), and each of the two main effects of sexual orientation on 2D:4D depicted in Hypothesis 2 (Figures 4.2A & 4.2B). We then computed the absolute values of these residuals. Finally, we performed Levene's tests for unequal variances to establish whether the degree of variance (magnitude of absolute residuals) in these effects were stronger in homosexual women than heterosexual women.

Homosexual women were not significantly more variable than heterosexual women in their self-reported childhood masculinity, $p = .21$, $\beta = .09$ [-.05, .23] (Figure 4.3A), but were significantly more variable than heterosexual women in both their self-reports of adulthood, $p = .02$, $\beta = .17$ [.03, .31] (Figure 4.3B), and observer-ratings of their videos from adulthood, $p < .001$, $\beta = .32$ [.18, .46] (Figure 4.3C).

Contrary to our hypothesis, homosexual women were significantly *less* variable in their left-hand 2D:4D than heterosexual women, $p = .01$, $\beta = -.18$ [-.33, .04] (Figure 4.4A). They were not significantly more (or less) variable in their right-hand 2D:4D than heterosexual women, $p = .73$, $\beta = .03$ [-.12, .17] (Figure 4.4B).

Thus, even though we did not find any evidence for homosexual women's increased variability in the marker of prenatal androgen exposure, they were more variable in two out of three measures of masculinity-femininity. This increased variance in masculinity-femininity within homosexual women still pointed to the possibility that there are different types of homosexual women, who may be differentiated by the measure of androgen exposure. We examined this possibility in our next analyses.

4.4.5 Hypothesis 5.

We hypothesized that homosexual women with the most male-typical 2D:4D show the greatest degree of masculinity, as compared to heterosexual women, and other homosexual women with less male-typical 2D:4D. We calculated three multiple regression analyses, predicting one of our three measures of masculinity-femininity. In each analysis, independent variables were sexual orientation, left-hand 2D:4D, and their interaction. If variation in 2D:4D explains why homosexual women are more variable in masculinity-femininity than heterosexual women, then this interaction between sexual orientation and left-hand 2D:4D will be significant.

The results of the regression analyses are summarized in Table 4.2. For all three measures of masculinity-femininity, sexual orientation was the only significant predictor of masculinity-femininity. Neither the effect of left-hand 2D:4D, nor the interaction of sexual orientation with 2D:4D were significant in any of the analyses, and the standardized regression coefficients of these effects were weak in magnitude.

4.5 Discussion

Present findings confirmed that homosexual women were more masculine than heterosexual women, on average. Furthermore, homosexual women were more variable in their masculinity in two out of three measures. However, contrary to our hypothesis, they were significantly less variable in their left-hand (but not right-hand) 2D:4D than heterosexual women, and we do not have any reasonable explanation for this. Furthermore, homosexual women did not have more male-typical digit ratios, nor did their 2D:4D mediate or moderate the relationship between sexual orientation and their degree of masculinity.

The finding that homosexual women are more masculine, in general, but also more variable in their masculinity-femininity than heterosexual women has been previously reported (Rieger et al., 2016). In combination, these findings point to the possible existence of different types of homosexual women, at least with respect to their masculinity-femininity. Hence, it seemed conceivable the most masculine homosexual women, especially, would show signs of increased androgen exposure in the form of more male-typical 2D:4D. However, this was not the case in the present sample. In general, 2D:4D was not significantly more masculine in homosexual women than heterosexual women, even though the effect was in the predicted direction. This is noteworthy, as such a pattern was previously confirmed in a meta-analysis (Grimbos et al., 2010). Perhaps our measure, 2D:4D, was not sensitive enough to robustly indicate prenatal androgen exposure. Yet, we consider this unlikely, as we have previously confirmed a sexual orientation difference in 2D:4D in a much smaller sample of women using the identical methodology (Watts et al., 2018a). Furthermore, as mentioned previously, computer-assisted measurement techniques, such as those employed in the current study, are highly reliable (Allaway et al., 2009) and this was also the case in the present study. Finally, even though men were not the focus of the present research, we had simultaneously gathered 2D:4D data from male participants for a different project. As predicted, these men had significantly lower (more male-typical) 2D:4D than women in both the left hand, d = .33 [.33, .34], p = .004, and right hand, d = .33 [.33, .33], p = .008. Thus, it seems less likely that present null findings are a result of measurement issues. Maybe, in the present study, homosexual women simply did not have more male-typical digit ratios than heterosexual women.

Perhaps the present research should have used a self-report measure of "butch" and "femme" identities, rather than degrees of masculinity-femininity, in order to elicit the hypothesized effects, as such an approach succeeded in previous work (Brown et al., 2002a). Yet, because we reasonably assumed that women who self-identify as "butch" would be more masculine compared to those who identifies as "femme" (Brown et al., 2002a), we expected that the present measures of masculinity-femininity would reveal predicted effects, if they were indeed present.

It is further possible that 2D:4D is not a sensitive enough measure to significantly explain the relationship between sexual orientation and masculinity-femininity. Indeed, there is ongoing debate about the utility and strength of 2D:4D as a measure of androgen exposure (McCormick & Carré, 2020; Swift-Gallant et al., 2020). Perhaps, future research may have more success using other biomarkers of prenatal androgen exposure, in addition to 2D:4D. These may include otoacoustic emissions (McFadden & Pasanen, 1998) or anogenital distance (Barrett et al., 2018). These measures show promise for measuring prenatal androgen exposure, with otoacoustic emissions producing a difference between heterosexual and homosexual women which had a larger effect size than the difference in 2D:4D found in a meta-analysis, *d* = .23 (left) and .29 (right) for 2D:4D, and *d* = .37 for otoacoustic emissions (Grimbos et al., 2010; McFadden & Pasanen, 1998).

A related limitation concerns statistical power in the present study. The sample size we chose was based on the weakest estimated effect: The relationship between left-hand 2D:4D and sexual orientation (*r* = .30). This estimate was taken from a previous research project which found a significant effect using the exact same measurement procedures conducted by the same researchers in the same lab (Watts et al., 2018a). However, this previous project used identical twins as participants. Although we treated these twins as unrelated (i.e. unpaired) in our power calculations for the present study, it was perhaps naïve of us to assume that the effect previously found in twins would be equally strong in unrelated participants. In the present research, the strongest effect was *r* (or *β*) = -.08 in the left hand. With this effect, post-hoc power analyses suggested that we would have needed a minimum sample size of 1634 women of different sexual orientations for it to become significant. If our a-priori sample size estimate had returned such a large number, we would have considered it an unreasonable goal for a lab-based study like ours.

Perhaps, also, we should have considered in advance the relationships between sexual orientation and 2D:4D calculated by the meta-analysis of Grimbos et al. (2010). In this respect, it is worth noting that for the present study, once effect sizes were converted into Hedge's g (the effect size used in the meta-analysis), our estimates of the relationship between sexual orientation and 2D:4D were -.16 for the left hand, and -.06 for the right hand. These estimates fall within the 95% confidence intervals (but closer to zero) of the unadjusted meta-analytic estimates; which were (scaled in the same direction as present effects), *g* = -.23 [-.51, -.06] for the left hand, and *g* = -.29 [-.43, -.04] for the right hand. Our estimates were also close to the publication bias-corrected estimates given in the same meta-analysis, which were -.07 for the left hand and -.13 for the right hand.

Additionally, as mentioned in the method section, it is possible that the present study was not sufficiently powerful for 2D:4D to mediate or moderate the relationship between sexual orientation and masculinity-femininity. However, the effect sizes for the computed mediations and moderations (e.g., Table 4.2) were so small in magnitude that the most parsimonious assessment from the present data is that it seems unlikely to detect such patterns even in much larger samples.

A final point concerns bisexual women, who were intermediate between heterosexual and homosexual women in both measures and variability of their masculinity-femininity. That is, our analyses indicated that the relationships of sexual orientation with masculinity-femininity were explained by simple linear effects, whereas we found no evidence for curvilinear effects that would, for instance, suggest that bisexual women are closer to homosexual women than heterosexual women in their masculinity-femininity. Although it seems sensible that bisexual women are between these two groups with respect to their masculinity-femininity, there are no strong hypotheses regarding what factors would cause this outcome (Rieger et al., 2020a).

In sum, the present study did not find evidence of a link between masculinity in homosexual women and exposure to androgens in the prenatal period, as reflected in finger length ratios. In fact, in the present study, homosexual women showed no clear signs of elevated prenatal androgen exposure, as compared to heterosexual women. Thus, our hypothesis that homosexual women’s male-typed traits were influenced by early hormonal influences remains unconfirmed.

Table 4.1 - *Multiple Regression Analyses for Sexual Orientation and Left-Hand 2D:4D predicting Self-Reported Childhood and Adulthood Masculinity (Step 1 N = 196, Step 2 N = 181) and Observer-Rated Adulthood Masculinity (Step 1 N = 191, Step 2 N = 180).*

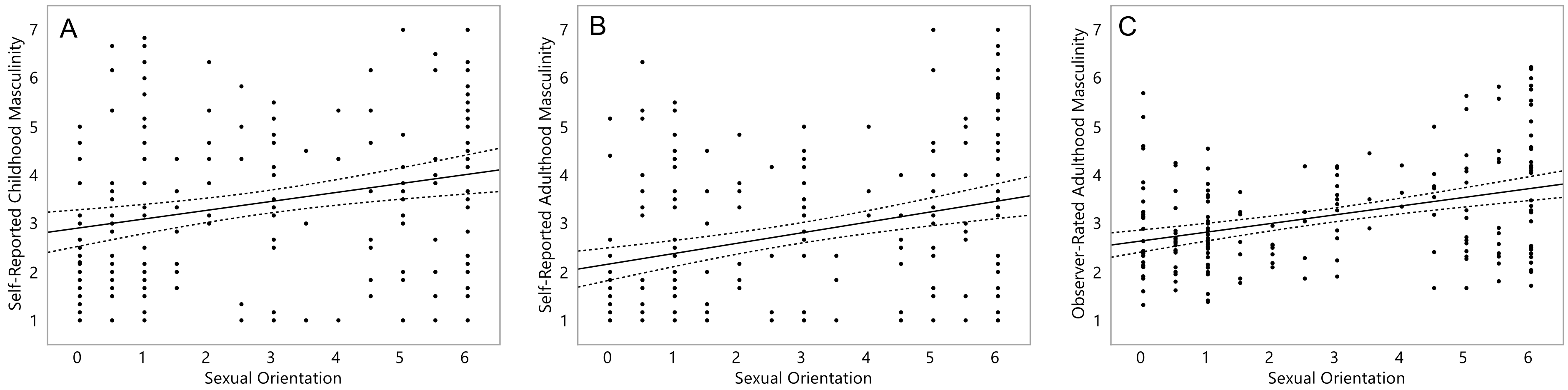
Step 1	**Self-Reported Childhood Masculinity**	**Self-Reported Adulthood Masculinity**	**Observer-Rated Adulthood Masculinity**
Variables	β	β	β
Sexual Orientation (SO)[1]	.24 [.11, .38]**	.31 [.18, .45]**	.38 [.25, .51]**
Step 2	**Self-Reported Childhood Masculinity**	**Self-Reported Adulthood Masculinity**	**Observer-Rated Adulthood Masculinity**
Variables	β	β	β
Sexual Orientation (SO)[1]	.25 [.10, .39]**	.29 [.15, .43]**	.36 [.22, .50]**
Left-Hand 2D:4D[2]	.07 [-.08, .21]	-.05 [-.19, .09]	-.03 [-.16, .11]

Note. R^2*'s* for the three models are .06, .10 and .14 in Step 1; .06, .09 and .12 in Step 2, respectively. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Higher scores indicate a more homosexual orientation. [2]Lower scores indicate a more male-typical 2D:4D. † $p < .1$, * $p < .05$, ** $p < .01$.

