Maynooth University
National University of Ireland Maynooth

Automatic Gender Binary Beliefs and their Role in Gender Inequality

Thesis submitted to the Department of Psychology, Faculty of Science and Engineering, in part fulfilment of the requirements for the degree of Doctor of Philosophy, Maynooth University

Aoife Cartwright, B.Sc. (Hons.), M.Sc.

February 2019

Thesis Supervisor: Dr Bryan Roche

Head of Department: Prof Andrew Coogan

Acknowledgements

I would first like to thank my supervisor, Dr Bryan Roche. Bryan, I really am so grateful. I would never have gotten through the last few years without your support. I don't know how many times I subjected you to a bout of panic or rage or tedium, and you somehow always managed to listen patiently and find a way to get me back on track. You have been an incredibly generous and knowledgeable supervisor, but also a very kind friend. Thank you so much.

I am also very thankful for the other mentors I was lucky enough to have during my PhD. To Noeline Blackwell, thank you for your invaluable insights around the NGO sector, for helping me contextualize so much, and for being so generous with your time and advice over the past few months. To Professor Julia Heiman, thank you the opportunity to visit your lab and for all of the useful career, research, and life advice you shared while I was there.

Thank you as well to the staff and students in the Maynooth University Psychology Department for their advice and assistance over the years. I am especially grateful to the final year students who assisted with data collection and who generously contributed their time and insights to this project: thank you very much Ciara, Elaine, Emer, Lauren, Roisin, and Grainne.

Ian, I'm not sure what category to put you in! But thank you from the bottom of my heart for being my sounding board, advisor, and friend for the past 5 years. I appreciate it all so much.

To my wonderful friends – Lisa, Dave, Elliott, Francesca, Marc, Yasmin, Ríoghán, and Claudia – I am so, so, so grateful for each of you. Thank you all for your

humour, kindness, optimism, and encouragement over the past few years. In particular, I want to thank Francesca for making our home such a safe and fun place and for being such a support throughout this whole process. I also owe a big thank you to all the Maynooth pals (Dylan, Jamie, Alan, Joanne, Conor, and the other postgrads) for keeping me sane, for sharing this with me, and for making the past few years so special.

To my boyfriend Alan, thank you so much for coming into my life and being such a comfort during this final stretch.

Finally, I want to thank my family. They understand more than anyone how difficult this process has been, and I know without a doubt I never have been able to pursue or finish this without their encouragement. Most of all, I would like to thank my amazing, compassionate, smart, patient, hilarious Mam. Thank you for raising me to value inclusion and equality, for teaching me to love learning, for never losing faith in me, and for always encouraging me in whatever I choose to do. You are my inspiration and my main source of support and I am so very grateful for everything you've done for me. I love you.

Abstract

The social construction of gender-as-binary plays an increasingly central role within gender equality research and activism. Despite its importance, however, there remain few empirical tools for assessing binarist beliefs, practices, or behaviours at the individual level. This thesis sought to address this gap by, first, proposing a new way to operationalize the gender binary, second, introducing the Implicit Relational Assessment Procedure (IRAP) as a potentially valid and reliable psychometric measure of automatic binarist beliefs. The current work had three broad aims: First, it aimed to conduct a comprehensive survey of self-reported and automatic binary beliefs in a sample of young Irish adults. Nine separate studies were conducted in the service of this ($N = 602$), which together provided clear evidence that gender is indeed structured in a binary, oppositional way (i.e., women are feminine but not masculine and men are masculine but not feminine). They also provided novel insights into the relational structure of gender roles, and the asymmetrical way in which we "gender" men relative to women. A second aim of this work was to examine the role of the binary in inequality. To this end, studies examined the relationship between IRAP effects and responses on three different measures of gender discrimination and prejudice: gendered hiring preferences (Chapter Three), androcentric bias (Chapter Four), and sexual harassment proclivity (Chapter Five). While studies in Chapter Three provided strong evidence that the binarisation of gender underpins discrimination in occupational contexts, effects in the remaining chapters were comparably weaker. Lastly, this thesis took the novel step of gathering a sufficiently large IRAP dataset for a set of pooled analyses. These analyses (Chapter Six) strengthened the conclusions drawn around the strength of the biases on the binary IRAP, provided novel insights into the magnitude and nature of gender

differences on this measure, and shed light on some of its psychometric properties. Overall, these findings have a number of broad implications: First, they add to the growing empirical literature around binarist ideologies and their role in gender inequality. Second, they inform our understanding of how gender is structured, and elucidate the oppositional, relational, and asymmetrical way in which gender categories are framed. Third, they reveal the IRAP to be an adequately reliable and valid tool for quantifying gender binary biases. Fourth, and last, they attest to the automaticity of binary beliefs and thus the centrality of the binary within gender cognition more broadly.

Chapter 1

General Introduction

1.1 Introduction

The social construction of gender-as-binary is the centre of considerable feminist debate, research and activism. To date, however, the field of psychology has struggled with the operationalization and assessment of binarist ideologies. The current work aims to address this by, first, introducing a new psychometric tool for quantifying gender binary beliefs and, second, investigating the potential role of these biases in gender-based discrimination and prejudice. Before introducing the specific studies planned for this thesis, a comprehensive overview of the gender binary and its measurement will be provided. First, this literature review will very briefly outline modern gender theory and the paradigmatic shift towards systematic rather than essentialist models of gender. Binary gender systems will then be introduced and described, and the link between binarist frameworks, social behaviour, and widespread inequality clearly outlined. This review will then turn its focus to measurement, looking at the various different fields, sub-disciplines, and theoretical perspectives that have analysed the binary at various levels of analysis. Finally, a novel way to conceptualise and measure binary biases will be put forward.

1.1.1 Epistemological Statement

Before reviewing the literature relevant to the current thesis, it is important to first clarify my own epistemological position. I am a cisgender, White woman and have lived in Ireland my whole life. I consider myself a feminist and would particularly align myself with the philosophy of Radical Feminism. While it can be difficult to find a unifying definition of any strand of feminism in the literature, Radical Feminism can be described as the effort to "challenge, change and ultimately end patriarchy" (Hooks, 2004, p.108). One of the guiding goals of radical feminist research in recent decades has been to expose gender structures and understand the

ways in which they relate to gender-based violence and inequality (see Mackay, 2015). This work was conducted to analyse one particular aspect of Western gender structures identified in the literature: the gender binary. As it is the focus of the current work, it should be clear that I am adopting a radical feminist interpretation of the gender binary as a meaningful, socio-historically constructed and malleable system for organising behaviour (e.g. Butler, 1990).

1.2 Gender: From Essentialist to Systemic Models

For a long time, gender was conceptualised as a natural and essential reality. From as far back as Ancient Greece to the advent of Christianity, differences between women and men have been attributed in varying degrees to innate, immutable, divine, or predetermined factors (Butler, 1988; Gelman, 2003). The introduction of Darwinian theory in the late 19th century and the extension of the sexual dimorphism framework to human behaviour provided a new biological backing to historical arguments, moving the immutable origins of sexual difference away from the divine and onto the genetic (Dar-Nimrod & Heine, 2011; DeLamater & Hyde, 1998). Or, as argued by Bem (1993), away from "God's grand creation [and on] to its scientific equivalent: Evolution's grand creation" (p. 68). Even the earlier waves of feminist activism assumed – and indeed endorsed – essential differences between the sexes. During feminism's first wave, for instance, prominent Suffragettes such as Elizabeth Cady Stanton and Emmeline Pankhurst based their platforms on traditional notions of women's moral and spiritual superiority over men (see Heilmann, 2011). Similarly, many strands of liberal or "different-but-equal" feminism (most popular during the second wave) largely sought to achieve equality through the celebration of women and men's distinct, natural, and complementary strengths (e.g., Daly, 1985; Friedan, 1963; Gilligan, 1982).

Towards the beginning of the early 1980s, however, the emerging schools of post-structuralist, phenomenological, postmodern, and social constructionist philosophy began to influence those working within the context of gender and women's activism. Judith Butler's highly influential theory of gender *performativity*, for instance, challenged the taken-for-granted assumption of gender as innate, instead reframing it as something which is continually performed, constructed, and reproduced in daily interaction (1990, 2002). Candace West and Don Zimmerman (1987) introduced the similar concept of *doing gender* to describe the myriad ways in which historical roles and norms are enacted (or "done") at both individual and social levels. Dorothy E. Smith (1987) and Patricia Collins (1990) further interrogated the origins and validity of essentialism by introducing feminist standpoint theory, or the broad argument that all knowledge (including gender knowledge) should be considered socially informed, situated, and governed. Of particular significance to the current work, however, was the emergence of power-based models, or those which examined the relationship of gender to structures of oppression. R.W. Connell's Gender Order Theory (1982, 1985), for instance, proposed that gender is essentially a hierarchical system that governs the differential distribution of rights and social resources to men (particularly traditionally masculine men) over women. According to Connell, the categories "man", "woman", "masculinity" and "femininity" are thus not biological truths, but rather different positions within a gender order.

Together, these models and frameworks brought about significant change within feminist scholarship and marked the beginning of what is now generally considered the third wave. With the rise of radical feminism (and the goal "to question everything": Baer, 2011), gender itself gradually became the subject of

analysis and the centre of a new standalone discipline: Gender Studies. Essentialism was slowly replaced by the social and power-based accounts, and many historic concepts were either abandoned, reinterpreted, or radically reconceptualised. For example, the construct of biological sex was now seen as separable from culturally constructed gender roles (De Beauvoir, 1979; Daly & Wilson. 1983), while ideologies around "natural" sexualities, roles or relationship structures were reframed as tools for enforcing women's subordination (Rich, 1980). Importantly, the concept of *patriarchy* was introduced during this time to broadly challenge and dismantle the assumption of a natural gender order, and to acknowledge socio-historically constructed systems of male dominance (Firestone, 1970; Dworkin & MacKinnon, 1988; Lorde, 1986).