Table 4.2 - *Multiple Regression Analyses for Sexual Orientation and Left-Hand 2D:4D predicting Self-Reported Childhood and Adulthood Masculinity (N = 181) and Observer-Rated Adulthood Masculinity (N = 180).*

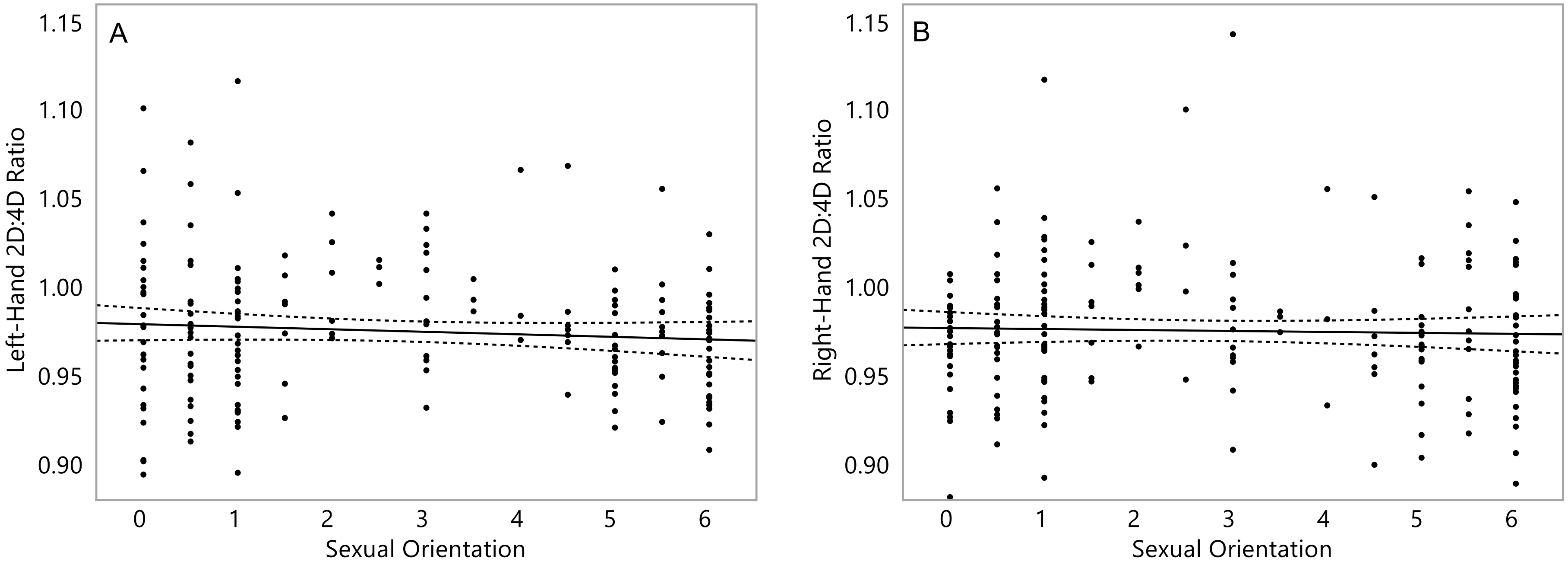
Step 1	**Self-Reported Childhood Masculinity**	**Self-Reported Adulthood Masculinity**	**Observer-Rated Adulthood Masculinity**
Variables	β	β	β
Sexual Orientation (SO)[1]	.25 [.10, .39]**	.28 [.14, .43]**	.36 [.22, .50]**
Left-Hand 2D:4D[2]	.07 [-.08, .22]	-.07 [-.22, .08]	-.03 [-.18, .12]
SO x 2D:4D	.01 [-.14, .16]	-.05 [-.20, .10]	-.01 [-.16, .13]

Note. R^2*'s* for the three models are .06, .09 and .13, respectively. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Higher scores indicate a more homosexual orientation. [2]Lower scores indicate a more male-typical 2D:4D. † $p < .1$, * $p < .05$, ** $p < .01$.



p < .001, *β* = .24 [.11, .38] *p* < .001, *β* = .31 [.18, .45] *p* < .001, *β* = .38 [.25, .51]

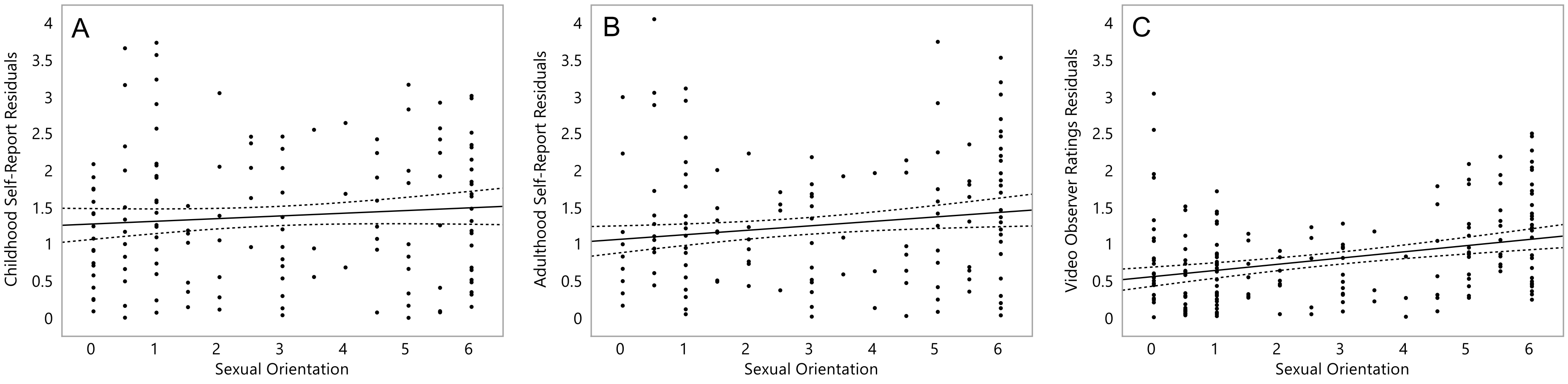
Figure 4.1 - Masculinity of 196 women (self-report from childhood, A; and adulthood, B) and 191 women (observer-ratings, C). On the Y axis, masculinity scores, with higher scores representing a greater degree of masculinity. On the X axis, 0 represents exclusive heterosexuality, 3 represents bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants' scores. Statistics represent linear effects.



p = .26, *β* = -.08 [-.23, .06]

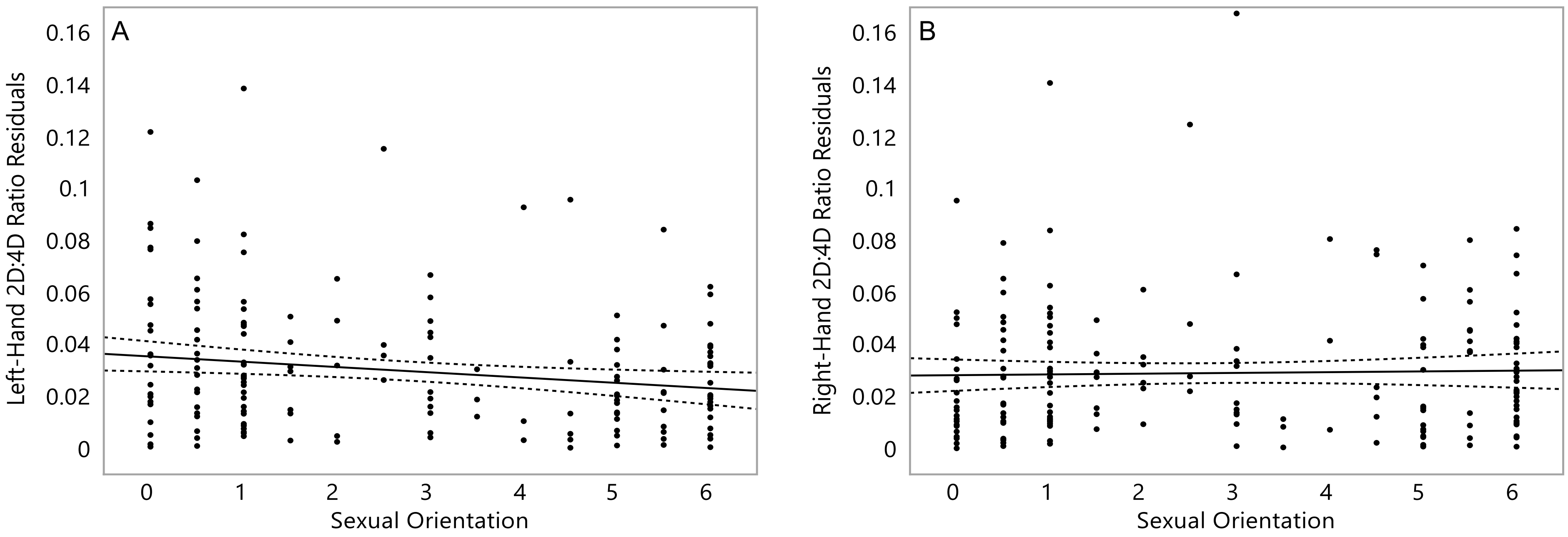
p = .67, *β* = -.03 [-.18, .11]

Figure 4.2 - 2D:4D of 182 women (left-hand 2D:4D, A; and right-hand 2D:4D, B). On the Y axis, 2D:4D is the length of the index finger divided by the length of the ring finger, with lower scores representing a more male-typical 2D:4D. On the X axis, 0 represents exclusive heterosexuality, 3 represents bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants' scores. Statistics represent linear effects.



p = .21, β = .09 [-.05, .23] p = .02, β = .17 [.03, .31] p < .001, β = .32 [.18, .46]

Figure 4.3 - Masculinity of 196 women. Absolute residuals derived from the effects displayed in Figure 1 for masculinity of 196 women (self-report from childhood, A; and adulthood, B) and 191 women (observer-ratings, C). On the Y axis, residuals for masculinity, with higher scores representing a greater degree of variance from the main effect. On the X axis, 0 represents exclusive heterosexuality, 3 represents bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants' residuals. Statistics represent linear effects.



p = .01, *β* = -.18 [-.33, .04]

p = .73, *β* = .03 [-.12, .17]

Figure 4.4 - 2D:4D of 182 women. Absolute residuals derived from the effects displayed in Figure 2 for 2D:4D (left-hand 2D:4D, A; and right-hand 2D:4D, B). On the Y axis, residuals for 2D:4D, with higher scores representing a greater degree of variance from the main effect. On the X axis, 0 represents exclusive heterosexuality, 3 represents bisexuality, and 6 represents exclusive homosexuality. Triple lines represent regression coefficients with their 95% confidence intervals. Dots represent participants' residuals. Statistics represent linear effects.

4.6 Appendix

Scale Items: Childhood Gender Nonconformity Scale (Women)

As a child I felt that I had more in common with boys than girls.

As a child I was perceived as masculine by my peers.

As a child I preferred playing with boys rather than girls.

I was more masculine than feminine.

As a child I sometimes wished I had been born a boy rather than a girl.

As a child I usually avoided feminine clothing (such as dresses).

Scale Items: Adulthood Continuous Gender Identity Scale (Women)

My mannerisms are not very feminine.

I assume most people see me as not very feminine.

I assume most people see me as masculine.

I consider myself very masculine in my behaviors and interests.

I do not consider myself very feminine in my behaviors and interests.

I consider myself more masculine than feminine.

Chapter 5 Sexual Arousal and Empathy in Women and Men

5.1 Abstract

Women, more than men, show physiological sexual arousal to both males and females. Moreover, compared to men, women are more empathetic. We hypothesized that women's elevated empathy for both sexes triggers their increased physiological sexual arousal to both sexes. We measured the sexual responses of 111 men and 147 women via genital arousal and pupil dilation to erotic videos, in addition to their self-reported trait empathy, and for a subsample (32 men and 33 women), their empathy towards the actors in the videos. Women had stronger sexual responses to both sexes than men, although this pattern was clearer for genital arousal than pupil dilation. Women also reported higher levels of trait empathy, and were somewhat, but not significantly, more empathetic towards both males and females in the videos. However, there was no evidence that the sex difference in sexual responses was influenced by a sex difference in empathy.

5.2 Introduction

Studies measuring physiological sexual arousal to sexually explicit stimuli suggest a robust difference between the sexes: The majority of men display strong sexual arousal to stimuli featuring their preferred sex and less to no arousal to stimuli featuring their non-preferred sex, whereas most women display substantial sexual arousal to stimuli featuring either sex, regardless of their sexual preferences. For this reason, sexual arousal to explicit stimuli correlates more strongly with self-reported sexual orientation in men than in women. This sex difference has been demonstrated most commonly with measures of penile circumference and vaginal pulse amplitude (Chivers et al., 2004; Rieger et al., 2015; Watts et al., 2018c). Other measures also confirm this sex difference, including the comparison of penile with clitoral responses (Suschinsky, Dawson, & Chivers, 2020), pupil dilation (Attard-Johnson et al., 2016; Rieger et al., 2015; Rieger & Savin-Williams, 2012), viewing time (Ebsworth & Lalumière, 2012; Israel & Strassberg, 2009; Lippa, 2012), and cerebral responses (Safron et al., 2019; Sylva et al., 2013) while viewing sexual stimuli. Some exceptions to this pattern exist: Bisexual men can show stronger arousal to both sexes, compared to heterosexual and homosexual men (Jabbour et al., 2020), and homosexual women respond more strongly to their preferred sex than the other sex, compared to heterosexual and bisexual women (Rieger et al., 2016). In general, however, stronger responses to one sex over the other sex is more commonly seen in men than in women (Bailey, 2009).