Around this same period, empirical research from the biological, psychological and social sciences began to cast further doubt on naturalistic explanations of sex/gender difference. Findings from anthropology and history, for instance, provided evidence of both temporal and cross-cultural variation in gender roles, thereby challenging the assumption of a universal gender (Mead, 1935, 1963). More recently, meta-analytic reviews from psychology and biology have found considerable variation in the strength, reliability, and replicability of gender differences across disciplines, often finding more variation within gender groups than between (Fausto-Sterling, 2000; Fine, 2010; Richardson, 2013). These findings are of course further complicated when different moderators, mediators, or explanatory frameworks are applied to the data and/or taken into account (e.g., neuroplasticity, epigenetics, socialisation, etc.: see Hyde, Bigler, Joel, Tate & Van Anders, 2019). Even the validity of the more basic anatomical or genetic categories (i.e., male or female, XX or XY) has been called into question in recent years by

statistics showing that approximately 1.7% of individuals are born intersex or with a disorder of sexual development (i.e., with ambiguous genitalia or an atypical sex-chromosome configuration: see Richards, Bouman & Barker, 2018). According to Oakley (2016), sexual difference should be considered more apparent than real, given that men and women ultimately "have the same body ground-plan… neither the phallus nor the womb are organs of one sex only: the female phallus (the clitoris) is the biological equivalent of the male organ, and men possess a vestigial womb, whose existence they may well ignore until it causes enlargement of the prostate gland in old age." (p. 26). While it is important to note these combined bodies of work do not (and are not intended to) discredit the role of biology in gender difference, they do demonstrate the complexity inherent in these categories and thus the shortcomings of an entirely radical essentialist framework (Dreger, 1999; Gelman, 2003).

1.3 The Gender Binary

Within mainstream gender studies, it is thus now increasingly common to view gender an ideological system rather than a biological fact (Ferree, Lorber, and Hess 1999; Lorber 1994; Nakano Glenn 1999; Ridgeway 1997; Ridgeway and Smith-Lovin 1999; Risman 1998). Though there are and have been many systems for organising gender throughout history (see Fausto-Sterling, 2000 and Vasey & Bartlett, 2007 for reviews of gender systems across cultures and time periods), in the context of the modern West, gender is now generally described using one broad and collective term: the gender binary. The binary (also referred to as binarism) encompasses and unites the myriad social, historic, ideological, epistemological, and institutional practices that categorise individuals into different groups (i.e., women and men), as well as those which organise systems of social relations, discrimination,

and inequality on the basis of that difference (Ridgeway & Correll, 2004). Research into the binary is therefore vast and includes multiple different bodies of work from a wide range of disciplines (e.g., anthropology, history, sociology, psychology, and so on: Hyde et al., 2019; Ridgeway, 2011). For brevity, and because this work is concerned with the measurement of the binary and its role in social discrimination and inequality, the below review will organise and structure material in the following way: First, a review of binary gender roles will be provided, along with a summary of the various different responsibilities, attributes, skills, and abilities that have become associated with different groups under the binary. Next, the myriad ways in which these roles underpin (and are shaped by) social behaviour will be outlined. Finally, the systems of power organised on the basis of real and complete gender difference will be reviewed.

1.3.1 Gender Roles

To look first at how individuals are categorised in a binary system, gender in the West has been summarized as a "two *and only two*" system (Garfinkel, 1967; Kessler & McKenna, 1978). Within the binary, there are two genders (female and male) that are divided on the basis of primary and secondary sexual characteristics (i.e., sex chromosomes, gonads, genitalia, and typical pubertal development: Richards et al., 2015). Each gender is associated with distinct sets of attributes, skills, traits and abilities, which together are generally referred to as masculine and feminine gender roles (Eagly, 1987; Risman, 2004). Decades of research suggests that masculinity tends to be associated more readily with *agentic* attributes (e.g., independence, dominance, aggression, ambition, logic, self-sufficiency, or leadership potential), while femininity is defined more by *communal* attributes (e.g., sensitivity, kindness, affection, interdependence, submission, and interpersonal care:

Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1994; Eagly & Karau, 2002; Newport, 2001; Gaucher, Friesen, & Kay, 2011; Madera, Hebl, & Martin, 2009; Moscatelli, Ellemers, Menegatti, & Rubini, 2016; Rudman, Moss-Racusin, Phelan, & Nauts, 2012; Spence & Buckner, 2000; Williams & Best, 1990).

There is thus an inherent "complementarity" to Western gender roles, with masculine strengths corresponding to feminine weaknesses and vice versa (Jost & Kay, 2005; Eagly & Karau, 2002). As described by Butler (1990), "one is only one's gender to the extent that one is not the other gender" (p. 22), and indeed a large body of evidence suggests that gender categories are relationally defined. That is, that which is feminine is also *not-masculine*, and that which is masculine is also *not-feminine*. For instance, numerous empirical studies show that attributes rated as typical or desirable in one gender are often considered atypical or undesirable in the other (Diekman & Eagly, 2000; Glick & Fiske, 1996; Hall & Carter, 1999; Spence & Helmreich, 1978; Williams & Best, 1990; Wood, Christensen, Hebi & Rothberger, 1997). Moreover, people generally prefer and expect men and women to behave in a manner consistent with their gender role (Carli, 2001; Rudman & Glick, 2001), and a large body of research documents the negative social consequences associated with role deviations (e.g., McCreary, 1994; O'Neil, Helms, Gable, David, & Wrightsman, 1986; Pleck, 1981; Smiler, 2004). In this way, binarised gender roles are not merely shared cultural beliefs about men and women, but rather the normative, prescriptive and proscriptive scripts for acceptable male and female behaviour (Keonig, 2018).

Eagly's widely-cited Social Role Theory (SRT: Eagly, 1987; Eagly & Karau, 2002) provides a useful explanatory framework for the ways in which gender roles underpin broader social structures and responsibilities. According to the SRT, society's expectation and preference that men and women do not behave like the

"opposite sex" explains societal gender stereotyping, segregation, and differentiation across numerous spheres. Specifically, Eagly and her colleagues propose that the socialization of men (but not women) as agentic and women (but not men) as communal explains why men are evaluated as more naturally suited for some roles (e.g., leadership, STEM careers, politics, or roles involving money or economic oversight) and women for others (e.g., teaching, care work, or assistant roles). Because gender is oppositionally structured, there is also an inherent *in*congruity between women and agentic roles and men and communal roles, potentially explaining why women and men are often negatively evaluated when they occupy positions deemed incongruous with their gender (male nurses or female CEOs; Eagly & Mladinic, 1989; Heilman, 1983, 2012).

The SRT also proposes that the observation of men and women in these different roles leads to the naturalisation of gender stereotypes within a culture over time (i.e., the belief that gender differences arise because of natural and immutable differences between women and men). While most contemporary feminist and gender theorists reject purely naturalistic explanations of gender roles or the argument that biology necessitates women's subordination (e.g., Butler, 2002), essentialist beliefs, models and practices remain widespread (Prentice & Miller, 2006; Smiler & Gelman, 2008). Studies suggest, for example, that people tend towards essentialist explanations of gender differences even when presented with additional information (e.g., about socialization or inequality-related factors: Brescoll, Uhlmann & Newman, 2013; Brescoll & LaFrance, 2004; Haslam, Rothschild & Ersnt, 2000, 2002; Yoder, Fischer, Kahn & Groden, 2007). Moreover, these effects seem to be strongest for domains or behaviours that have historically been very gender-differentiated (e.g., aggression or helping: Brescoll, Uhlman &

Newman, 2013; Eagly & Wood, 1991), thus supporting the SRT argument that roles and stereotypes tend to become naturalized over time. Essentialism is also evident at a societal level. Within contemporary medical and psychological research, for example, it is still common practice to use "sex" and "gender" interchangeably, and indeed to collect and analyse gender information in a binary way (i.e., using only male/female response options: Bouchel et al., 2018; Fine, 2010; Fraser, 2018). Sex and gender can be similarly conflated in the legal system, with many laws and public policies defining gender on a biological rather than socio-cultural basis (if at all: see Van Anders, Schudson, Abed, Beischel, Dibble, Gunther et al., 2017).

The fusion of biological sex with sociocultural gender under the binary (i.e., the assumed concordance between genetic makeup, anatomy, and social roles) has a range of significant consequences for women and men, some of which will be explored later in this thesis. However, the "two *and only two*" assumption also has considerable consequences for individuals who fall outside of the binary; that is, those who cannot be meaningfully accounted for within a binary framework of knowledge (Murjan & Bouman, 2015; Young, 1994). These include, but are not limited to, individuals who are transgender (those who have a different gender identity than what was assigned at birth), gender-fluid (those with multiple or flexible gender identities, or an identity that may change over time), non-binary/genderqueer (those whose identity does not align reliably with either gender category), and intersex individuals or those born with a Disorder of Sexual Development (DSD; i.e., people with ambiguous sexual characteristics or with a condition affecting the development or expression of sexual characteristics: Hyde et al., 2018 and Hegarty et al., 2018 for comprehensive reviews of the above). Though the estimated combined prevalence of these phenomena is between 2-8% (depending

on the criteria applied: Bouchel et al., 2018; Olyslager & Conway, 2007) individuals who do not fit a binary definition of gender still experience significant legal and medical difficulties (see Richards & Barker, 2015). It remains near impossible to live legally without a binary gender identity, for example, with nearly all societies requiring *either* a male or female gender to be assigned at birth (Fraser, 2018). Furthermore, it is still mainstream practice to surgically normalize those who present as ambiguous, including intersex individuals or people born with DSD (Roen, 2015). While some societies have provided legal recognition for "third gender" groups (e.g., the *Hijra* of Pakistan or *Fa'fafine* of Samoa: Lorber, 1997; Vasey & Bartlett, 2007) or legal/medical avenues for changing gender (e.g., gender reassignment surgery), feminists and activists remain conflicted about whether these practices serve to reinforce or dismantle rigid systems of gender categorization (Fine, 2010; Hird, 2000; Richards & Barker, 2015).

1.3.2 Gender Relations

The above review clearly outlines the impact of the binary on individuals, and the many different rights, roles, and responsibilities associated with different gender groups under the binary. In addition to prescribing gender norms, however, the binarisation of gender is also considered highly significant for social and interpersonal behaviour (Butler, 1988; Sedgwick, 1991). According to Ridgeway and Correll (2004), the binary acts as a sort of axiomatic, omni-relevant and primary frame for organising social relations, both within and between gender groups. Lorber (2011) similarly argues for the relevance of gender to our daily interactions, comparing it to that of water to a fish: highly significant and yet so deeply ingrained in our consciousness that it is almost invisible. While gender is similar to (and indeed interacts with) many other systems of social categorization such as race,

class, ability, or age, gender is considered an especially powerful system for organizing behaviour with particularly well-established patterns of interaction (Eagly & Steffen, 1984). This is primarily because of the frequency with which men and women interact, as well as the centrality of gender to social identity and categorization (Kachel, Steffens & Niedlich, 2016).

A wealth of social psychological research demonstrates the significance of gender to social and identity processes. Studies of interpersonal recognition and categorization, for instance, show that gender information tends to be processed significantly faster than information pertaining to age, race, or ethnicity (see Young & Burton, 2018). Analyses suggest gender has similar primacy in identity development and construction, with studies reliably identifying gender as the most central or highly ranked identity dimension (i.e., people tend to identify as a man or women before they identify as a Black person, or Irish person, and so on: Aboud, 1984; Jones & McEwan, 2000: Haim & Ruble, 2010). Developmental research indicates these abilities emerge very early in life and before the capacity to categorize or identify according to other dimensions (e.g., by race, language group, nationality, or religion; Bigler & Liben, 2007; Kinzler, Shutts & Correll, 2010; Levy & Killen, 2008). Rudimentary categorization abilities found have been found among infants as young as 6 months old, for instance, while identity processes are thought to begin around 2 years of age (see Baron, Schmader, Cvencek, & Meltzoff, 2014). Often described as "gender-typing" (Bem, 1981), these cognitive and behavioural processes are considered significant because they form the basis of gender stereotyping and identity processes, and thus gender knowledge more broadly (Bussey & Bandura, 1999; Martin et al., 2002; Richards & Barker, 2015; Turner, 2000).

This interactive process of learning to adhere to or behave in accordance with gender norms is broadly referred to as *gender socialization* (Leaper & Friedman, 2007). Gender role socialization is theorized to begin very early in social development, with children exposed to vastly different forms of play, parenting, activities, and educational instruction depending on their gender (e.g., Richards & Barker, 2015). While the specifics of gender socialisation will be discussed later in this chapter, broadly speaking, gender theorists note the ways in which boys are encouraged towards activities based on problem-solving, competition, and physical strength, and girls towards those involving aesthetics, beauty, cooperation, and domestic abilities (e.g., Bigler, 1995; Gelman, Taylor, & Nguyen; 2004; Leaper, 1994; Leaper, Breed, Hoffman, & Perlman, 2002; Maccoby, 1998; Marsh & Kleitman, 2003; Thompson & Zerbinos, 1995).

The enforcement and enactment of norms is believed to continue across the lifespan through the "social-relational context" of daily interactions (Ridgeway, 2011). According to Butler (2002), there are clear and historic rules for all gender relations under the binary, which are both predicated on and reproduce traditional gender roles. Between women and men, for example, a large body of work suggests these interactions remain largely *heteronormative* in nature; that is, consistent with the traditional belief that masculine men and feminine women form a natural heterosexual dyad (Warner, 1991). Heteronormativity was a concept that was developed to describe the societal expectation and normalization of heterosexuality (Herek, 2004, Rubin, 1993; Rich, 1980), as well as the more general conflation of sex, gender roles, gender identity and sexual orientation under the binary (e.g., Habarth et al., 2019). This conflation, which has been referred to as the *sex/gender system* (Rubin, 1993) or the *sex-gender-sexuality complex* (Segal, 2006), is argued to

create an overarching "script" for gender relations, where the dominant male is expected to pursue, manage, control, or provide for the passive or subordinate female (Habarth, 2015; Jackson, 2006; Kitzinger, 2005; Massey, 2009). While outside of the scope of this thesis, evidence abounds that men and women adhere to and endorse this script in a variety of contexts, including romantic and sexual relationships (Byers, 1995; Masters et al., 2013), the workplace (Losert, 2008), educational settings (Ward & Schneider, 2009), and platonic friendships (Cronin, 2015).

The binary also sets out clear rules for *intra*-gender relations, or relations between men and between women. Several typologies and frameworks have been developed to describe the different gender dynamics which may operate within gender groups (e.g., Kimmel, 1993, 1996; Lorber, 1998), including the previously mentioned Gender Order Theory (GOT; Connell 1987). In brief, the GOT proposes that gender in the West is hierarchically structured, with traditional dominant masculinity at the top, followed by minority or non-traditional masculinity (e.g., gay men, non-White men, effeminate men, etc.) and lastly by traditional and non-traditional femininity. Similar to theories of gender performativity, Connell argues that masculinity is something which is perceptually achieved and reconstructed through male-male social interactions. Citing evidence from discursive, cognitive, behavioural and sociological studies, proponents of this model note the many ways in which men are socialized to *perform* their masculinity (e.g., through displays of aggression, control and dominance towards those lower down in the hierarchy: Connell & Messerschmidt, 2005; Connell et al., 1982; Donaldson, 1993; Hunt, 1981; Kessler et al., 1992; Willis, 1977). This "hegemonic masculinity" framework has been used to explain why male friendships can be characterized by competition, emotional detachment, or the shared objectification and/or sexualisation of women

(Bird, 1996), and to account for the relatively severe social backlash experienced by men when they violate prescribed gender norms (e.g., engage in same-sex behaviour or pursue a stereotypically feminine interest or occupation: Gelman et al., 2004; McCreary, 1994; Smiler, 2006). At a societal level, sociologists have extended this framework to a range of gendered social phenomena, including sexual and gender-based violence, gender disparities in health and help-seeking behaviour, and the disproportionately high rates of suicide in men (see Jewkes et al., 2015 for a review).

Gender relations between women are theorized to be somewhat different, although they function to reproduce and regulate gender norms in much the same way (Harding, 1983). According to Coates (2015), female relations both reflect and shape the communal feminine role, and the ways in which women are socialised towards emotional awareness, verbal ability and interpersonal skills. Studies of female peer groups, for instance, show they are characterized by significantly more emotional intimacy, self-disclosure, reciprocal support, frequency of contact, and longevity than their male equivalents (Fischer & Oliker, 1983; Hey, 1996; Johnson & Aries, 1983; O'Connor, 1992; Wright, 1988; Wright & Scanlon, 1991). Moreover, when compared to men, studies show that women experience low rates of gender role stress (i.e., reduced pressure to adhere to traditional feminine gender roles; Eisler, Skidmore & Ward, 1988; Levant, 2011) and experience weaker social punishment for gender role transgressions (McCreary, 1994; Smiler, 2004). Women do still encounter intra-gender backlash and conflict when they violate gender norms, however, particularly those related to compliance, submission, and sexual purity. For instance, studies show that women are equally (and sometimes more) likely than men to negatively evaluate women who are sexually promiscuous, agentic,

overweight, or choose not to have children (Acker, 2009; Campbell, 2004; Mavin, 2006; Mavin, Williams & Grandy, 2014).

1.3.3 Gender-Power Dynamics

While the research reviewed thus far clearly outlines how the binary separates the genders and underpins social relations, it is important to now connect these literatures to the more pressing question of gender inequality. As mentioned previously, many feminists root women's oppression within our broader gender ideology and frameworks of knowledge (Brownmiller, 1975; Butler, 2002; Connell & Messerschmidt, 2005; Dworkin, 1975; Rich, 1980). While the micro and macro social processes that enact equality are likely numerous, according to Bem (1996), separating the genders lays the groundwork for inequality in two main ways: First, it allows for the establishment of gender orders or hierarchies over time (i.e., through the "gendering" of various socially-valued traits, roles, and abilities). Second, it leads to the relegation of certain experiences, viewpoints, and knowledge over others, and thus to the gradual normalisation of one gender over the other. Within the specific spatio-temporal context of the modern West, the binary thus enables two main types of male privilege: (1) male *supremacy*, and (2) male *centrality*.

To look first at male supremacy, there is a wealth of evidence to show that the current gender order prioritizes men. The most recent statistics by the World Health Organization, United Nations, and the World Economic Forum shows that women fare worse than men in nearly every nation in the world on nearly all metrics of gender parity (UN, 2018; WEF, 2018; WHO, 2007). These include reduced access to education, underrepresentation in political or decision-making spheres, and lower economic participation and success. These results hold true for Europe, with the most recent Gender Equality Index Report finding superior outcomes for men across

work, money, knowledge, time, power, health, violence, and "intersecting inequality" domains (European Institute for Gender Equality, 2017). Ireland ranks slightly higher than the EU average overall (69.5 out of a possible 100) but is notably lower in the domains of economic/political power (48.6) and gender-based violence (25.6) (see also the Women and Men in Ireland Report: Central Statistics Office, 2016).

According to Ridgeway (2011), men also have ideological supremacy under patriarchy, in that they tend to be more readily associated with culturally-valued traits, abilities, and characteristics. Studies generally show that higher status groups tend to be viewed as possessing more socially valued attributes (Berger, Rosenholtz & Zelditch, 1980; Cheng, Tracy & Henrich, 2010; Eagly & Wood, 1982), and a large body of research suggests an alignment between cultural values and masculine stereotypes. Leadership stereotypes, for example, are defined by many of the same agentic traits associated with traditional masculinity (e.g., independence, assertiveness, etc.: see Koenig, Eagly, Mitchell & Ristikari, 2011 for a meta-analysis). Studies have found an overlap between masculine stereotypes and those for other lucrative or high-status roles, including a successful scientist, politician, doctor, athlete, or financial manager (Burton, Barr, Fink & Bruening, 2009; Carli, 2016; Himmelstein, 2016; Schneider & Bos, 2014). By contrast, feminine stereotypes have been shown to overlap with roles associated with lower status and/or income (e.g., nursing, teaching, care work, full-time parenting, or assistant roles: Conway, Pizzamiglio & Mount, 1996; Eagly & Steffen, 1984; Glick, Wilk & Perreault, 1995).