Several broad reasons for this difference in sexual arousal patterns between men and women have been proposed (see review by Chivers, 2017), including that women's sexuality may be more affected by contextual and social factors than men's (Baumeister, 2000), that women may be more fluid in their sexual attractions than men

(Diamond, 2016), or that women's sexual responses change according to their levels of fertility, whereas men are not affected by such a mechanism (Gangestad & Simpson, 2000). However, no definitive explanation has yet been found. The most prominent proposal for the described sex difference in physiological sexual response is the “preparation hypothesis” (Suschinsky & Lalumière, 2011). This hypothesis holds that since forced copulation is common in many animal species, including humans, and because forced copulation can lead to genital injury, women may have evolved to respond to any sexual stimulus with genital arousal, and consequently vaginal lubrication, in order to mitigate this risk. Not all findings for female sexual arousal are apparently consistent with this proposal, but on the weight of all available evidence, the preparation hypothesis remains the favored theory for explaining women’s non-specific patterns of arousal (see review by Lalumière et al., 2020).

The present study did not focus on an ultimate explanation for the sex difference in sexual response. Rather, unlike previous work, we examined potential personality differences between men and women, to determine whether this could explain the difference in their sexual responses. One possible reason for the sex difference in sexual arousal is a sex difference in empathy. Empathy has been defined as the capacity to understand and respond to the mental states of others (Lamm, Batson, & Decety, 2007), such as an observer’s ability to notice and directly share in the internal state of others, even without seeing them in person (Decety & Jackson, 2004). Several studies have pointed out that women have a greater degree of empathy, on average, than men. This difference emerges in childhood, widens in adolescence, and carries into adulthood (Hoffman, 1977; Lam et al., 2012; Rose & Rudolph, 2006). In adulthood, women are also more likely than men to adopt and synchronize emotions with an observed individual (Doherty, Orimoto, Singelis, Hatfield, & Hebb, 1995). Furthermore, while women show steady increases in empathy towards both sexes during childhood and onwards from adolescence, men exhibit an increase in empathy towards females but a decrease in empathy towards other males in the same time frame (Bryant, 1982; Mestre, Samper, Frías, & Tur, 2009; Olweus & Endresen, 1998; Stuijfzand et al., 2016).

It has been hypothesized that the explanation for most differences in brain and behavior between men and women is their differing degrees of exposure to androgens during prenatal development (Breedlove, 2010), and a range of studies using both prenatal and postnatal measures have linked androgens to empathy. In one study, mothers underwent amniocentesis (assessment of amniotic fluid) during their second trimester of pregnancy, and those children who were exposed to higher levels of androgens during prenatal development were scored lower on tasks measuring empathy at ages 6-8 (Chapman et al., 2006). Additionally, bloodstream androgen levels at age 9 were found to correlate negatively with empathy (Pascual-Sagastizabal et al., 2013), and the administration of testosterone into the bloodstream reduces cognitive empathy temporarily in adult women (van Honk et al., 2011). Since men in general are exposed to far greater levels of androgens than women, both during prenatal development and in their bloodstreams throughout their lifespan (van Honk et al., 2011), it is possible that the sex difference in empathy is explained by the sex difference in androgen exposure.

In sum, androgen exposure provides a plausible explanation for why men may exhibit less empathy than women, as well as a plausible explanation for why most men show physiological sexual arousal to only one sex, and most women show roughly equal sexual arousal to both sexes. It is therefore possible also that these differences are inter-related. If so, women's greater degree of empathy, compared to men's, may be linked to why most women show physiological sexual arousal to both sexes, while most men show sexual arousal to only one, preferred sex. The present study combined these two separate lines of research, and measured sexual response with genital measures (vaginal pulse amplitude and penile circumference) and pupil dilation to sexual stimuli. In addition, we assessed empathy as a trait via a standardized questionnaire plus, for a subsample of participants, questions about their empathy for the male and female actors in the sexual videos.

We therefore tested the following hypotheses:

Hypothesis 1. Women will show stronger sexual responses towards both sexes than men.

Hypothesis 2. Women will report greater empathy than men. This will be the case on a trait level, and also with respect to their reactions to both male and female sexual stimuli.

Hypothesis 3. Women's stronger sexual responses to both sexes, compared with men's, will be linked to (mediated by) women having greater empathy than men.

5.3 Method

5.3.1 Participants

With the used variables of sexual response, we expected the effect size (Cohen's *d)* of the sex difference in genital arousal to males or females to be, at a minimum, *d* = 1.18 (estimated from previous data from our research group), the corresponding sex difference in pupil dilation to be, at minimum, *d* = 1.00 (Rieger & Savin-Williams, 2012), and the sex difference in trait empathy to be *d* = .64 (Mestre et al., 2009). Because no previous study has conducted a mediation with these variables, we erred on the side of caution with participant numbers by conducting power analyses using the more conservative power value of .90. Using this figure, a power analysis using G*Power determined that at a minimum, a total of 106 men and women would be necessary for the smallest estimated sex difference (*d* = .64) to reach significance. However, because this experiment ran concurrently with other experiments needing larger sample sizes, and we saw no reason to exclude participants from this particular study, we continued recruiting and for the majority of our analyses we had substantially more participants (N = 258) than the power analysis recommended.

A total of 258 participants (111 men and 147 women) were recruited via pride festivals, online magazines, and university fairs and mailing lists, and reported their sexual orientation using a 7-point scale (Kinsey et al., 1953). The 111 recruited men self-identified as "exclusively straight" (n = 50), "mostly straight" (n = 7), "bisexual leaning straight" (n = 17), "bisexual" (n = 11), "bisexual leaning gay" (n = 5), "mostly gay" (n = 10), or "exclusively gay" (n = 11). The 147 women self-identified as "exclusively straight" (n = 31), "mostly straight" (n = 33), "bisexual leaning straight" (n = 12), "bisexual" (n = 15), "bisexual leaning lesbian" (n = 11), "mostly lesbian" (n = 16), or "exclusively lesbian" (n = 29). The mean (SD) age was 24.90 (10.14) for men and

24.52 (7.62) for women. For men, 80% were White, followed by 4% Chinese, 4% Indian, and other ethnicities. For women, 78% were White, 6% Chinese, 4% Black, and other ethnicities.

Self-reported trait empathy data were available for all 258 participants. Due to some participants opting out of components involving genital measurements, genital arousal data were available for 247 of the 258 participants. For 40 of the 65 individuals for whom we assessed their empathy for the actors (see next paragraph), eye data were not collected, because the relevant computer ports were in use by equipment for another study at the time. Thus, pupil data were available for 218 of the 258 participants, and the number of participants varies across analyses.

In a sub-sample of 65 participants, we examined a sex difference in empathy towards each sexual stimulus rather than as a general trait, by asking each participant how much they empathized with each actor in each video. Such an examination has (to our knowledge) never been conducted previously, and as such, we could not predict whether this number was sufficiently powerful to detect a sex difference. Because of this, we considered this particular study component to be exploratory. We originally intended to continue testing until our other projects concluded. Unfortunately, our testing was interrupted by the COVID-19 pandemic. As such, the final number of participants for this study component was 65, possibly underpowered, and its results must be interpreted with caution. Regardless, we included findings from this study component, because we believed it could potentially add to the understanding of a link between sex differences in empathy and in sexual arousal. Within this subsample, of the 65 participants, 33 were female and 32 were male. Of the 32 men, 24 were exclusively or mostly heterosexual, 6 were bisexual, and 2 were mostly homosexual.

Of the 33 women, 24 were exclusively or mostly heterosexual, 7 were bisexual, and 2 were mostly or exclusively homosexual.

5.3.2 Measures and Materials

Trait empathy. This was measured using the 28-item Interpersonal Reactivity Index, answered on a 7-point scale ranging from “Does not describe me well” to “Describes me very well” (Davis, 1983). This scale was considered particularly suitable for the present research because we believed that any effect would be driven specifically by perspective-taking, which is directly measured by one subscale of this index (Perspective-Taking), and is conceptually related to another (Fantasy). Example items include “Before criticizing someone, I try to imagine how I would feel if I were in their place” and “I sometimes try to understand my friends better by imagining how things look from their perspective.” This scale is comprised of four 7-item subscales: Perspective-Taking, Fantasy, Empathetic Concern and Personal Distress. All items were coded in a way such that higher scores indicated greater degree of empathy. Item reliability (Cronbach’s alpha) was .82 across the entire scale for men and women (separately). For the subscales, item reliability ranged from .75 for Personal Distress in male participants to .82 for Perspective-Taking in female participants. For each participant, item responses were averaged for each of the four subscales, and further averaged again across the entire scale to create a composite trait empathy score. Both this composite score and the four sub-scales were used for analyses.

Per-video empathy. After each explicit video, a subsample of 65 participants were asked “How much did you empathize with this person's feelings?” and “How well were you able to see things from this person's perspective?” Item reliability (Cronbach’s alpha) for the six per-video empathy and perspective-taking questions across all male stimuli was .91 for men and .88 for women. For the six questions across the three female stimuli, it was .94 for men and .90 for women. For the per-video empathy questions only, item reliability across all three male stimuli was .92 for men and .78 for women, and across all three female stimuli it was .86 for men and .81 for women. For the per-video perspective taking questions only, item reliability across all three male stimuli was .88 for men and .83 for women, and across all three female stimuli it was .92 for men and .79 for women. The per-video empathy and perspective taking questions were strongly correlated, ranging from $p < .0001$, $r = .67$, 95% CI [.42, .82] for male participants viewing male stimuli, to $p < .0001$, $r = .89$, [.78, .94] for male participants viewing female stimuli. We averaged responses in a way to create a) per-video empathy, and b) per-video perspective taking across male stimuli and, separately, across female stimuli for each participant. We further averaged these two average responses together to create c) a composite per-video empathy score towards males and towards female for each participant, and all 6 of these averages were used in analyses.

Sexual orientation. Participants reported both their sexual orientation and degree of attraction to men or women using 7-point scales (Kinsey et al., 1953). These two scales were highly correlated in both men, $p < .0001$, $r = .98$, 95% CI [.97, .99], and women, $p < .0001$, $r = .97$, [.96, .98], and averaged within participants. For this average, a score of 0 represented exclusive heterosexuality, a score of 3 bisexuality with equal attractions to women and men, and 6 exclusive homosexuality.

Stimuli. The sexual stimuli consisted of 3-minute videos, three featuring a female model and three featuring a male model, each of them masturbating in a bedroom. These stimuli were selected in a previous study in which 200 videos were rated on their sexual appeal by men and women of different sexual orientations (Rieger et al., 2015), and the top three female and male videos were used in the present study. Neutral stimuli to assess baseline genital responses were 2-minute clips taken from a nature documentary. Their engaging but nonsexual content facilitated participants' return to an unaroused level. However, these nature videos were not used for pupil dilation baseline, as their engaging content might elicit dilation for reasons other than sexual arousal. Thus, two 1-minute animations of clouds were used to obtain a pupillary baseline. All videos were edited using MPEG Streamclip and Final Cut Pro to be of similar luminance.

Genital arousal. For all participants, the genital arousal signal was recorded at 200 Hz in 16-bits resolution using a BIOPAC MP150 data acquisition unit and AcqKnowledge software. For male participants, genital arousal was measured via changes in penile circumference using a penile plethysmograph, and the signal was low-pass filtered at 10 Hz. For female participants, it was measured as changes in peak-to-trough vaginal pulse amplitude (VPA) using a vaginal photoplethysmograph, high-pass filtered at 0.5 Hz. The VPA exhibits both convergent and discriminant validity for the measurement of women's sexual responses (Suschinsky et al., 2009).

Pupil dilation. Pupil dilation data was measured with a SR Research EyeLink 1000 infrared eye tracking unit. A 35 mm lens focused on the participants' right eye, positioned approximately 60 cm from the participants' head and sampling at a rate of 500 Hz. The infrared light emitted by the eye tracker is reflected by the pupil, and the number of pixels reflected were recorded. Because raw pupil area data included "0's" for missing values, for instance from blinks or head movements, these values were removed prior to analyses.

5.3.3 Procedure

Participants provided written informed consent before completing a survey which included demographic information and self-report measures of empathy and sexual orientation. They were then seated in a booth with dim lighting conditions kept the same for all participants. The eye tracking unit was positioned approximately 600 mm from their face, and was calibrated by participants fixating on dots outlining the screen. Participants were asked to keep their eyes on the screen throughout the experiment regardless of whether they enjoyed the video, and were instructed in the use of the genital device. Participants had complete privacy, and communication was possible via an intercom. Men placed the gauges midway around their penises and women inserted the vaginal photoplethysmograph, and the experimenter checked the signal from the genital device to ensure it was correctly placed before commencing the experiment.

Participants first viewed an animation of clouds, followed by alternating sexual and nature videos. These were displayed in a random order, but a sexual video was always followed by a nature video. Following the sixth nature video, another animation of clouds was displayed, and the experiment was concluded.

For a subsample of 65 participants, each sexual video was followed by two questions on how strongly the participant empathized with the actor in the video. The next nature video began immediately after these questions were answered. The entire procedure took approximately 90 minutes. At the end of the experiment participants were debriefed and compensated.

Following previous procedures (Watts et al., 2018c), genital data and pupil data were averaged across the duration of each stimulus and for each participant. These averages were then standardized within participants, producing a z-score for each participant and stimulus. For genital data, standardized responses to the 5 seconds preceding each sexual stimulus (following the display of a neutral stimulus, and after the participant had returned to baseline) were subtracted from the standardized response to the sexual stimulus. For pupil data, standardized responses to neutral stimuli (the animated clouds) were subtracted from standardized responses to all sexual stimuli. We then computed, for each participant, average responses across all sexual stimuli of a given type (female or male), which reflected their responses to each sex as compared to baseline.

These standardized response scores were used to calculate the two experimental variables. The first is the absolute difference between each participant's responses to males and females, calculated by deducting one mean from the other, such that zero indicates equal responses to males and females, and deviation from zero means a stronger response to one sex over the other. We expected men to have a large difference in their sexual arousal to one sex and the other. For example, a heterosexual man should show strong arousal to females, and little to males. Conversely, a homosexual man should show strong arousal to males and little to females. In each case, there would be a notable absolute difference in their responses to males or females (but see our below comment on bisexual men). In contrast to heterosexual and homosexual men, women of all sexual orientations were expected to show smaller absolute differences, because they would respond similarly to both males and females. Thus, regardless of their sexual orientations, we expected men's absolute differences in their arousal to males or females to be larger than women's (Rieger et al., 2015).

The second variable used in the present study was participants' responses to their less-arousing sex, or their minimum arousal. To calculate this, we selected the mean response to whichever stimulus category (male or female) each participant had a lower response to. For men (relative to women), we expected their response to their less-arousing sex to be low, because they usually respond strongly to one sex and weakly to the other, regardless of their sexual orientation. For women, more than men, we expect their response to their less-arousing sex to be higher, because women of all sexual orientations are more likely to respond to both sexes, including their less-arousing sex (Rieger et al., 2015).

There was the possibility of an interaction between sex and sexual orientation affecting sexual response, because bisexual men's responses can be more bisexual than the responses of heterosexual or homosexual men, whereas in women, bisexual arousal is more consistently observed across all sexual orientations (Rieger et al., 2016; Slettevold et al., 2019). In fact, such a pattern was found in the present data. We therefore systematically checked whether the inclusion or exclusion of bisexual (and, further, homosexual) individuals would impact the main effects being explored in the study. We determined that overall, effect sizes (*d's*) increased if only heterosexual men and women were compared, with the largest change being $d = 1.59$ to $d = 2.48$ for absolute difference in genital arousal, and the smallest change being $d = .25$ to $d = .27$ for pupil dilation to the less-arousing sex. The sex difference in empathy also strengthened slightly, from $d = .64$ to $d = .72$ when only heterosexuals were included in analyses. Crucially, the inclusion or exclusion of bisexual (and homosexual) individuals did not change the direction of main findings with respect to sex differences in any measure of sexual arousal or empathy. Because the present research had no main focus on sexual orientation, but rather on sex differences, we therefore did not concentrate on complex analyses that differentiated bisexual from heterosexual and homosexual men and women, and we decided to focus on sex differences across all participants, regardless of their sexual orientations.

For the subsample of 65 participants for whom per-video empathy data were available, we had three scales: Per-video empathy, per-video perspective taking, and a composite per-video empathy score consisting of the average of these two scales. Using these, we computed variables akin to the ones for sexual arousal: For both the two subscales and their composite, an absolute difference score was first calculated by deducting one mean from the other, and second, a minimum empathy score was calculated by selecting whichever category, male or female, each participant reported lower empathy towards. In keeping with our predictions for sexual arousal, and in light of the research suggesting that women show greater empathy towards both sexes while men favor one sex (Bryant, 1982; Olweus & Endresen, 1998), we expected that men, more than women, would show a larger difference in absolute empathy. However, we expected that women, would score higher than men in their responses to whichever sex they showed lower empathy towards – in other words, women would have a greater minimum empathy.

5.4 Results

5.4.1 Hypothesis 1.

We hypothesized that women will show stronger sexual responses towards both sexes than men. We first examined the absolute difference in sexual arousal (genital arousal or pupil dilation) to one sex over the other, expecting it to be greater in men than in women. For genital arousal, an independent-samples t-test indicated that the absolute difference score in men (*M* = 1.32, *SD* = .67) was significantly greater than the absolute difference score in women (*M* = .79, *SD* = .58), *t*(245) = 6.68, *p* < .001, *d* = .87, 95% CI [.79, .94] (Figure 5.1A). For pupil dilation, the absolute difference score in men (*M* = .42, *SD* = .31) was in the expected direction but not significantly greater than the absolute difference score in women (*M* = .35, *SD* = .29), *t*(214) = 1.84, *p* = .07, *d* = .26 [.22, .29] (Figure 5.1B).

We then examined sexual response to the less-arousing sex, expecting it to be greater in women than in men. For genital arousal, men's responses to the less-arousing sex (*M* = .30, *SD* = .49) were significantly lower than women's responses to the less-arousing sex (*M* = 1.17, *SD* = .59), *t*(245) = -12.27, *p* < .001, *d* = -1.59 [-1.66, -1.52] (Figure 5.2A). For pupil dilation, men's responses to the less-arousing sex (*M* = .54, *SD* = .59) were not significantly lower than women's (*M* = .38, *SD* = .61), and if anything, this trend was in the opposite direction to the hypothesized, *t*(214) = 1.94, *p* = .05, *d* = .27 [.19, .35] (Figure 5.2B).

5.4.2 Hypothesis 2.

We hypothesized that women will report greater empathy than men, both as a trait and with respect to their reactions to both male and female sexual stimuli.

Composite trait empathy was significantly lower in men (*M* = 4.36, *SD* = .66) than in women (*M* = 4.84, *SD* = .66) *t*(256) = -5.75, *p* < .001, *d* = -.72 [-.81, -.65] (Figure 5.3A).

Men scored significantly lower than women in all four trait empathy subscales: For Perspective-Taking, men (*M* = 4.91, *SD* = 1.00) scored significantly lower than women (*M* = 5.17, *SD* = 1.09) *t*(256) = -1.99, *p* = .048, *d* = -.25 [-.38, -.12]. Similarly, men (*M* = 4.71, *SD* = 1.12) scored significantly lower in Fantasy than women (*M* = 5.14, *SD* = 1.14) *t*(256) = -3.01, *p* = .002, *d* = -.38 [-.52, -.24]. This was also true for Empathetic Concern, where men (*M* = 4.72, *SD* = 1.04) scored significantly lower than women (*M* = 5.54, *SD* = 1.03) *t*(256) = -6.31, *p* < .001, *d* = -.80 [-.92, -.67], and Personal Distress, where men (*M* = 3.11, *SD* = 1.00) again scored significantly lower than women (*M* = 3.51, *SD* = 1.16) *t*(256) = -2.93, *p* = .004, *d* = -.37 [-.50, -.28].

For the subsample of 65 participants, we then examined their composite per-video empathy scores. The absolute difference in composite per-video empathy was not, as hypothesized, significantly greater in men (*M* = 1.72, *SD* = 1.41) than in women (*M* = 1.76, *SD* = 1.09), *t*(63) = -.12, *p* = -.90, *d* = -.03 [-.33, .27] (Figure 5.3B). The minimum per-video composite empathy score was, as hypothesized, lower in men (*M* = 2.43, *SD* = 1.24) than women (*M* = 2.92, *SD* = 1.04), albeit not significantly so, *t*(63) = -1.71, *p* = .09, *d* = -.43 [-.70, -.15] (Figure 5.3C).

Similarly, no significant differences were found in the per-video subscales. That is, for the per-video empathy subscale, the absolute difference in per-video empathy was not, as hypothesized, significantly greater in men (*M* = 1.71, *SD* = 1.52) than in women (*M* = 1.63, *SD* = 1.20), *t*(63) = .21, *p* = .83, *d* = .05 [-.27, .38]. Minimum per-video empathy was not significantly lower in men (*M* = 2.41, *SD* = 1.35) than women (*M* = 2.97, *SD* = 1.19), *t*(63) = -1.78 *p* = .08, *d* = -.45 [-.75, -.14]. Similarly, the absolute difference in per-video perspective taking was not significantly greater in men (*M* = 1.75, *SD* = 1.60) than in women (*M* = 1.94, *SD* = 1.22), *t*(63) = -.54, *p* = .59, *d* = -.16 [-.48, .21], and the minimum per-video perspective taking was not significantly lower in men (*M* = 2.45, *SD* = 1.29) than in women (*M* = 2.84, *SD* = 1.06), *t*(63) = -1.33, *p* = .19, *d* = -.34 [-.62, -.05].

5.4.3 Hypothesis 3.

We hypothesized that women's stronger sexual responses to both sexes, compared with men's, would be linked to women having greater empathy than men. In order to investigate this, we computed a series of regression analyses predicting genital arousal or pupil dilation to sexual stimuli. We predicted either the absolute differences in sexual response, or the responses to the less-arousing sex. In all analyses, in Step 1, sex was the only predictor of sexual response. In Step 2, one empathy variable was included as predictor alongside sex – either self-reported composite trait empathy, absolute difference in composite per-video empathy, or minimum composite per-video empathy. If empathy explained differences between men and women in sexual response, then the inclusion of these empathy variables should mediate (weaken) any sex difference in sexual response. For the subsample of participants who answered per-video empathy questions, analyses only applied to genital responses (because for them, pupil dilation had not been measured).

We first computed four multiple regression analyses, predicting the genital arousal or pupil dilation absolute difference score by sex and, in a second step, by both sex and composite trait empathy. The inclusion of composite trait empathy as a predictor had little effect on the relationship between sex and absolute difference in genital arousal or pupil dilation (Table 5.1). To fully examine a potential mediation of empathy (or the lack of it), we computed analyses on the basis of 10,000 bootstrapped samples (Preacher & Hayes, 2008). Composite trait empathy did not significantly mediate the relationship between sex and absolute difference in sexual response, as the confidence intervals of the indirect effects included zero, both for genital arousal, β = .03 [-.06, .11], and pupil dilation, β = -.02, [-.11, .09].

We then computed four further regression analyses predicting genital arousal and pupil dilation to the less-arousing sex. The inclusion of composite trait empathy as a predictor had almost no effect on the differences between men and women in their genital responses or pupil dilation to the less-arousing sex (Table 5.2). We again conducted analyses based on 10,000 bootstrapped samples. Trait empathy did not significantly mediate the relationship between sex and minimum arousal score for either genital response, β = .05, [-.02, .12], or pupil dilation β = .03, [-.06, .14].

We conducted 16 further mediation analyses using one of four self-reported empathy sub-scales as a mediator. None of them significantly mediated the relationship between sex and either measure of genital arousal or pupil dilation. In almost all these analyses, the introduction of an empathy subscale as a further predictor variable did not even change standardized regression coefficient for the relationship between sex and sexual arousal by .01. Of all the models, the biggest change was for the sex difference in absolute difference in genital arousal once Empathetic Concern was included as a covariate; the change was from β = -.39 [-.51, -.28] to β = -.50 [-.76, -.24]. This strengthening of the relationship suggest suppression rather than mediation, but regardless, it was not found to be significant when checked systematically via bootstrapping analyses. The indirect effect included zero in its 95% confidence interval, β = .01, [-.09, .12]. In the interest of brevity, these analyses are not included in the tables.

Likewise, the difference between men and women in their genital response was not weakened by including a measure of per-video empathy as a predictor. In these analyses, we focused on whether the sex difference in genital arousal was mediated by the corresponding sex difference in per-video empathy – in order words, we checked if absolute difference in genital arousal was mediated by absolute difference in per-video empathy, and whether arousal to the less-arousing sex was mediated by minimum per-video empathy. The sex differences in sexual response remained identical when a corresponding measure of video-rated empathy was included in a second step of the regression analyses (Tables 5.3 and 5.4). Bootstrapping analyses also did not show significant mediation effects. Of these, the biggest change for the sex difference in minimum genital arousal happened once minimum per-video empathy was included as a covariate; the change was from β = .55 [.34, .76] to β = .50 [.29, .70]. Again, this mediation was not significant when checked systematically β = .11, [-.02, .26], and again, in the interest of brevity, these analyses are not included in the tables.

5.5 Discussion

Present data confirmed that women are, on average, less sex-specific in their sexual arousal than men. Women had a smaller absolute difference between sexes in both genital arousal and pupil dilation (albeit for the latter not significantly so). Furthermore, women had stronger genital responses to their less-arousing sex. However, hypothesized sex differences were not at all confirmed was in pupil dilation to the less-arousing sex. Women further reported more trait empathy, but this increased empathy did not explain the sex difference in sexual arousal patterns. Furthermore, there was very limited (and not significant) evidence that men and women differed in their empathy towards male and female models, but this difference, too, did not affect the sex difference in sexual responses.