Further evidence for ideological male supremacy comes from the fact that gender inequality is remarkably resistant to outside intervention. To look at the

example of occupational inequality, for instance, research shows that it has persisted in spite of significant socioeconomic transformations, including women's entry into the workforce, improved educational attainment, and the introduction of many forms of antidiscrimination legislation (Ridgeway, 2011; Verniers & Vala, 2018). In Ireland alone, for instance, considerable gender imbalances remain among socially valued and senior employment categories, with women comprising just 22.2% of Teachtaí Dála (CSO, 2016) and 16% of Chief Executives positions in Irish financial institutions (Central Bank, 2018). This is despite the fact that roughly comparable percentages of women and men now enter the workforce (59.5% and 69.9% respectively: CSO, 2016), and indeed a higher percentage of women now complete third-level education (55.1% relative to 42.9%). Given that the historic argument that women are inherently ill-suited to or disinterested in senior management roles does not tend to be supported by empirical research (see Eagly & Karau, 2002), the above gender barriers would appear to be ideological.

Gender-power dynamics have been further examined in the context of non-binary identities. As mentioned previously, there are several groups of people who fall outside of the binary system, including transgender, non-binary, gender-fluid, and intersex individuals (Hyde et al., 2018; Richards & Barker, 2015). In addition to the aforementioned legal exclusion that can be encountered by non-binary individuals (i.e., not being able to legally work, marry, or access medical and social resources without a male or female identity), research suggests these groups are subject to profound social discrimination and inequality. For instance, when compared to their "cisgender" counterparts (i.e., those whose gender identity aligns with the gender assigned to them at birth), studies show these individuals experience disproportionally high occupational prejudice, sexual violence and harassment,

physical or verbal violence or threats, and psychological and physical health issues, as well as significant underrepresentation in spheres of power or influence (e.g., politics and media: see Grant et al., 2010; Richards, Bouman & Barker, 2017 for reviews of the above evidence). According to Butler (2004), this intolerance is not merely a reaction to deviance or perceived difference; it is a powerful regulatory tool for reinforcing the binary. She notes, "this violence emerges from a profound desire to keep the order of binary gender natural or necessary; to make of it a structure, either natural or cultural, or both, that no human can oppose, and still remain human" (p. 35).

The above review demonstrates how the gender binary may underpin a broad system of male or masculine supremacy. As mentioned, however, another body of work examines a separate yet important manifestation of male privilege under patriarchy: male *centrality* (Bem, 1996). Centrality in this case refers to the societal tendency to normalize maleness or masculinity, or to place a masculine point of view at the centre of society (Hegarty & Pratto, 2006). Feminist theory has long acknowledged the andro (i.e., male) centric nature of Western society and cultural practices (Gillman, 1911; De Beauvoir, 1979). As argued by Monique Wittig, "only one gender exists: the feminine… The masculine is not the masculine, but the general" (1985, p. 8), and indeed the literature does suggest we tend to normalise men relative to women. Within the English language, for instance, it is still conventional to use masculine universals (e.g., *man*, *mankind, guys,* etc.) and generics (i.e., he/his/him in place of she/they; Silveira, 1980), add a feminine suffix to distinguish a female role or occupation (e.g., *actor* vs *actress*; see Bodine, 1975; Hyde, 1984), or default to a masculine gender in gender-neutral or ambiguous contexts (Lambdin et al., 2003). Similarly, in Western cultures at least, default

symbols for “human” and “man” are generally interchangeable (e.g., those used for social media avatars or safety/healthcare signs; see Bailey & LaFrance, 2016), and studies indicate that men remain the prototypical exemplar for several social categories (Eagly & Kite, 1987; Stroessner, 1996; Ng, 2007; Zárate & Smith, 1990). More broadly, feminists have critiqued the dominance of the male standpoint or “epistemological stance” across social systems of knowledge and representation (Harding, 1986; Mackinnon, 1982; Smith, 1988). This extends to the overreliance of male narrators, experiences or terminology in historical records, literature, law and media (Bem, 1993; Hegarty, 2006; Smith & Choueit, 2010). Equally, it includes modern scientific and academic practices (e.g., generalizing findings from male-only research studies to women, or explaining gender differences in terms of female attributes, differences, or deficits: see Bruckmuller et al., 2012; Hegarty & Buechel, 2006; Pratto et al., 2007).

While seemingly benevolent, analyses suggest these practices have significant consequences for women’s equality. As argued by both Bem (1996) and Lucal (1999), women’s demarcation as the “other” under patriarchy means their experiences, contributions, and needs tend to be relegated relative to men’s. Studies show that the conventional use of masculine generics, symbols and gender-blind terms significantly increases the likelihood of a male bias and/or male-centric imagery or behaviours (e.g., attributing a male gender to an ambiguous person or piece of text, drawing or selecting an image of a man, writing stories about a man: etc.; for a review, see Braun, Sczesny, & Stahlberg, 2005). This practice of viewing women as “gender-specific” and men as “gender-neutral” (also referred to as the *default male hypothesis*: Smith & Zaraté, 1992) is evidenced across numerous contemporary social practices. Accessible examples can be found in sport, where

women's sports are generally framed as a niche or subtype that is in some way separate from mainstream sporting activity (Pilcher & Whelehan, 2004) and indeed female athletes and teams have been shown to receive significantly less state funding, media coverage, and corporate sponsorships than their male counterparts (Kian & Hardin, 2009). Analyses of political or scientific funding practices highlight a similar tendency to under-fund work conducted by women or into women's experiences (see Ceci & Williams, 2011 for a comprehensive review), or indeed work concerning issues of particular relevance to women (e.g., maternal healthcare or sexual violence: Fisk & Atun, 2009; Waechter & Van Ma, 2015).

It is important to note at this stage that the gender-power dynamics reviewed above do not preclude the existence of positive female stereotypes. Stereotype theory and research shows there are positive and negative attributes associated with both femininity and masculinity (e.g., nurturing but nagging, independent but aggressive, and so on), and indeed studies show that femininity may in fact be more positively valenced in the abstract (Glick & Fiske, 2001). Often dubbed the "women-are-wonderful" effect, studies of stereotype and trait ratings reliably show a pro-feminine bias, for both male and female participants (Eagly & Mladinic, 1989; Eagly, Mladinic & Otto, 1991; Hosoda & Stone, 2000). While seemingly paradoxical, many radical and third-wave feminists consider this entirely coherent within a patriarchal value system, and similar to the recent rise in popularity of neoliberal (i.e., "different-but-equal") feminism among conservative groups (see Babar, 2017). According to Rudman (2001), women are not devalued in all spheres under patriarchy, just those related to status or power (i.e., lucrative or decision-making roles in society).

1.3.4 The Gender Binary: A Summary

Gender under the binary may be summarized as follows: with regards to the categorical make up, there are two genders (males and females) who are characterized by distinct sets of attributes, skills, and abilities (masculine and feminine gender roles). Gender is relationally and oppositionally structured, such that men (but not women) are expected to be agentic, and women (but not men) are expected to be communal. Role violations are generally undesirable, and behaviour or identities which fall outside of the binary are typically met with an array of negative social and legal consequences. The social-relational context plays an important role in the gender binary, for both the construction and maintenance of gender roles. Referred to broadly as gender role socialization, gender normativity appears to be regulated through a variety of inter and intra-gender relations across the lifespan (e.g., early childhood experiences, familial relationships, peer groups, educational settings, romantic relationships, the workplace, and so on). The binary is considered significant not only for the limits or boundaries it places on individuals and their behaviour, but for the role it plays in shaping and maintaining inequality. Specifically, given that (a) men and women are framed as completed opposites, and (b) men have historically been associated more socially valued attributes, then the binarization of gender essentially lays the groundwork for a patriarchal and androcentric social order.

1.4 Measuring the Gender Binary: Traditional Methods

Despite the centrality of the binary within contemporary gender theorizing, the measurement of binarist ideologies, practices, and beliefs remains somewhat abstract. While there are many different tools for measuring components of the gender system (e.g., stereotypes, identity, socialization practices, etc.), there are few direct measures of binarism that could be used in an experimental or individual-level

analysis. In order to give some background to the experimental methodology proposed in this thesis, an overview will be provided of the existing approaches for studying the social construction of gender-as-binary. This includes methodologies or approaches that assess or focus on how gender categories are related or framed; the accepted roles for men and women; how ambiguous or non-binary identities are interpreted and dealt with in society; the direct or indirect fusion between sex, gender, and sexuality; and the relationship of the binary to male dominance. For ease, these various bodies of research will be reviewed under the following analytic category headings: (1) culture; (2) history; (3) development; (4) language; and (5) beliefs.

1.4.1 Analysing Culture

Cultural analysis broadly refers to the study of a culture's beliefs, practices, values, symbols, or assumptions (see King, 2016, for an introduction). Among gender theorists, such analyses are frequently used to examine the current or historic influences of society on gender role maintenance and socialization, as well as the binarization of society in accordance with traditional gender roles. Though there are many methodological and data sources that may be used for a cultural analysis of the binary, national or global statistics have been especially useful in exposing the gendered divisions in society. In the context of labour, for example, research reliably shows that the majority of unpaid or low-income care roles are occupied by women (i.e., child-rearing or acting as a full-time carer to a family member or partner) while most high-risk and physically dangerous jobs are held by men (e.g., forestry, mining and fishing: UN, 2018). Similar gender divisions have been found in the context of health, with socio-policy analyses noting the significant impact of gender on life expectancy, disease prevalence and outcomes, mental health diagnoses, and health-

related behaviours (WHO, 2007). Societal crime and violence are also highly gendered, with Irish and global statistics revealing a significant percentage of crimes are perpetrated by men and against men (CSO, 2016; UNODC, 2018). This is with the exception of sexual or intimate partner violence, which is predominantly perpetrated by men against women.