Thus, we found some support for the hypotheses that men and women differed in specificity of sexual arousal and in empathy; yet, these two sex differences were not further related to each other. As mentioned previously, there is sufficient evidence in the literature to suggest that sex differences exist in empathy and, separately, in sexual arousal patterns (Lam et al., 2012; Rieger et al., 2016; Stuijfzand et al., 2016). On this basis, the present study speculated that the two patterns may be related, but no evidence of such a link was found. Given the fact that the sample size was far above the recommended value, even given our conservative power analysis at .90, it appears that there is indeed no relationship between empathy, at least with respect to self-reported trait empathy, and sex differences in sexual arousal patterns. Perhaps, if we had more data for the per-video empathy data, they would elicit hypothesized patterns. Unfortunately, due to the COVID-19 pandemic impacting on testing, our participant sample fell short of the one recommended by the power analysis.

Further work should also consider other mechanisms and measurements. The mirror neuron system, which may be the basis for empathy in humans, is known to activate both when an individual performs an action, and when observing another individual performing the same action (Rizzolatti & Craighero, 2004). The strength of this "mirroring" can be predicted by psychometric measures of empathy (Jabbi et al., 2007). One fMRI study suggested that in men, degree of penile erection correlated with mirror neuron activation while viewing stimuli depicting heterosexual intercourse; thus, the same neurons may be activated if men experience sexual arousal themselves or view other people being sexually aroused (Mouras et al., 2008). Future research could therefore focus on employing a measure of mirror neuron activation, which has already been shown to relate in some fashion to sexual arousal in men, but has never been studied with regards to sex differences.

It is further possible that a sex difference in empathy (or in mirror neuron activation) is underpinned by a sex difference in androgen exposure, either during prenatal development or in the bloodstream as an adult (Chapman et al., 2006; van Honk et al., 2011). Therefore, another possible avenue for future research may be to employ a robust measure of androgen exposure – either in the bloodstream or during the prenatal period – alongside a measure of empathy – either via self-report or via direct measurement of the mirror neuron system – to examine whether these measures can illuminate a link between sex differences in empathy and sex differences in sexual arousal.

One limitation of the present study was that sexual response patterns measured with pupil dilation did not fully reflect those assessed via genital arousal. While patterns seen in pupil dilation did broadly match those seen in genital arousal (in that men showed a stronger response to one sex over the other), men also dilated more strongly than women to their less-preferred sex, which was the opposite of our hypothesis. While there is substantial evidence for the validity of genital arousal as a measure of sexual interest (Janssen, 2002; Suschinsky et al., 2009). Pupil dilation is a newer measure of sexual response, with a comparatively smaller body of evidence. In some studies it has produced results which match with those found through genital arousal (Rieger et al., 2015), but in others it has produced results which are not fully reflective of genital arousal patterns (Watts et al., 2018c). Because pupil dilation reflects not only sexual arousal, but also emotion, cognition, or non-sexual interest in stimuli (Bradley et al., 2008; Goldinger & Papesh, 2012), for some participants, pupillary responses could have been driven by factors other than sexual interest.

In sum, the present study suggested that women have less specific sexual arousal patterns, on average. In some analyses, women also reported having higher level of empathy than men. However, there was no evidence that these two patterns are linked with one another. There might be other, unknown personality characteristics or cognitive or emotional factors which explain the lack of specificity in the sexual responses of women.

Table 5.1 - *Multiple Regression Analyses for Sex and Composite Trait Empathy Predicting Absolute Difference to One Sex Over the Other in Genital Arousal (N = 274) and Pupil Dilation (N = 216).*

Step 1	**Absolute Difference in Genital Arousal**	**Absolute Difference in Pupil Dilation**
Variables	β	β
Sex[1]	-.39 [-.51, -.28]**	-.12 [-.26, .01] †
Step 2	**Absolute Difference in Genital Arousal**	**Absolute Difference in Pupil Dilation**
Variables	β	β
Sex[1]	-.42 [-.54, -.30]**	-.11 [-.25, .04] †
Composite Trait Empathy[2]	.08 [-.04, .20]	-.05 [-.19, .09]

Note. R^2*'s* for the two models are .15 and .02 in Step 1, and .16 and .02 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Males were coded as 0, females as 1. [2]Higher scores indicate greater self-reported composite trait empathy as measured on the Interpersonal Reactivity Index. † $p < .10$, ** $p < .01$.

Table 5.2 - *Regression Analyses for Sex and Composite Trait Empathy Predicting Sexual Response to the Less-Arousing Sex in Genital Arousal (N = 247) and Pupil Dilation (N = 216).*

Step 1	**Genital Response to Less-Arousing Sex**	**Pupil Dilation to Less-Arousing Sex**
Variables	β	β
Sex[1]	.62 [.52, .72]**	-.13 [-.26, .00] †
Step 2	**Genital Response to Less-Arousing Sex**	**Pupil Dilation to Less-Arousing Sex**
Variables	β	β
Sex[1]	.59 [.49, .70]**	-.14 [-.29, .00]*
Composite Trait Empathy[2]	.07 [-.04, .17]	.03 [-.11, .18]

Note. R^2*'s* for the two models are .38 and .017 in Step 1, and .38 and .018 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Males were coded as 0, females as 1. [2]Higher scores indicate greater self-reported composite trait empathy as measured on the Interpersonal Reactivity Index. † $p < .10$, * $p < .05$, ** $p < .01$.

Table 5.3 - *Multiple Regression Analyses for Sex and Per-Video Empathy Predicting Absolute Difference in Genital Arousal to One Sex Over the Other (N = 65).*

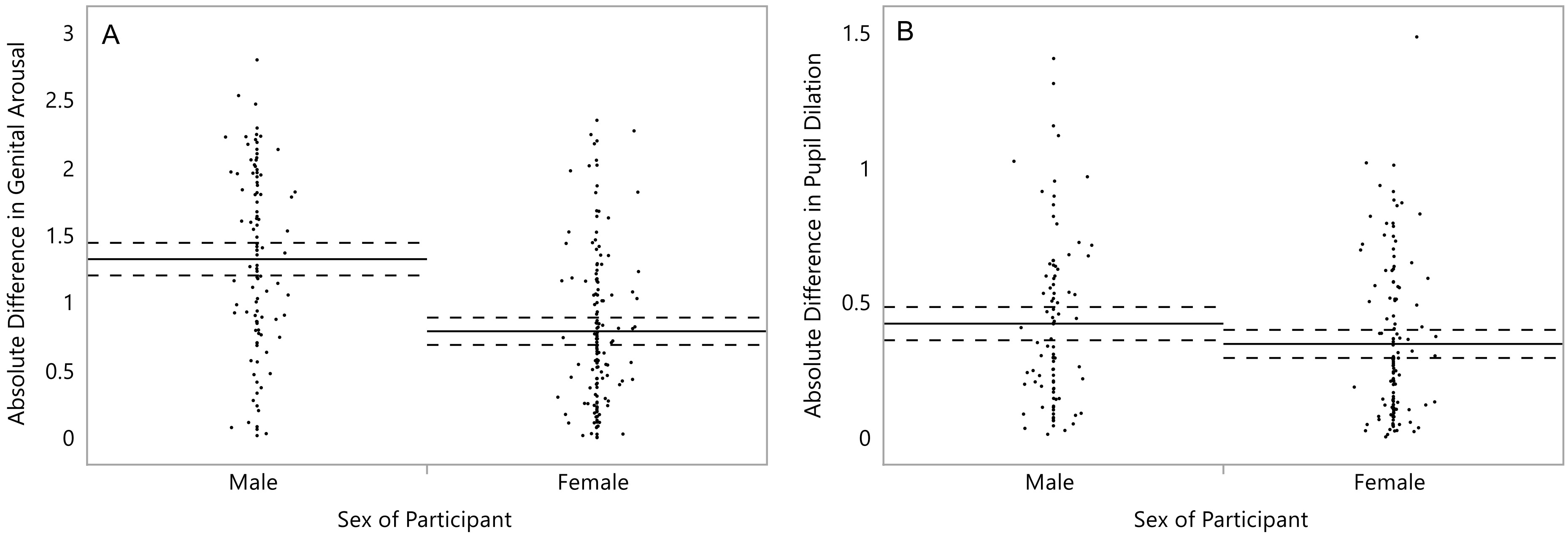
Step 1	**Absolute Difference in Genital Arousal**
Variables	β
Sex[1]	-.38 [-.61, -.14]**
Step 2	**Absolute Difference in Genital Arousal**
Variables	β
Sex[1]	-.38 [-.61, -.14]**
Absolute Difference in Per-Video Empathy[2]	-.11 [-.13, .34]

Note. R^2 for the model is .14 in Step 1 and .15 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Males were coded as 0, females as 1. [2]Higher scores indicate a greater difference between per-video empathy reported towards male stimuli and female stimuli, on average. ** $p < .01$.

Table 5.4 - *Multiple Regression Analyses for Sex and Per-Video Perspective-Taking Predicting Genital Response to the Less-Arousing Sex (N = 65).*

Step 1	**Genital Response to Less-Arousing Sex**
Variables	β
Sex[1]	.55 [.34, .76]**
Step 2	**Genital Response to Less-Arousing Sex**
Variables	β
Sex[1]	.51 [.30, .72]**
Minimum Per-Video Empathy[2]	.18 [-.03, .39] †

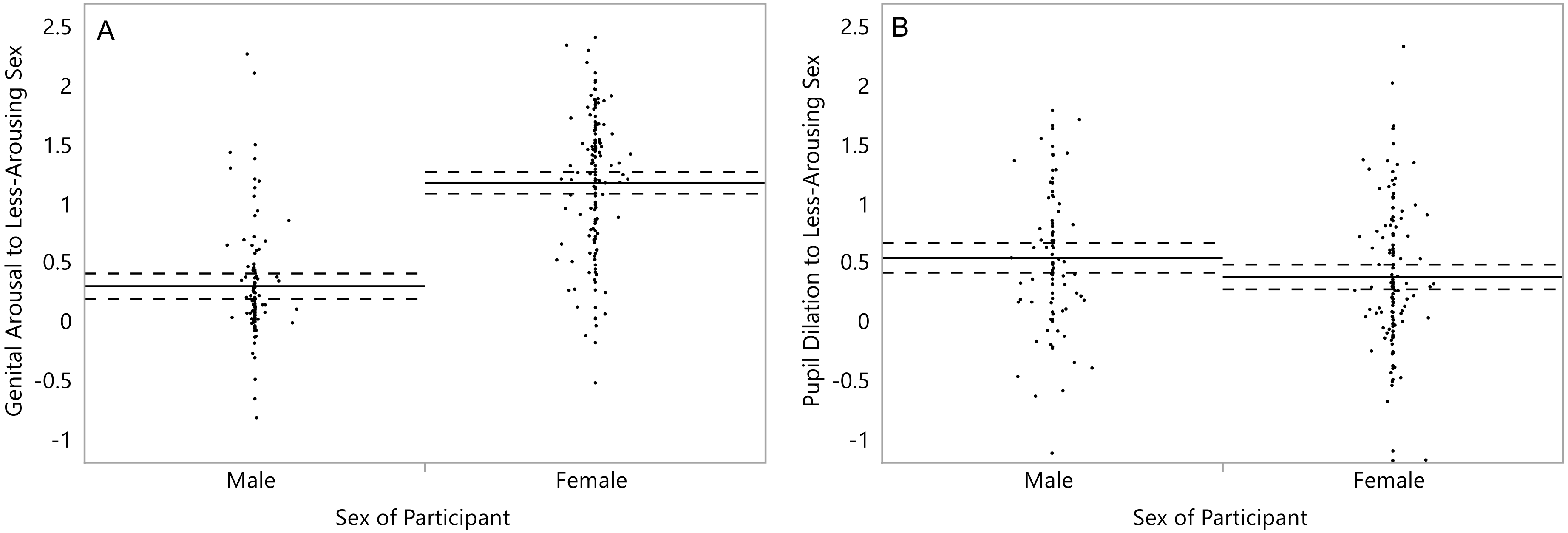
Note. R^2 for the model is .30 in Step 1 and .33 in Step 2. Numbers in brackets represent 95% confidence intervals of the standardized regression coefficient, β. [1]Males were coded as 0, females as 1. [2]Higher scores indicate a greater degree of empathy reported to male stimuli or female stimuli – whichever of the two was lower. † $p < .10$, ** $p < .01$.



p < .001, *t* = 6.68, *d* = .87 [.79, .94]

p = .07, *t* = 1.84, *d* = .26 [.22, .29]