Cultural analyses have also shed light on the socio-political systems which may contribute to or reinforce these divisions, such as the law. Since the 1950s, feminist scholars have argued that the law plays an important practical role in inequality, both in subordinating women and maintaining the binary (Brown, 1990; Harris, 1990; MacKinnon, 2005). Across the world, the vast majority of lawmakers and law enforcers are men (see Baer, 2011), and many feminists argue this has resulted in an inherently gender-biased, heteronormative, and male-dominated legal system (Fineman, 2010; 2013). Feminist analyses of the law have examined, for instance, the comparably low conviction rates for crimes disproportionately affecting women (e.g., rape and domestic abuse: Anderson, 2003; Burgess-Jackson, 1996; Dowrkin, 1997), or the legal frameworks around contraception, abortion or maternal healthcare (Barnett, 1998; Smart, 2002). Feminists have also analysed how the law may reinforce traditional binary gender roles for both women and men. In many jurisdictions, for instance, it remains common to have laws or provisions that are gender-specific (i.e., afford special privileges or punishments depending on the gender of the perpetrator or victim: see Baer, 2011). Some examples from Irish law would be the constitutional protections afforded to families or women in the home (Bunreacht na hÉireann, 1937), or that incest only carries a maximum sentence of life imprisonment only when the perpetrator is male (Sexual Offences Act, 2017).

The media has been subject to similar scrutiny for its role in both shaping and maintaining gender ideologies. Analyses have identified several common media practices which reify traditional gender norms, including the reliance on conventional gender tropes (e.g., women's domestication or the conflation of abuse and romance: Byerly & Ross, 2006), underrepresentation or misrepresentation of women in leading, powerful or desirable roles (e.g., heroes, scientists, presidents: Leaper, Breed, Hoffman, & Perlman, 2002; Thompson & Zerbinos, 1995), or the tendency to use women (or their suffering) as a plot device to further a male character's development (Thornham, 2007). Within feminist media analysis, scholars have been particularly critical of the tendency to objectify and sexualise women, or to depict their lives and bodies through the perspective of the "male gaze" (Carilli & Campbell, 2005). According to Laura Mulvey (1975, 2004), the widespread practice of exposing idealized, sexualised versions of women's bodies contributes significantly to women's subordinated status in society, primarily through reinforcing the idea that women exist as sexual objects for male consumption. Gender norms may also be reproduced through advertising (e.g., targeting cleaning, beauty or dieting adverts to women only, etc.: Gallagher, 2013; Signorielli, 2001), and in the reporting of news and current events. A recent report demonstrated that just 10% of news stories focus on women or women's issues, and that women comprise just 20% of the spokespeople or experts interviewed (UNESCO, 2018). Gender stereotypes or norms can also be subtly reinforced through storytelling "templates" in media. In the context of domestic and sexual violence, for instance, a recent analysis by Berns (2017) explored how male aggression and violence is often normalized in the news through (a) shifting the focus onto the victim; (b) reframing the violence as a display of over-eager or misconstrued romance/sexual arousal; or

(c) describing the behaviour as aberrant rather than normative. Other media analyses have focused on the gender divides in the industry, such as the unequal representation of female actors, directors and executives, or the widely discussed gender pay gap (Hollinger, 2012; Leavy, 2007).

More recently, feminists have critiqued scientific culture, particularly for its role in reifying the "ideology of difference" (Richards, Bouman & Barker, 2017). Harding (2003) argues that science and academia have contributed significantly to the social construction of gender-as-binary and the extension of sexual dimorphism explanations to nearly all aspects of social life. These include critiques of early biological, psychiatric, and evolutionary explanations of gender and sexual behaviour (e.g., associated with Darwin and Galton), as well as early and modern theories within the field of sexology (from Havelock Ellis' work in the 19th century up to modern neuroscience explanations of gender difference: see Fausto-Sterling, 1992 and Schmitz & Höppner, 2014 for critiques). Scientific theories have also come under scrutiny within feminist circles for naturalising male violence and sexual aggression as a normal mating behaviour (as opposed to a form of abuse: see McPhail, 2016 for a comprehensive review). Contemporary scientific practices have also been argued play a role in upholding the binary, and more generally androcentric systems of knowledge. As mentioned previously, these include (but are not limited to) the tendencies to analyse by gender without a rationale, attribute gender differences to women rather than men, explain gender differences in terms of biology rather than bio-psycho-social causes, and collect gender information in a binarized way; see Hegarty & Pratto, 2006; Fraser, 2018; Seal, 2017).

In addition to demonstrating how the roles of women and men are shaped and reproduced through social practice, cultural research has demonstrated how the

binary may be enforced through widespread social intolerance of gender ambiguity. These include ethnomethodological analyses conducted by non-binary individuals (e.g., Garfinkel, 1967; Lucal, 1999), as well as the large-scale surveys documenting the violence, discrimination, and prejudice these groups experience (e.g., Budge, Adelson, & Howard, 2013). While there has been a significant increase in efforts to challenge the binary in recent years (e.g., through raising awareness of different gender identity categories, or introducing gender-neutral pronouns such as they/ze/zhe, etc.: see Darr & Tyler, 2016), research suggests linguistic and societal change is quite slow. For example, a recent analysis of Facebook's practices found that, although users can choose from over 50 gender identification categories (as of 2017), non-binary users are re-classified as male or female in the deep code of the database (Bivens, 2017).

1.4.2 Analysing History

As argued by Butler (1988), gender should always be considered a historical situation rather than a natural fact, and indeed historical analyses have provided valuable insights into how gender systems may be established or maintained over time. Within the West, for example, analyses of historical records suggest that women's oppression as we currently understand it is a relatively new phenomenon (Bennett, 1989). Before the emergence of class-based social systems around the 14th Century, for example, some historians posit that gender divisions in labour, care work, and power were present but significantly less pronounced than they are now (Brenner, 1989; Thane, 1992). The separation of labour and childcare roles after the industrial revolution is considered another important precipitating factor to modern inequality, given that it increased the child-rearing burden on women and laid the foundations for their exclusion from the workplace (Sharpe, 2002). An additional

significant historical event is the institutionalization of marriage. While records show that marital-style practices have ancient roots in many cultures (Pederson, 1999), the transition from marriage as a ceremonial practice to a patriarchal legal system that regulated property ownership and reproduction had significant consequences for women's equality (see Feldstein, 2002 for a detailed analysis).

Historical analysis has also informed our understanding of how gender roles and power dynamics may be constructed over time. For instance, re-analyses of historical records and texts show how the masculine stereotype evolved over time to fit changing social roles and power structures. Kimmel (2005), for instance, notes how in many ancient texts and epics, masculinity was associated with sexual virility, strength and heroism, while Victorian ideals of manhood largely encompassed Christian purity and chivalry. Records show that at the beginning of the 20th century, masculinity underwent another significant transformation, coming to be defined largely by participation and success in the paid labour force, as well as the capacity to provide for a family (Connell, 2005). Masculine sexual norms have also changed considerably over time, particularly with regards to same-sex behaviour. For example, while heterosexuality is now considered a defining feature of masculinity in many cultures, there have been numerous instances of normative homosexual or homo-erotic behaviour across history (see Foucault, 1986).

The related fields of Women's History and Feminist History have been similarly important for understanding how women have been oppressed in both history and historical records (Smith, 2008). Scholars from these fields often analyse (or re-analyse) texts through a feminist lens, seeking to better understand the contextual factors moderating women's subordination or the "mechanisms of separation" (e.g., changes in the labour market, famine, urbanisation, etc.), the social

roles held by women across history, or cross-cultural comparisons of inequality through the ages (see Bennett, 1989). Feminist historiographers have also made visible androcentric practices within history, including the tendencies of male historians to focus on men's achievements or indeed attribute success to men over women (Sutherland, 2002; Frank, 2017). A notable example includes the paleontological and anthropological interpretations of cave drawings from the Palaeolithic age. While initially, these early examples of art were attributed to men, recent re-analyses of the hand size and digit length indicates these drawings were likely done by women (Gelder & Sharpe, 2009). Other re-analyses have explored the minimization of women's contributions in science or the arts (e.g., Lady Mary Wortley Montagu who brought inoculation to Britain in 1822, or Rosalind Franklin's contribution to the discovery of DNA: Grundy, 2000; Maddox, 2002). According to Pederson (2004), these assumptions and related tendency to overemphasise the contributions of men relative to women has significantly shaped modern conceptions of men as natural leaders (see also feminist critiques of the "Great Man Theory" in earlier historical research: e.g., Vetter, 2010).

1.4.3 Analysing Development

As mentioned, gender and feminist theorists place considerable importance on gender role socialization, and developmental research has been useful in elucidating how women and men "learn to gender" in a range of contexts. Observational, naturalistic and longitudinal studies have demonstrated, for example, that children are exposed to vastly different toys and activities depending on their gender, with parents more likely to give action-based or problem-solving toys to boys (e.g., toy trucks, puzzles, or action figures) and nurture or beauty-based toys to girls (e.g., dolls, toy kitchens, or beauty sets: see Steffens & Viladot, 2015 for a

recent review of this literature). The familial context has also been analysed, with studies demonstrating how norms may be passed from parent to child through direct gender-stereotyped feedback and information (e.g., hearing statements like "dresses are only for girls", or "boys don't play with dolls", and so on: Blaise, 2005; Leaper & Friedman, 2007), as well as through more covert observation and inference (e.g., modelling, being spoken to or parented differently depending on their gender, or seeing their parents engage in gender-stereotyped behaviour and activities: Bandura, 1977; Epstein & Ward, 2011; Leaper, 2014; Witt, 1997). Early education settings are another important context for socialization, with a wealth of empirical and longitudinal studies exploring the origins of gender-stereotyped subjects and interests (e.g., through differentiated feedback from teachers and peers, or the segregation of different school-based or extracurricular activities: see Bigler, Hayes & Hamilton, 2013 and Goodwin and Kyratzis, 1997 for comprehensive reviews). Gender differences have also been observed in the context of moral development (see Jaffee & Hyde, 2000 for a meta-analysis of this literature). According to a now-seminal analysis by Carol Gilligan (1982), the two gender groups are socialised in culture to have vastly different codes of ethics, with women encouraged to develop and "ethics of care" (focused on responsibilities and relationships), and men and "ethics of personal responsibility" (focused on individual rights and justice).

Research has also measured and modelled the *consequences* of gender socialization, for both the individual and society (Yoder, 1995). To look first at women, communal socialisation practices have been argued to explain women's enhanced emotional and interpersonal skills or increased life expectancy, but equally their socioeconomic disadvantage, low or subordinate social status and assumed responsibility for the care of others. For men, socialisation is believed to account for

their increased license to develop and display socially valued traits (e.g., wit, scientific ability, and leadership) but then their proclivity towards violence and crime, impaired emotional awareness, and increased risk of suicide or social isolation (Eagly & Karau, 2002; Kimmel, 2004; Pleck, 1981; Levant & Pollack, 1995; Moss-Racusin et al., 2010: see also the recent literature on "toxic masculinity", e.g., Haider, 2016). These outcomes have been measured at multiple levels of analysis using a variety of methods. At the societal level, as mentioned earlier, national and global reports are often used to measure societal gender divisions (e.g., around health, labour, caregiving, and crime). They have also been used to estimate the economic costs associated with gendered practices (e.g., of domestic violence or suicide: Shepard, Gurewich, Lwin, Reed & Silvrman, 2015; Walby, 2009). At the level of the individual, large bodies of qualitative and quantitative research from psychology, sociology, applied social studies and the health sciences demonstrates the effects of communal and agentic socialisation. While outside the scope of this review, these include the well-established gender differences in verbal, emotional, and spatial reasoning abilities, personality, aggression, help-seeking behaviours, and sexual behaviour (see Richardson, 2013 for a comprehensive review). Neuroscience and neuroendocrinology have more recently examined how gender socialisation may literally become embodied (e.g., by looking at the hormonal or neural changes which may occur in response to gender-typical learning: see Fine, 2010 and Van Anders et al., 2015 for summaries).

In addition to the empirical and cultural research, feminists have critiqued traditional developmental theory for its role in maintaining the binary. According to Miller (2015), traditional theories played a significant role in scientific and lay conceptualisations of gender, largely by popularising the notion of healthy or

"successful" gender development. Early psychometric measures of gender identity such as Terman's Masculinity-Femininity (M-F) scale (1936) and later androgyny measures (e.g., Bem's Sex Role Inventory, 1974) were heavily critiqued for assuming homogenous and natural gender categories and then measuring an individual's scores against them (see Fraser, 2018). Similarly, models proposed by Freud (1905) and Parsons (1955) assumed essential or natural gender identities, which boys and girls attain or "master" as they develop. Kimmel (2012) proposes that, over time, these models shaped cultural norms around gender-appropriate behaviour and influenced how psychologists, medics, educators, and parents interpret and respond to children's development. This extends to the gender-normative socialisation practices discussed above, but also the broader societal pathologization of "deviant" or non-binary behaviours (e.g., the diagnosis of gender dysphoria or criminalisation of same-sex behaviours, etc.: see Harper, 2007; Richards & Barker, 2013).

1.4.4 Analysing Language

Language plays a central role within feminist and gender theory, and thematic, content, conversation, and phenomenological analysis methods remain popular in modern gender studies (Cheshire & Trudgill, 1998; Holmes & Meyeroff, 2003; Lakoff, 1975; Tannen, 1994). Because discourse is often conceptualised as the medium through which the social world is created and reproduced (e.g., Foucault, 1978; Butler, 1990), gender theorists frequently turn to language and discourse to understand the social construction of gender. Within linguistics, for instance, a large body of work has explored gendered speech patterns (or "genderlects": see Tannen, 1994), and the ways in which they may reflect communal and agentic gender roles. Studies show that, relative to men, women's speech tends to be characterised by

more emotionality (i.e., reliant on emotional topics or inclusive of more emotional intensifiers), self-disclosure, active listening and minimal responses (e.g., "mm" or "yeah" when another person is speaking), the use of tag questions or mitigating words/sounds (e.g., "isn't it?" or "you know?" after a question or between topics), frequent questions, and indirect phrasing (Coates, 2015; Holmes, 1992; Lakoff, 1975; Menegatti & Rubini, 2017; Tannen, 1990, 1993; West & Zimmerman, 1975). By contrast, these studies found men's speech tends to be more direct, fact-based, focused on external rather than personal information, low in emotional content, and characterised by fewer questions, linking statements or mitigating words.

In addition to the differences in speech, feminist linguists have explored how heteronormative power dynamics may play out in conversation and social interaction. These include naturalistic analyses of gendered verbal practices like street remarks (Gardner 1980; Kissling 1991; Kramarae, 1992), sexist slang (Grossman & Tucker, 1997), sexual or harassing language in the workplace (Holmes, 2005; Ragan et al. 1996), coaxing or coercion prior to sexual activity (Muehlenhard et al., 1991), and online abuse (Herring 1999; Herring & Stoerger, 2017), as well as male acts of "conversational dominance" (e.g., talking over women, denigrating or making jokes about women's issues, or a low level of uptake of women's topics: Fishman 1983; Ochs and Taylor 1995; Spender 1985). Other laboratory studies have examined how gender expectations may bias listeners or evaluators. For example, studies show that although men are significantly more likely than women to interrupt in a dyadic or group context, women are often perceived as having interrupted more frequently and less appropriately than men (Anderson & Leaper, 1998; Robinson & Reis, 1989). Similar results have been found for volubility (or "talkativeness") whereby women are often perceived as

having spoken more than men in mixed groups, even when they are described as having contributed an identical amount (Brescoll, 2001, 2011; Cutler & Scott, 1990). The structure of a conversation has also been shown to be influenced by the gender of the participants, and these patterns again seem to reflect broader socialisation practices. For example, men are significantly less likely than women to engage in conversational turn-taking or simultaneous speech, particularly in mixed gender dyads or groups (Coates, 2015).

Other feminist analyses have focused on the formal properties of language, exploring how the binary may be embedded in a language's grammatical or syntactical rules (Menegatti & Rubini, 2017). While English does not grammatically mark gender in the same way as some other languages (i.e., ascribe a gender to nouns or their dependent linguistic forms, as is done in French, Italian or German: see Braun, Irmen, & Sczesny, 2007), linguists suggest the binary is still constructed or reproduced through various grammatical norms. For example, linguistic convention dictates that we both designate and qualify individuals according to their gender (i.e., he or she: see Wittig, 1985), and it remains rare in the English language to use the gender neutral "they" despite increased activism in this area (see Richards & Barker, 2015). Theorists have also identified other grammatical norms that may reify gender roles, including objectification language (e.g., women-object associations in language: Kissling, 1991) and the previously discussed androcentric linguistic practices (see Bailey & LaFrance, 2018). More broadly, linguists and cultural critics have critiqued the Western tendency to frame social concepts around "binary oppositions"; that is, pair social categories together in language or thought as related, oppositional concepts (Cameron, 1997). While not a grammatical rule as such, evidence suggests people readily dichotomize and polarise many social

categories (e.g., male-female, gay-straight, mind-body, good-evil, etc.: Bing & Victoria, 1996; Utaker, 1974; Westen, 2001).

Another popular analytic method within gender studies is discourse analysis (Zimmen & Hall, 2016). Discourse may be broadly defined as language in context, and its study encompasses the analysis of discursive practices (e.g., text or conversations) as well as the many contextual, historical, personal, or situational factors that may influence them (Fairclough, 1992; Wetherell & Potter, 1992). With regards to the construction of gender-as-binary, analyses have explored how sex/gender categories may be framed in language (i.e., as relational opposites: see Ehrlich, Meyeroff & Holmes, 2014), as well as the extent to which gender, sex, and sexual orientation may become fused in popular discourse (e.g., "a man needs the love of a woman": Livia & Hall, 1997). Discourse analyses have also been used to examine the construction of gender self-concept or identities. For example, studies have looked at the verbal construction of gender identities in early life (e.g., "I am a boy, and boys have short hair and like cars": e.g., Leaper & Friedman, 2007; Litosseliti & Sunderland, 2002), and also how prevailing cultural norms about gender may spill over into individual gender self-concepts (e.g., "Like most women, I don't have an aptitude for science"; Bacchi, 1999).

The role of the social-relational context in gender construction has also been elucidated using discourse analyses. Analyses of conversations between male peer groups, for instance, have shown how traditional markers of masculinity (e.g., heterosexuality, sexual dominance, violence, financial success, and a rejection of emotionality) may be frequently regulated through verbal practice (Benwell, 2017; Blaise, 2005; Gilmore, 1995; Kiesling, 1997; Woodward, 2000). Similarly, studies of female conversations show that they can similarly reflect and enforce patriarchal

ideas of femininity (e.g., through frequent discussions of weight or body image, romantic relationships, and child-rearing: Cameron, 1997; Heilburn, 1988; Wilton, 1992), though they may also provide a space for consolation, intimacy, and subversive discourse and resistance (Coates, 2015; Green, 1998). Others have examined the role of the specific social context on gendered discourse, or how gender norms and expectations may interact with other social or power structures (e.g., by comparing discourses of masculinity across racial, class, or ethnicity groups: Bucholtz, 1999; Eckert and McConnell-Ginet 1992; Lave and Wenger 1991).