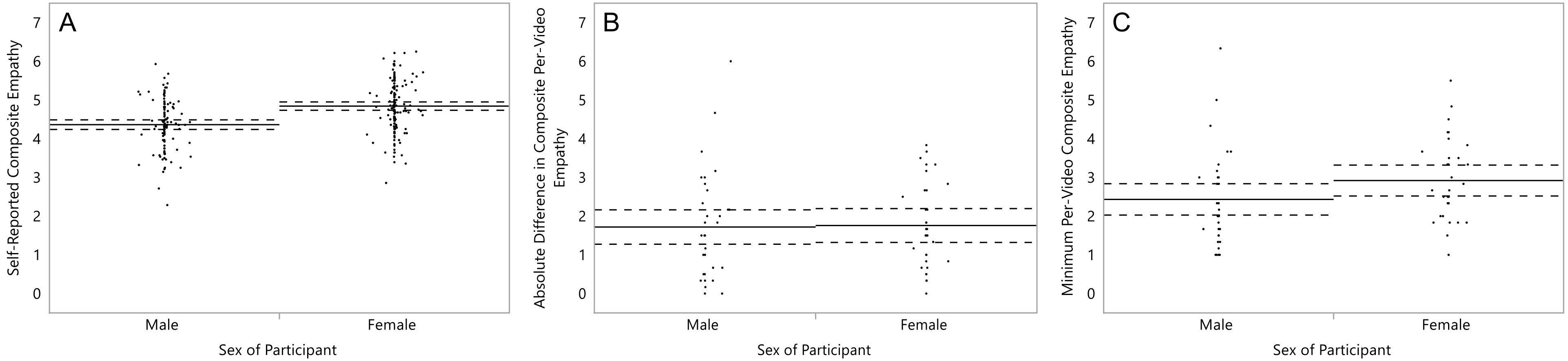
Figure 5.1 - Genital arousal and pupil dilation of men and women to sexual stimuli. Absolute difference between stimuli featuring males and stimuli featuring females in (A) genital responses of 102 men and 145 women and (B) pupil dilation of 90 men and 126 women. On the Y axis, scores reflect the absolute difference between sexual arousal to males and females, standardized within participants. Solid lines represent group means, and dashed lines their 95% confidence intervals. Dots represent participants' average scores. Captions are independent-samples t-tests.



p < .001, t = -12.27, d = -.1.59 [-1.66, -1.52]

p = .05, t = 1.94, d = .27 [.19, .35]

Figure 5.2 - Genital arousal and pupil dilation of men and women to sexual stimuli. Minimum arousal to stimuli featuring males and stimuli featuring females – whichever is lowest - in (A) genital responses of 102 men and 145 women and (B) pupil dilation of 90 men and 126 women. On the Y axis, scores reflect minimum arousal values, standardized within participants. Solid lines represent group means, and dashed lines their 95% confidence intervals. Dots represent participants' average scores. Captions are independent-samples t-tests.



p < .001, *t* = -5.75, *d* = -.72 [-.81, -.65]   *p* = .90, *t* = -.12, *d* = -.03 [-.33, .27]   *p* = .09, *t* = -1.71, *d* = -.43 [-.70, -.15]

Figure 5.3 - Self-reported empathy of men and women. Figure A depicts self-reported trait empathy as measured on the Interpersonal Reactivity Index for 111 men and 147 women. Figure B depicts the absolute difference in per-video empathy between stimuli featuring males and stimuli featuring females in 32 men and 33 women. Figure C depicts minimum per-video empathy to stimuli featuring males and stimuli featuring females – whichever is lowest – in 32 men and 33 women. On the Y axis, scores reflect mean self-reported composite empathy (A), absolute difference in composite per-video empathy (B) and minimum composite per-video empathy (C). Solid lines represent group means, and dashed lines their 95% confidence intervals. Dots represent participants' average scores. Captions are independent-samples t-tests.

Chapter 6 General Discussion

In general, the present work suggested that the sex difference in 2D:4D was significant in the predicted direction, with women having more female-typical (higher) 2D:4D in both hands than men. Sex differences in sexual arousal were also significant in the predicted directions, with women showing a less specific pattern of sexual arousal than men, both in terms of their genital arousal and pupil dilation. Despite this, the sex difference in 2D:4D did not explain the sex difference in sexual arousal patterns. Women also reported higher levels of empathy than men, on average, but again, empathy did not explain the sex differences in sexual arousal patterns.

In women, we found that homosexual women had more male-typical (specific) sexual arousal patterns than heterosexual women, and that homosexual women had more male-typical self-concepts and behaviours than heterosexual women. We did not confirm that homosexual women had more male-typical 2D:4D than heterosexual women, or that variation in 2D:4D explained in any way the difference between heterosexual and homosexual women in their gender nonconformity or sexual arousal.

In humans, there is indirect evidence of 2D:4D reflecting prenatal exposure to androgens. This includes that a sex difference in 2D:4D in foetal humans as early as 9 weeks gestation, which rules out the idea that social influences might be responsible (Galis et al., 2010; Malas et al., 2006). Through various means, we can also eliminate other potential explanations for a sex difference in 2D:4D, including the direct effect of sex chromosomes themselves, or the antimullerian hormone which prevents the development of the female genital tract in males. This leaves prenatal androgen exposure as the most possible explanation other than some as-yet unknown biological factor (Swift-Gallant et al., 2020). This assertion is further supported by evidence from conditions that affect an individual’s exposure to androgens, with individuals with CAH (which causes excessive androgen exposure) having lower 2D:4D, and individuals who have CAIS (which causes immunity to androgens) having higher 2D:4D (Brown et al., 2002b; Wisniewski et al., 2000).

Experimentally controlled causal evidence of a link between prenatal androgens and both 2D:4D and adulthood sexual behaviour does exist, but not from human models. At least in rodents, exposure to high levels of androgens prenatally masculinises the 2D:4D ratio and causes a predicted sex difference in 2D:4D (Zheng & Cohn, 2011). In primates, exposure to prenatal androgens can produce masculinisation of both sexual and non-sexual behaviours, although this study did not measure 2D:4D (Goy et al., 1988). However, it is possible that animal models – even those from closely-related primates – may not be applicable to humans because of biological differences between the species. One cannot rule this out, and so such studies must be interpreted with caution.

The lack of a link between sexual orientation and 2D:4D in women in the present sample is puzzling, as such an effect was previously confirmed in two separate meta-analyses (Grimbos et al., 2010; Hönekopp et al., 2007). Assuming for a moment that a true link between women's sexual orientation and 2D:4D exists, one could therefore reason that it would further explain the relationship between women's sexual orientation and other variables, such as their sexual arousal patterns. However, one of these meta-analyses (Grimbos et al., 2010) did report the possibility of publication bias in the literature, and speculated that the true link between 2D:4D and sexual orientation, if it exists, is possibly weaker on average than what is reported.

Even without considering this potential literature bias, it is also the case that the link between women's sexual orientation and 2D:4D is not perfect. For example, individual studies on sexual orientation and 2D:4D often find a relationship in one hand but not the other, and no explanation has yet been offered for this (Grimbos et al., 2010). It is therefore plausible that other, unknown factors could affect the precision of the relationship between prenatal androgen exposure and 2D:4D (McCormick & Carré, 2020). If true, this could explain not only the lack of link between sexual orientation and 2D:4D in the present research, but also the fact that 2D:4D did not explain the link between sexual orientation and other variables.

In general, it is possible that 2D:4D is sufficiently sensitive as a measure to detect a difference in prenatal androgen exposure between men and women, but not sufficiently sensitive to (at least reliably) detect an effect of sexual orientation in women. It is therefore possible that other factors which affect the length of digits (but not sexual orientation) or sexual orientation (but not the length of digits) create enough noise to render the measure ineffective except, perhaps, in very large samples.

If this is the case, could a stronger indicator of prenatal androgen exposure, stronger than 2D:4D, explain the masculinisation of sexual arousal patterns? This has been tried previously using gender nonconformity itself as an indicator of prenatal masculinisation (which should therefore also impact on sexual arousal patterns), but there was no conclusive evidence on either occasion that gender nonconformity and sexual arousal patterns are related (Rieger et al., 2016). Another possibility arising from the study on primates is that different aspects of sexual and non-sexual behaviour may be masculinised separately by exposure to androgens at different stages of foetal development (Goy et al., 1988). It is therefore possible that sexual orientation, sexual arousal patterns and 2D:4D are all masculinised by exposure to androgens during the prenatal period, just during independent “critical windows”.

Or, are male-typical versus female-typical sexual arousal patterns not even affected by prenatal androgenization, but instead by yet-unknown factors? One possibility is independent biological factors influencing sexual arousal patterns – these could be genetic, but independent of androgens, or they could be other hormones entirely, and thus not reflected in 2D:4D. For example, we know that sexual orientation in men is shaped to some degree by the fraternal birth order effect, where each successive male child in a family has a higher chance of developing a homosexual orientation because of his mother's growing resistance to H-Y antibodies produced during pregnancy (Bogaert et al., 2018). There is no similar effect known in women, but this at least presents evidence in favour of biological processes which would likely not be reflected in finger length ratios influencing sexual orientation (and therefore, potentially, on sexual arousal).

Another possible explanation for the masculinized sexual arousal patterns seen in homosexual women is social influences, although this seems very unlikely. Although we expect that a social explanation could be constructed to explain a sex difference in sexual arousal, in that men and women are unquestionably subject to different social environments, it seems unrealistic to us that any social factors would be the true reasons for differences in female vasocongestion patterns (which are often unknown to women), especially. Furthermore, any potential social explanation would fall short of feasibly being able to explain why arousal patterns would differ between heterosexual and homosexual women, or between subgroups of homosexual women. Thus, we favour the biological explanation even in the absence of any concrete evidence for what the biological factors may be.

We have also considered the possibility that part of the issue may be a weakness in how we measure sexual arousal in women, in particular our measure of vasocongestion, which assesses vaginal pulse amplitude. This measure was recently found not to significantly correlate to genital lubrication to sexual stimuli. Instead, lubrication correlated with subjective arousal (Sawatsky et al., 2018). This raises questions about its ecological validity as a measure of female sexual arousal, although it is still possible that vasocongestion is a necessary precursor to lubrication, and that lubrication only occurs when cues to potential penetration are present in the stimuli (Lalumière et al., 2020). Despite these limitations, we stress that vasocongestion produces a link between sexual orientation and specificity of arousal patterns in women which is similar to other measures such as pupil dilation, viewing time, and brain imaging (Ebsworth & Lalumière, 2012; Rieger et al., 2015; Safron et al., 2019). Vasocongestion is also highly male-specific in (male-to-female) transgender men, suggesting that this measure can pick up male-typical arousal patterns, if they exist, in biological females (Raines, In Press).

In theory, it is possible that other genital measures would produce more valid results. Two relatively new genital measures are clitoral blood volume and genital thermography (Huberman & Chivers, 2015; Suschinsky et al., 2020), both of which have already produced non-specific patterns of sexual arousal among heterosexual women, consistent with the patterns found with the measure used in present research, but have not yet been tested on homosexual women. Thus, one avenue for future research is to assess the use of multiple measures of sexual arousal in a sample of women of varying sexual orientations, to determine whether homosexual women have male-typical patterns of sexual responses, and to test whether (potential) multiple male-typical patterns of sexual arousal link to 2D:4D, or some other measure of prenatal androgenisation (some of which might be difficult to assess) such as anogenital distance or otoacoustic emissions (Barrett et al., 2018; McFadden & Pasanen, 1998).

Taken together, these separate lines of evidence indicate that the non-specific pattern of arousal seen in natal females is unlikely to be a product of the sexual arousal measures themselves. As such, although we speculate that vasocongestion may be one reason for our null results, it is also entirely possible that other measures would produce similar results.

Chapter 7 References

Allaway, H. C., Bloski, T. G., Pierson, R. A., & Lujan, M. E. (2009). Digit ratios (2D:4D) determined by computer-assisted analysis are more reliable than those using physical measurements, photocopies, and printed scans. *American Journal of Human Biology, 21*, 365-370. doi: 10.1002/ajhb.20892

Arnold, A. P. (2009). The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Hormones and behavior, 55*, 570-578. doi: 10.1016/j.yhbeh.2009.03.011

Attard-Johnson, J., Bindemann, M., & Ó Ciardha, C. (2016). Pupillary response as an age-specific measure of sexual interest. *Archives of Sexual Behavior, 45*, 855-870. doi: 10.1007/s10508-015-0681-3

Bailey, J. M. (2009). What is sexual orientation and do women have one? In D. A. Hope (Ed.), *Contemporary perspectives on lesbian, gay, and bisexual identities* (Vol. 54, pp. 43-63). New York, NY: Springer.