1.4.5 Analysing Beliefs

The final analytic category of relevance to the current thesis is beliefs. Research into gender-related beliefs is vast and encompasses the study of attitudes, stereotypes, biases, norms, and identities (Waylen, Celis, Kantla & Weldon, 2013). Each of these constructs may be defined in numerous ways depending on the discipline, or even specific text or article, but generally speaking they may be defined as follows: attitudes typically refer to valenced evaluations or associations about a specific social group (e.g., women are bad); stereotypes are beliefs or assumptions about a group's typical, natural, or essential behaviour (e.g., women are nurturing; men should be strong); social bias describes the cognitive, perceptual or behavioural tendency to favour or be prejudiced against one group over another (e.g., hiring a man over a woman); and norms are the broad beliefs about what is normal, normative, or appropriate for a particular social group (or system) (Greenwald, Rudman, Nosek, Banaji, Farnham & Mellott, 2002; Paluck & Ball, 2010). Gender identity then refers to the broad set of associations or beliefs that make up a person's gender self-concept, as well as the specific labels a person may use to refer to

themselves within a particular culture (Richards & Barker, 2015). As the bodies of research for each construct are broad, the following review will focus on the areas most relevant to a discussion of the binary: essentialist and/or gender-as-binary beliefs; anti-women or pro-male bias; gender stereotypes and/or the endorsement of traditional gender roles; and gender identity/self-concept. Moreover, while beliefs may be assessed using any number of qualitative or quantitative methods, this review will limit itself to literature relying on explicit self-report measures (e.g., scales and questionnaires). This is because self-reports are by far the most widely used measurement tool for analysing beliefs (see Deaux & Snyder, 2012), and also because this review has already covered other analytic paradigms and techniques for measuring the binary (e.g., cultural criticism, discourse analysis, etc.).

To look first at essentialism, research has measured different aspects of essentialist beliefs in various ways. Some studies have focused on biological determinism, examining, for example, how strongly people endorse biological explanations of gender (Keller, 2005), encourage essentialist research or ideas (Morton et al., 2009), or believe that biology necessitates women and men's roles in society (Tinsley et al., 2015). Others have focused on the immutability of gender differences. Brescoll, Uhlmann and Newman (2013), for example, measured how strongly participants endorsed biological/natural explanations of gender difference and also the extent to which they believed these differences could be changed by environmental influence. A small number of questionnaires specifically intended to measure essentialism and/or binarism have been developed in recently years, including the Heteronormativity Attitudes and Beliefs Scale (HABS: Habarth, 2015). The HABS assesses both gender and sexuality-related beliefs but includes a specific subscale for measuring the binary structure and composition of gender (with items

such as “There are only two genders”). Skewes, Fine and Haslam (2018) recently developed the Gender Essentialism Scale (GES) as a way to measure the various different components of essentialist or binarist thinking. Specifically, this scale is intended to capture the beliefs that gender differences are discrete, biologically based, immutable, inherent, historically invariant, and highly informative.

With regards to anti-women sentiment, a range of measures have been developed over the past 50 years to assess sexism in its various forms. These include classic misogyny questionnaires, such as the Attitudes towards Women Scale (AWS: Spence & Heilmrich, 1972). The AWS was initially developed as a way to measure hateful or negative views about women relative to men, including items such as “Swearing and obscenity are more repulsive in the speech of a woman than of a man.” Since the 1970s, measures have evolved to assess more contemporary sexist beliefs as well as the more explicit prejudice. The Modern Sexism Scale (MS: Swim et al., 1995), for example, includes separate subscales for “old-fashioned” sexism (with items similar to the AWS) and modern sexism, characterised more by a rejection of women’s rights movements (e.g., “Discrimination against women is no longer a problem in the United States”). The Ambivalent Sexism Inventory (ASI: Glick & Fiske, 2000) similarly attempts to capture different types of sexist attitudes, measuring both *hostile* (e.g., “Women are too easily offended”) and *benevolent* (but still harmful) beliefs about women (e.g., “Women should be cherished and protected by men”). More recently, researchers have sought to measure more nuanced or subtle forms of anti-women prejudice, such as an individual’s level of endorsement with various structural inequalities. For example, studies have recently examined individual differences in “choice” explanations of workplace inequality that is, the extent to which they believe inequality results from women’s own life choices

(Skewes, Fine & Haslam, 2018). Other studies have focused on how women are evaluated or punished when they deviate the prescribed communal gender role (i.e., opt not to have children, behave in an agentic manner, or occupy a traditionally masculine social role). As discussed previously, these "backlash effects" are well-documented, particularly in the context of women in leadership (Eagly & Karau, 2002; Rudman, 2012).

Stereotypes are another widely measured construct within gender studies and social psychology. As reviewed at length above, numerous studies have shown that women and men are associated with vastly different attributes, abilities, traits, skills and interests, with women more readily associated with subordinate or care roles and men with dominant or leadership roles (e.g., Keonig et al., 2010). These stereotypes are both pervasive and broad, with stereotyping found across virtually all demographic groups (i.e., age, gender, ethnicity, etc.) and most societal spheres and domains (e.g., play, education, work, politics, and so on: see Waylen et al., 2013). In addition to understanding the content of these gendered associations, research has examined the dimensions or structure of the stereotypes themselves. For instance, theorists have assessed the extent to which female and male stereotypes respectively map onto established stereotype dimensions of communion and agency (discussed previously: Eagly & Karau, 2002), warmth and competency (Cuddy, Fiske & Glick, 2008), or expressivity and instrumentality (Kachel et al., 2016). Other research has focused on the functions of gender stereotypes, and the ways in which they set out rules for appropriate gender behaviour (Ridgeway, 2011). Evidence suggests, for example, that gender stereotypes are not merely descriptions of men and women's typical behaviour; rather, they are comprehensive, prescriptive, and proscriptive

norms about appropriate male and female behaviour (e.g., men shouldn't cry, women should be nice, etc.: Keonig, 2018; Prentice & Carranaza, 2002).

As with gender beliefs, individual variation in gender identity or self-concept has historically been measured using self-report techniques. These include the early masculinity-femininity and androgyny scales discussed previously (e.g., Terman, 1936: Bem, 1974) as well as more recent assessments of self-gender trait associations (Kachel et al., 2016). Gender identity may also be assessed in clinical settings using both observational and clinical assessment measures. These assessments would normally take place within the context of a gender dysphoria diagnosis (defined in the DSM-V as the distress associated when a person's gender identity does not align with their assigned sex at birth), and thus tend to focus as much on the person's well-being as their gender concordance (see Schneider et al., 2016). Popular assessments include the Utrecht Gender Dysphoria Scale (Cohen-Kettenis & van Goozen, 1997) and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (Deogracias et al., 2007), as well as sub-clinical assessments of distress like the Masculine Gender Role Stress scale (MGRS: Eisler, 1987). While many of these measures assess gender identity along masculine-feminine lines (e.g., as a position on a unipolar masculine-feminine dimension, or as scores on two separate intersecting dimensions: see Kachel et al., 2016), theorists have begun to acknowledge the complexity of identification process, as well as the degree to which gender intersects with other forms of identity. The recently developed Sexual Configurations Theory (SCT: Van Anders, 2015), for example, proposes gender identity intersects with various other dimensions within a person's sex/gender self-concept (including sexual attractions, desires, preferences within partnered and solo sexuality, and so on). According to Van Anders, the failure of

traditional theory to acknowledge the diversity and complexity of these intersecting dimensions (or indeed their contextual variation) has led to a reductionist, heteronormative gender identity framework (i.e., that can only meaningfully explain traditional gender/sex dyadic pairings).

While gender attitude, stereotype, and identity processes are viewed as distinct psychological constructs and have their own bodies of literature and evidence (Richards & Barker, 2015), it is important at this stage to review how they relate to one another, and indeed other forms of social prejudice and discrimination. A key assumption made within the feminist literature is that a binarist, patriarchal ideology has consequences for gender equality (e.g., Ridgeway, 2011), and indeed a review of the research suggests many of the above beliefs inter-correlate or overlap. Essentialism, for instance, has been shown to predict negative evaluations or beliefs about women (Keller, 2005; Skewes et al., 2018), endorsement of traditional breadwinner/provider gender roles (Gaunt, 2006; Tinsley et al., 2016) and broader gender stereotyping (Meyer & Gelman, 2016). Similarly, sexism is associated with a range of other discriminatory attitudes (e.g., racism, ageism, ableism: see Baldwin, 2017) and prejudicial gender behaviours, such as hiring discrimination (Fiske & lee, 2008), androcentric bias (Bailey & LaFrance, 2018), and sexual harassment proclivity (Rudman & Mescher, 2012). A number of studies have also examined the relationship between gender beliefs and the broader endorsement of conservative worldviews. Skewes et al. (2018), for example, found a relationship between essentialism and an endorsement of structural gender inequalities, while Christopher and Mull (2006) showed that ambivalent sexism predicted participants' alignment with a conservative ideology. Traditional gender views have similarly been found to correlate with right wing authoritarianism (Sibley, Wilson & Duckitt, 2007), as well

as system-justification or status-quo explanations of gender inequality (Keller, 2005; Morton et al., 2016).

In addition to research showing how these beliefs may coalesce to form and reflect a patriarchal value system, it would also be useful at this stage to review how a person's own gender may influence their gender world view. Generally speaking, research suggests that men have more conservative, sexist, and androcentric gender beliefs than women. Men are also more likely to use sexist language, discriminate based on gender, disregard or downplay feminist efforts, and more generally endorse a politically conservative worldview (see Cameron, 1998 and Weatherall, 2005 for reviews). While these effects are generally robust, it should be noted that the reliability and magnitude of gender differences varies considerably across studies and contexts. For example, essentialist beliefs do not always differ across gender groups (e.g., Skewes et al., 2018), and both women and men have been shown to endorse gender-normative expectations (e.g., by negatively evaluating those who deviate from prescribed gender norms: Acker, 2009; Campbell, 2004; McCreary, 1994; Smiler, 2004). Similarly, though the strength of stereotypes can vary across genders, studies show the content tends to be broadly similar (Deaux et al., 1985; Fiske, 2010; Koenig, 2018). According to Glick and Fiske (2000), this is largely because women and men have access to the same pervasive, dominant ideological information. Moreover, studies suggest that women are often rewarded for expressing traditional gender views, rejecting feminist/egalitarian perspectives, or engaging in sexist "banter" and behaviour (e.g., Bearman, Korobov & Thorne, 2009; Ficher, 2006). Gaunt (2013) therefore argues against treating gender as a simple moderator of discriminatory attitudes and behaviour. Rather, it should be conceptualised as a mediating or participating variable, which influences a person's

gendered experiences and exposures over the lifespan, and thus their core beliefs and assumptions about gender.