Bailey, J. M., Kim, P. Y., Hills, A., & Linsenmeier, J. A. W. (1997). Butch, femme, or straight acting? Partner preferences of gay men and lesbians. *Journal of Personality and Social Psychology, 73*, 960-973. doi: 10.1037/0022-3514.73.5.960

Bailey, J. M., Vasey, P. L., Diamond, L. M., Breedlove, S. M., Vilain, E., & Epprecht, M. (2016). Sexual orientation, controversy, and science. *Psychological Science in the Public Interest, 17*, 45-101. doi: 10.1177/1529100616637616

Bailey, J. M., & Zucker, K. J. (1995). Childhood sex-typed behavior and sexual orientation: A conceptual analysis and quantitative review. *Developmental Psychology, 31*, 43-55. doi: 10.1037/0012-1649.31.1.43

Barrett, E. S., Hoeger, K. M., Sathyanarayana, S., Abbott, D. H., Redmon, J. B., Nguyen, R. H. N., et al. (2018). Anogenital distance in newborn daughters of women with polycystic ovary syndrome indicates fetal testosterone exposure. *Journal of Developmental Origins of Health and Disease, 9*, 307-314. doi: 10.1017/s2040174417001118

Baumeister, R. F. (2000). Gender differences in erotic plasticity: The female sex drive as socially flexible and responsive. *Psychological Bulletin, 126*, 347-374. doi: 10.1037/0033-2909.126.3.347

Berenbaum, S. A., Bryk, K. K., Nowak, N., Quigley, C. A., & Moffat, S. (2009). Fingers as a marker of prenatal androgen exposure. *Endocrinology, 150*, 5119-5124. doi: 10.1210/en.2009-0774

Bogaert, A. F., Skorska, M. N., Wang, C., Gabrie, J., MacNeil, A. J., Hoffarth, M. R., et al. (2018). Male homosexuality and maternal immune responsivity to the Y-linked protein NLGN4Y. *Proceedings of the National Academy of Sciences, 115*, 302. doi: 10.1073/pnas.1705895114

Bossio, J., Suschinsky, K., Puts, D., & Chivers, M. (2014). Does menstrual cycle phase influence the gender specificity of heterosexual women's genital and subjective sexual arousal? *Archives of Sexual Behavior, 43*, 941-952. doi: 10.1007/s10508-013-0233-7

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*, 602-607. doi: 10.1111/j.1469-8986.2008.00654.x

Breedlove, S. M. (2010). Minireview: Organizational hypothesis: instances of the fingerpost. *Endocrinology, 151*, 4116-4122. doi: 10.1210/en.2010-0041

Breedlove, S. M. (2017). Prenatal influences on human sexual orientation: Expectations versus data. *Archives of Sexual Behavior, 46*, 1583-1592. doi: 10.1007/s10508-016-0904-2

Brown, W. M., Finn, C. J., Cooke, B. M., & Breedlove, S. M. (2002a). Differences in finger length ratios between self-identified “butch” and “femme” lesbians. *Archives of Sexual Behavior, 31*, 123-127. doi: 10.1023/A:1014091420590

Brown, W. M., Hines, M., Fane, B. A., & Breedlove, S. M. (2002b). Masculinized finger length patterns in human males and females with Congenital Adrenal Hyperplasia. *Hormones and Behavior, 42*, 380-386. doi: 10.1006/hbeh.2002.1830

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development, 53*, 413-425. doi: 10.2307/1128984

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the Empathy Quotient (EQ) and the “Reading the Mind in the Eyes” Test. *Social Neuroscience, 1*, 135-148. doi: 10.1080/17470910600992239

Chivers, M. L. (2017). The specificity of women’s sexual response and its relationship with sexual orientations: A review and ten hypotheses. *Archives of Sexual Behavior, 46*, 1161-1179. doi: 10.1007/s10508-016-0897-x

Chivers, M. L., Rieger, G., Latty, E., & Bailey, J. M. (2004). A sex difference in the specificity of sexual arousal. *Psychological Science, 15*, 736-744. doi: 10.1111/j.0956-7976.2004.00750.x

Chivers, M. L., Seto, M. C., & Blanchard, R. (2007). Gender and sexual orientation differences in sexual response to sexual activities versus gender of actors in sexual films. *Journal of Personality and Social Psychology, 93*, 1108-1121. doi: 10.1037/0022-3514.93.6.1108

Ciumas, C., Hirschberg, A. L., & Savic, I. (2008). High fetal testosterone and sexually dimorphic cerebral networks in females. *Cerebral Cortex, 19*, 1167-1174. doi: 10.1093/cercor/bhn160

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113. doi: 10.1037/0022-3514.44.1.113

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*, 71-100. doi: 10.1177/1534582304267187

Diamond, L. M. (2003). What does sexual orientation orient? A biobehavioral model distinguishing romantic love and sexual desire. *Psychological Review, 110*, 173-192. doi: 10.1037/0033-295X.110.1.173

Diamond, L. M. (2016). Sexual Fluidity in Male and Females. *Current Sexual Health Reports, 8*, 249-256. doi: 10.1007/s11930-016-0092-z

Doherty, R. W., Orimoto, L., Singelis, T. M., Hatfield, E., & Hebb, J. (1995). Emotional contagion: Gender and occupational differences. *Psychology of Women Quarterly, 19*, 355-371. doi: 0.1111/j.1471-6402.1995.tb00080.x

Ebsworth, M., & Lalumière, M. L. (2012). Viewing time as a measure of bisexual sexual interest. *Archives of Sexual Behavior, 41*, 161-172. doi: 10.1007/s10508-012-9923-9

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi: 10.3758/bf03193146

Galdikas, B. M. F. (1985). Subadult male orangutan sociality and reproductive behavior at Tanjung Puting. *American Journal of Primatology, 8*, 87-99. doi: 10.1002/ajp.1350080202

Galis, F., Ten Broek, C. A., Van Dongen, S., & Wijnaendts, L. D. (2010). Sexual dimorphism in the prenatal digit ratio (2D:4D). *Archives of Sexual Behavior, 39*, 57-62. doi: 10.1007/s10508-009-9485-7

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences, 23*, 573-587. doi: 10.1017/S0140525X0000337X

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21*, 90-95. doi: 10.1177/0963721412436811

Gottschalk, L. (2003). Same-sex sexuality and childhood gender non-conformity: A spurious connection. *Journal of Gender Studies, 12*, 35-50. doi: 10.1080/0958923032000067808

Goy, R. W., Bercovitch, F. B., & McBrair, M. C. (1988). Behavioral masculinization is independent of genital masculinization in prenatally androgenized female rhesus macaques. *Hormones and Behavior, 22*, 552-571. doi: 10.1016/0018-506X(88)90058-X

Grimbos, T., Dawood, K., Burriss, R. P., Zucker, K. J., & Puts, D. A. (2010). Sexual orientation and the second to fourth finger length ratio: A meta-analysis in men and women. *Behavioral Neuroscience, 124*, 278-287. doi: 10.1037/a0018764

Hamann, S., Stevens, J., Vick, J. H., Bryk, K., Quigley, C. A., Berenbaum, S. A., et al. (2014). Brain responses to sexual images in 46,XY women with complete androgen insensitivity syndrome are female-typical. *Hormones and Behavior, 66*, 724-730. doi: 10.1016/j.yhbeh.2014.09.013

Hines, M. (2009). Gonadal hormones and sexual differentiation of human brain and behavior. In D. W. Pfaff, A. P. Arnold, A. M. Etgen, S. E. Fahrbach & R. T. Rubin (Eds.), *Hormones, Brain and Behavior (Second Edition)* (pp. 1869-1910): San Diego: Academic Press.

Hines, M., Ahmed, S. F., & Hughes, I. A. (2003). Psychological outcomes and gender-related development in complete androgen insensitivity syndrome. *Archives of Sexual Behavior, 32*, 93-101. doi: 10.1023/a:1022492106974

Hines, M., Brook, C., & Conway, G. S. (2004). Androgen and psychosexual development: Core gender identity, sexual orientation, and recalled childhood gender role behavior in women and men with congenital adrenal hyperplasia (CAH). *Journal of Sex Research, 41*, 75-81. doi: 10.1080/00224490409552215

Hoffman, M. L. (1977). Sex differences in empathy and related behaviors. *Psychological Bulletin, 84*, 712. doi: 10.1037/0033-2909.84.4.712

Hönekopp, J., Bartholdt, L., Beier, L., & Liebert, A. (2007). Second to fourth digit length ratio (2D: 4D) and adult sex hormone levels: new data and a meta-analytic review. *Psychoneuroendocrinology, 32*, 313-321. doi: 10.1016/j.psyneuen.2007.01.007

Huberman, J. S., & Chivers, M. L. (2015). Examining gender-specificity of sexual response with concurrent thermography and plethysmography. *Psychophysiology, 52*, 1382–1395. doi: 10.1111/psyp.12466

Israel, E., & Strassberg, D. (2009). Viewing time as an objective measure of sexual interest in heterosexual men and women. *Archives of Sexual Behavior, 38*, 551-558. doi: 10.1007/s10508-007-9246-4

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage, 34*, 1744-1753. doi: 10.1016/j.neuroimage.2006.10.032

Jabbour, J., Holmes, L., Sylva, D., Hsu, K. J., Semon, T. L., Rosenthal, A. M., et al. (2020). Robust evidence for bisexual orientation among men. *Proceedings of the National Academy of Sciences, 117*, 18369-18377. doi: 10.1073/pnas.2003631117

Janssen, E. (2002). Psychophysiological measurement of sexual arousal. In M. W. Wiederman & B. E. Whitley, Jr (Eds.), *Handbook for conducting research on human sexuality* (pp. 139 –171). Mahwah N J: L. Erlbaum.

Johnson, K. L., Gill, S., Reichman, V., & Tassinary, L. G. (2007). Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology. *Journal of Personality and Social Psychology, 93*, 321-334. doi: 10.1037/0022-3514.93.3.321

Kinsey, A. C., Pomeroy, W. B., Martin, C. E., & Gebhard, P. H. (1953). *Sexual behavior in the human female*. Philadelphia, PA: Saunders.

Kraemer, B., Noll, T., Delsignore, A., Milos, G., Schnyder, U., & Hepp, U. (2006). Finger length ratio (2D: 4D) and dimensions of sexual orientation. *Neuropsychobiology, 53*, 210-214. doi: 10.1159/000094730

Kratochvíl, L., & Flegr, J. (2009). Differences in the 2nd to 4th digit length ratio in humans reflect shifts along the common allometric line. *Biology Letters, 5*, 643-646. doi: 10.1098/rsbl.2009.0346

Kurki, H. K. (2011). Pelvic dimorphism in relation to body size and body size dimorphism in humans. *Journal of Human Evolution, 61*, 631-643. doi: 10.1016/j.jhevol.2011.07.006

Lalumière, M. L., Sawatsky, M. L., Dawson, S. J., & Suschinsky, K. D. (2020). The empirical status of the preparation hypothesis: Explicating women's genital responses to sexual stimuli in the laboratory. *Archives of Sexual Behavior*, 1-20. doi: 10.1007/s10508-019-01599-5

Lam, C. B., Solmeyer, A. R., & McHale, S. M. (2012). Sibling relationships and empathy across the transition to adolescence. *Journal of Youth and Adolescence*, 1657-1670. doi: 10.1007/s10964-012-9781-8

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19*, 42-58. doi: 10.1162/jocn.2007.19.1.42

Li, G., Kung, K. T. F., & Hines, M. (2017). Childhood gender-typed behavior and adolescent sexual orientation: A longitudinal population-based study. *Developmental Psychology, 53*, 764-777. doi: 10.1037/dev0000281

Lippa, R. A. (1991). Some psychometric characteristics of gender diagnosticity measures: Reliability, validity, consistency across domains, and relationship to the Big Five. *Journal of Personality and Social Psychology, 61*, 1000-1011. doi: 10.1037/0022-3514.61.6.1000

Lippa, R. A. (2003). Are 2D:4D finger-length ratios related to sexual orientation? Yes for men, no for women. *Journal of Personality and Social Psychology, 85*, 179-188. doi: 10.1037/0022-3514.85.1.179

Lippa, R. A. (2005). Sexual orientation and personality. *Annual Review of Sex Research, 16*, 119-153. doi: 10.1080/10532528.2005.10559831

Lippa, R. A. (2008). Sex differences and sexual orientation differences in personality: Findings from the BBC Internet survey. *Archives of Sexual Behavior, 37*, 173-187. doi: 10.1007/s10508-007-9267-z

Lippa, R. A. (2012). Effects of sex and sexual orientation on self-reported attraction and viewing times to images of men and women: Testing for category specificity. *Archives of Sexual Behavior, 41*, 149-160. doi: 10.1007/s10508-011-9898-y

Lippa, R. A. (2015). *Assessing sexual orientation and category specificity in a representative sample of 2,825 United States adults.* Paper presented at the Puzzle Of Sexual Orientation Meeting, Lethbridge, Alberta, Canada.

Lippa, R. A. (2020). Interest, Personality, and Sexual Traits That Distinguish Heterosexual, Bisexual, and Homosexual Individuals: Are There Two Dimensions That Underlie Variations in Sexual Orientation? *Archives of Sexual Behavior, 49*, 607-622. doi: 10.1007/s10508-020-01643-9

Malas, M. A., Dogan, S., Evcil, E. H., & Desdicioglu, K. (2006). Fetal development of the hand, digits and digit ratio (2D:4D). *Early Human Development, 82*, 469-475. doi: 10.1016/j.earlhumdev.2005.12.002

McCarthy, M. M., Herold, K., & Stockman, S. L. (2018). Fast, furious and enduring: Sensitive versus critical periods in sexual differentiation of the brain. *Physiology & Behavior, 187*, 13-19. doi: 10.1016/j.physbeh.2017.10.030

McCormick, C. M., & Carré, J. M. (2020). Facing off with the phalangeal phenomenon and editorial policies: A commentary on Swift-Gallant, Johnson, Di Rita and Breedlove (2020). *Hormones and Behavior, 120*, 104710. doi: 10.1016/j.yhbeh.2020.104710

McFadden, D., & Pasanen, E. G. (1998). Comparison of the auditory systems of heterosexuals and homosexuals: Click-evoked otoacoustic emissions. *Proceedings of the National Academy of Sciences, 95*, 2709-2713. doi: 10.1073/pnas.95.5.2709

Mestre, M. V., Samper, P., Frías, M. D., & Tur, A. M. (2009). Are women more empathetic than men? A longitudinal study in adolescence. *The Spanish Journal of Psychology, 12*, 76-83. doi: 10.1017/S1138741600001499

Meyer-Bahlburg, H. F., Dolezal, C., Baker, S. W., & New, M. I. (2008). Sexual orientation in women with classical or non-classical congenital adrenal hyperplasia as a function of degree of prenatal androgen excess. *Archives of Sexual Behavior, 37*, 85-99. doi: 10.1007/s10508-007-9265-1

Morris, J. A., Jordan, C. L., & Breedlove, S. M. (2004). Sexual differentiation of the vertebrate nervous system. *Nature Neuroscience, 7*, 1034-1039. doi: 10.1038/nn1325

Motta-Mena, N. V., & Puts, D. A. (2017). Endocrinology of human female sexuality, mating, and reproductive behavior. *Hormones and Behavior, 91*, 19-35. doi: 10.1016/j.yhbeh.2016.11.012

Mouras, H., Stoléru, S., Moulier, V., Pélégrini-Issac, M., Rouxel, R., Grandjean, B., et al. (2008). Activation of mirror-neuron system by erotic video clips predicts degree of induced erection: an fMRI study. *NeuroImage, 42*, 1142-1150. doi: 10.1016/j.neuroimage.2008.05.051

Ökten, A., Kalyoncu, M., & Yariş, N. (2002). The ratio of second-and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Early Human Development, 70*, 47-54. doi: 10.1016/S0378-3782(02)00073-7

Olweus, D., & Endresen, I. M. (1998). The importance of sex-of-stimulus object: Age trends and sex differences in empathic responsiveness. *Social Development, 7*, 370-388. doi: 10.1111/1467-9507.00073

Palmer, C. (1989). Is rape a cultural universal? A re-examination of the ethnographic data. *Ethnology, 28*, 1-16. doi: 10.2307/3773639

Pascual-Sagastizabal, E., Azurmendi, A., Sánchez-Martín, J. R., Braza, F., Carreras, M. R., Muñoz, J. M., et al. (2013). Empathy, estradiol and androgen levels in 9-year-old children. *Personality and Individual Differences, 54*, 936-940. doi: 10.1016/j.paid.2013.01.019

Pasterski, V. L., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2005). Prenatal hormones and postnatal socialization by parents as determinants of male-typical toy play in girls with congenital adrenal hyperplasia. *Child Development, 76*, 264-278. doi: 10.1111/j.1467-8624.2005.00843.x

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891. doi: 10.3758/BRM.40.3.879

Puts, D., & Motta-Mena, N. V. (2018). Is human brain masculinization estrogen receptor-mediated? Reply to Luoto and Rantala. *Hormones and Behavior, 97*, 3-4. doi: 10.1016/j.yhbeh.2017.07.018

Putz, D. A., Gaulin, S. J. C., Sporter, R. J., & McBurney, D. H. (2004). Sex hormones and finger length: What does 2D:4D indicate? *Evolution and Human Behavior, 25*, 182-199. doi: 10.1016/j.evolhumbehav.2004.03.005

Rahman, Q. (2005). Fluctuating asymmetry, second to fourth finger length ratios and human sexual orientation. *Psychoneuroendocrinology, 30*, 382-391. doi: 10.1016/j.psyneuen.2004.10.006

Rahman, Q., & Wilson, G. D. (2003). Sexual orientation and the 2nd to 4th finger length ratio: evidence for organising effects of sex hormones or developmental instability? *Psychoneuroendocrinology, 28*, 288-303. doi: 10.1016/S0306-4530(02)00022-7

Raines, J. (In Press). Genital Sexual Arousal Patterns of Transgender Men. *Scientific Reports*.

Rieger, G., Cash, B. M., Merrill, S. M., Jones-Rounds, J., Dharmavaram, S. M., & Savin-Williams, R. C. (2015). Sexual arousal: The correspondence of eyes and genitals. *Biological Psychology, 104*, 56-64. doi: 10.1016/j.biopsycho.2014.11.009

Rieger, G., Holmes, L., Watts-Overall, T. M., Gruia, D. C., Bailey, J. M., & Savin-Williams, R. C. (2020a). Gender Nonconformity of Bisexual Men and Women. *Archives of Sexual Behavior*. doi: 10.1007/s10508-020-01766-z

Rieger, G., Linsenmeier, J. A. W., Gygax, L., & Bailey, J. M. (2008). Sexual orientation and childhood gender nonconformity: Evidence from home videos. *Developmental Psychology, 44*, 46-58. doi: 10.1037/0012-1649.44.1.46

Rieger, G., Linsenmeier, J. A. W., Gygax, L., Garcia, S. C., & Bailey, J. M. (2010). Dissecting "gaydar": Accuracy and the role of masculinity-femininity. *Archives of Sexual Behavior, 39*, 124-140. doi: 10.1007/s10508-008-9405-2

Rieger, G., & Savin-Williams, R. C. (2012). The eyes have it: Sex and sexual orientation differences in pupil dilation patterns. *PLoS ONE, 7*, e40256. doi: 10.1371/journal.pone.0040256

Rieger, G., Savin-Williams, R. C., Chivers, M. L., & Bailey, J. M. (2016). Sexual arousal and masculinity-femininity of women. *Journal of Personality and Social Psychology, 111*, 265-283. doi: 10.1037/pspp0000077

Rieger, G., Watts-Overall, T. M., Holmes, L., & Gruia, D. C. (2020b). Gender Nonconformity of Identical Twins with Discordant Sexual Orientations: Evidence from Video Recordings. *Archives of Sexual Behavior*. doi: 10.1007/s10508-020-01709-8

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192. doi: 10.1146/annurev.neuro.27.070203.144230

Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin, 132*, 98. doi: 10.1037/0033-2909.132.1.98

Safron, A., Sylva, D., Klimaj, V., Rosenthal, A. M., & Bailey, J. M. (2019). Neural responses to sexual stimuli in heterosexual and homosexual men and women: Men's responses are more specific. *Archives of Sexual Behavior, 49*, 433-445. doi: 10.1007/s10508-019-01521-z

Safron, A., Sylva, D., Klimaj, V., Rosenthal, A. M., Li, M., Walter, M., et al. (2017). Neural correlates of sexual orientation in heterosexual, bisexual, and homosexual men. *Scientific Reports, 7*, 41314. doi: 10.1038/srep41314

Sawatsky, M. L., Dawson, S. J., & Lalumière, M. L. (2018). Genital lubrication: A cue-specific sexual response? *Biological Psychology, 134*, 103-113. doi: 10.1016/j.biopsycho.2018.02.003

Slaughter, L., Brown, C. R. V., Crowley, S., & Peck, R. (1997). Patterns of genital injury in female sexual assault victims. *American Journal of Obstetrics and Gynecology, 176*, 609-616. doi: 10.1016/S0002-9378(97)70556-8

Slettevold, E., Holmes, L., Gruia, D. C., Nyssen, C. P., Watts-Overall, T. M., & Rieger, G. (2019). Bisexual men with bisexual and monosexual genital arousal patterns. *Biological Psychology, 148*, 107763. doi: 10.1016/j.biopsycho.2019.107763

Stuijfzand, S., De Wied, M., Kempes, M., Van de Graaff, J., Branje, S., & Meeus, W. (2016). Gender Differences in Empathic Sadness towards Persons of the Same- versus Other-sex during Adolescence. *Sex Roles, 75*, 434-446. doi: 10.1007/s11199-016-0649-3

Suschinsky, K. D., Dawson, S. J., & Chivers, M. L. (2020). Assessing gender-specificity of clitoral responses. *The Canadian Journal of Human Sexuality, 29*, 57-64. doi: 10.3138/cjhs.2019-0061

Suschinsky, K. D., & Lalumière, M. L. (2011). Prepared for anything? An investigation of female genital arousal in response to rape cues. *Psychological Science, 22*, 159-165. doi: 10.1177/0956797610394660

Suschinsky, K. D., Lalumière, M. L., & Chivers, M. L. (2009). Sex differences in patterns of genital sexual arousal: Measurement artifacts or true phenomena? *Archives of Sexual Behavior, 38*, 559-573. doi: 10.1007/s10508-008-9339-8

Swift-Gallant, A., Johnson, B. A., Di Rita, V., & Breedlove, S. M. (2020). Through a glass, darkly: Human digit ratios reflect prenatal androgens, imperfectly. *Hormones and Behavior, 120*, 104686. doi: 10.1016/j.yhbeh.2020.104686

Sylva, D., Safron, A., Rosenthal, A. M., Reber, P. J., Parrish, T. B., & Bailey, J. M. (2013). Neural correlates of sexual arousal in heterosexual and homosexual women and men. *Hormones and Behavior, 64*, 673-684. doi: 10.1016/j.yhbeh.2013.08.003

Timmers, A. D., Bouchard, K. N., & Chivers, M. L. (2015). Sexual arousal patterns to audiovisual stimuli among bisexual women. *Journal of Bisexuality, 15*, 180-203. doi: 10.1080/15299716.2015.1022924

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review, 17*, 72-86. doi: 10.1177/1088868312461308 van Honk, J., Schutter, D. J., Bos, P. A., Kruijt, A.-W., Lentjes, E. G., & Baron-Cohen, S. (2011). Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proceedings of the National Academy of Sciences, 108*, 3448. doi: 10.1073/pnas.1011891108

Wallien, M. S., Zucker, K. J., Steensma, T. D., & Cohen-Kettenis, P. T. (2008). 2D: 4D finger-length ratios in children and adults with gender identity disorder. *Hormones and Behavior, 54*, 450-454. doi: 10.1016/j.yhbeh.2008.05.002

Watts, T. M., Holmes, L., Raines, J., Orbell, S., & Rieger, G. (2018a). Finger Length Ratios of Identical Twins with Discordant Sexual Orientations. *Archives of Sexual Behavior*, 1-10. doi: 10.1007/s10508-018-1262-z

Watts, T. M., Holmes, L., Raines, J., Orbell, S., & Rieger, G. (2018b). Gender nonconformity of identical twins with discordant sexual orientations: Evidence from childhood photographs. *Developmental Psychology, 54*, 788. doi: 10.1037/dev0000461

Watts, T. M., Holmes, L., Raines, J., Orbell, S., & Rieger, G. (2018c). Sexual arousal patterns of identical twins with discordant sexual orientations. *Scientific Reports, 8*, 1-7. doi: 10.1038/s41598-018-33188-2

Williams, T. J., Pepitone, M. E., Christensen, S. E., Cooke, B. M., Huberman, A. D., Breedlove, N. J., et al. (2000). Finger-length ratios and sexual orientation. *Nature, 404*, 455. doi: 10.1038/35006555

Wisniewski, A. B., Migeon, C. J., Meyer-Bahlburg, H. F., Gearhart, J. P., Berkovitz, G. D., Brown, T. R., et al. (2000). Complete androgen insensitivity syndrome: long-term medical, surgical, and psychosexual outcome. *Journal of Clinical Endocrinology & Metabolism, 85*, 2664-2669. doi: 10.1210/jcem.85.8.6742

Xu, Y., Norton, S., & Rahman, Q. (2019). Early life conditions and adolescent sexual orientation: A prospective birth cohort study. *Developmental Psychology, 55*, 1226-1243. doi: 10.1037/dev0000704

Xu, Y., & Zheng, Y. (2015). The digit ratio (2D: 4D) in China: A meta-analysis. *American Journal of Human Biology, 27*, 304-309. doi: 10.1002/ajhb.22639

Zheng, Z., & Cohn, M. J. (2011). Developmental basis of sexually dimorphic digit ratios. *Proceedings of the National Academy of Sciences, 108*, 16289-16294. doi: 10.1073/pnas.1108312108

Zucker, K. J., Bradley, S. J., Oliver, G., & Blake, J. (1996). Psychosexual development of women with congenital adrenal hyperplasia. *Hormones and Behavior, 30*, 300-318. doi: 10.1006/hbeh.1996.0038