1.5 Measuring the Gender Binary: A Novel Approach

Taken together, this large body of research highlights the binary construction of gender and the relationship of binarization to gender inequality. It also demonstrates the many different paradigms, data sources, and theoretical perspectives that can be adopted when studying these practices at varying levels of analysis (i.e., cultural, contextual, or individual). To summarise, at the cultural level, analyses have been useful in demonstrating the pervasiveness of binarist practices and beliefs across different social spheres. They have also highlighted how binarism may underpin many systems of knowledge, influence, and power. Developmental research has shown how these systems may spill over and shape a person's gender knowledge, both for themselves and others, and how the cultural context may reinforce the binary in subtle and direct ways. At the individual level of analysis, a large body of work has explored how the binary may be reproduced through, or reflected in, various linguistic, behavioural and cognitive processes.

In outlining this variety of existing approaches, the previous section highlighted that there are, however, few direct measures of gender binarism that can be used in experimental or individual-level analysis. The current section therefore introduces a broad class of measures that may serve to fill this gap. Broadly referred to as *implicit measures*, this diverse set of paradigms was developed as a way to study various cognitive processes or biases that may be difficult to capture using conventional methods (Gawronski & De Houwer, 2014). These include those which a person might wish to conceal, alter, or suppress (e.g., due to social desirability concerns), in addition to those which could be outside of a person's conscious

awareness or intentional control (Moors, 2011). Importantly, implicit measures are considered a useful way to analyse *automatic* social-cognitive processes, defined as those which are especially well-entrenched, basic, immediate, and efficient (Gawronski, 2015). Given that the binary is considered such a foundational and primary frame for organising gender information (e.g., Bem, 1996; Ridgeway, 2011), it is likely that binary biases would be highly automatic (or at least have an automatic component). However, to date, no research has been conducted that has either (a) conceptualised the binary in terms of automatic or implicit cognition, or (b) investigated whether binary biases are readily demonstrated on an implicit measure. The current thesis aims to address this gap in the literature by conducting the first in-depth investigation into the automaticity of the gender binary. Using an implicit measure, this research will explore several important features of binary biases, and specifically the tendency to frame women and men as opposites with distinct and mutually exclusive traits and abilities.

1.5.1 Implicit Measures

Before outlining how implicit measures will be used in the current thesis, it is necessary to first give a brief procedural overview of these paradigms and their evidence base. Implicit measures were first introduced in the mid-1980s, when researchers extended and modified sequential priming tasks for the study of automatic social cognitions and behaviours (Fazio & Olsen, 2003; Gawronski & De Houwer, 2014). One of the earliest measures developed for this purpose was the Implicit Association Task (IAT: Greenwald, McGhee & Schwartz, 1998), which, in brief, provides an index of a person's automatic attitudes or biases by comparing their performance on two opposing tasks. In one task, participants are required to pair or relate sets of stimuli assumed to be consistent with a particular cultural

stereotype (e.g., women as submissive and men as dominant), and in the other, sets of stimuli that are deemed to be inconsistent (e.g., men as submissive and women as dominant). Participants usually relate these stimuli together by pressing one of two response keys on a keyboard (e.g., the letter *z* for "men" and "good", or *m* for "women" and "bad"). Any differences in performance across the two tasks (e.g., in terms of speed or accuracy) are taken as a metric of the automaticity, strength, or coherence of one task relative to the other, and thus as a broad metric of a cognitive bias (De Houwer, 2014). Other widely used paradigms include the Affective Misattribution Procedure (Payne, 2014), which infers automatic bias from participants' misattributions about the sources of their affect or cognitions, and the Go/No-go Association Task (Nosek & Banaji 2001), in which participants are required to either respond or not respond to different stimulus-valence pairings (e.g., black-bad versus black-good). Inferring psychological content from behavioural performances in this way (rather than self-reports) is what leads to many researchers referring to these measures as implicit or automatic (although see Gawronski & De Houwer, 2014 for a comprehensive review of the various terms in implicit measures research).

Effects on implicit measures (i.e., the difference in performance across the two types of task) are typically produced using response latencies, or the length of time before a participant presses a key to categorise or relate the stimuli on-screen (Fazio & Olsen, 2003). Specific algorithms or scoring procedures vary within the literature, but generally speaking it is conventional to calculate the mean response latencies for each task (and occasionally means across multiple blocks for each task type) and produce a single score based on the effect size of that difference. The conventional IAT score, for example, for example, is the difference in mean reaction

times between two block types divided by the standard error of the pooled reaction times. This effect size is related to (but distinct from) Cohen's *d*, and therefore referred to as the *D* score (Greenwald, Banaji, & Nosek, 2003). Other scoring algorithms have relied on accuracy differences (e.g., Payne, 2014) or some combination of speed and accuracy (also called fluency: see Gavin, Roche & Ruiz, 2008). Research using implicit measures usually includes a period of practice or training before the test phase, intended to familiarise participants with the task and/or get them to a certain level of speed or accuracy (Nosek, Greenwald & Banaji, 2005). Generally, scores would only be based on latencies for the testing phase.

It is worth noting at this stage that there has been considerable debate and variation in the literature regarding the interpretation of the effects on these tasks; that is, which psychological mechanism, process, or construct the effect is believed to represent (De Houwer & Moors, 2007). To give some background, implicit-style paradigms have been developed and used by a number of psychological sub-disciplines, and this theoretical eclecticism means the effect is often interpreted in very different ways. Within social cognition, for example, they are generally conceptualized as measures of unconscious, automatic, indirect, or implicit associations (Nosek, Greenwald & Banaji, 2005). Contemporary learning psychologists have used these paradigms to measure various associative, propositional, analogical or evaluative processes (De Houwer, Teige-Mocigemba, Spruyt & Moors, 2009), while in behaviour analysis they may be used to examine individual histories of relating different classes of stimuli (Roche et al., 2008). Many, however, have simply suggested that paradigms are a useful way to measure socially-relevant behaviours under conditions of automaticity (Greenwald, Poehlman, Uhlmann & Banaji, 2009; Hughes, Barnes-Holmes & De Houwer, 2011).

The current thesis adopts a similar theoretical stance to these researchers and is broadly agnostic to the psychological processes involved. To expand, given that the aim of this work is to examine the automaticity of the binary and explore whether implicit measures can inform our understanding of how gender is framed or constructed, it is neither necessary nor relevant to connect effects to a specific cognitive process.

To return to the potential relevance and utility of implicit measures, there is now a very large body of research that has examined their utility in a range of different contexts. Meta and systematic analyses of this literature have been conducted on the several million individuals who have participated in studies using implicit measures over the last two decades (Xu et al., 2014), in a variety of different socially relevant domains (e.g., race, religion, political identification, and sexuality: see Kurdi et al. 2018 for a recent meta-analysis). This includes the context of gender, where they have been used extensively to examine stereotypes around agency and communion (e.g., Rudman & Glick, 2001; Rudman & Kilianski, 2000), scientific ability (e.g., Smyth & Nosek, 2015), competency (e.g., Latu, Stewart, Myers, Lisco, Estes & Donohue, 2011) or objectification (e.g., Rudman & Mescher, 2012). They have also seen extensive use outside of social psychology, such as within clinical research, political science, or legal studies (see Roefs, Huijding, Smulders, et al, 2011 a review and meta-analysis). A range of studies and meta-analyses have also analysed the psychometric properties of these paradigms (e.g., Golijani-Moghaddam, Hart, & Dawson, 2013; Nosek et al 2007; Vahey et al 2015). While outside the scope of this review, the above-cited review generally suggest they have adequate validity, internal consistency, and predictive utility, although often lower than traditional self-reports measures.

1.5.1.1 The Implicit Relational Assessment Procedure

Given the number of implicit measures that have been developed in recent years, it was important to ensure one was selected which best suited the needs of the current thesis. Following a review of the literature, the Implicit Relational Assessment Procedure (IRAP: Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2010) was considered the most appropriate for an assessment of binary biases. To give a brief overview of this paradigm, the IRAP shares many procedural properties with other mainstream implicit measures. As in the IAT and other measures, a typical IRAP examines a participant's ability to relate two different stimulus categories (e.g., men and women) with two other stimulus categories (e.g., stereotypically masculine or feminine traits). It has a two-block structure, with participants required to relate the categories in different ways in different blocks, usually under one of two response "rules". To continue with the example of the above categories, participants in an IRAP would relate stimuli in one block according to the rule "Respond as if men have stereotypically masculine traits and women have stereotypically feminine traits", and the other to the rule "Respond as if women have stereotypically masculine traits and men have stereotypically feminine traits". Blocks are made up of multiple trials, in which participants relate a specific pair of stimuli from each category. Stimuli may be presented on-screen as a pair (e.g., "Men" and "Dominant") or in the form of a statement (e.g., "Men are dominant"). Participants then respond using a set of relational terms (corresponding to different response keys) such as "Similar/Different" or "True/False".

It is this trial structure and format which makes the IRAP particularly well-suited to an assessment of the binary. Most implicit measures require participants to categorise or pair stimuli together using a common key press. In the IAT, for

example, the evaluation or attribute categories are at the top right and left of the screen and participants categorise a particular stimulus (e.g., "Men") as one or the other (e.g., masculine or feminine). As such, the measure can only produce a relativistic, either/or assessment of bias (e.g., men are more masculine than feminine / women are more feminine than masculine: Greenwald et al., 1998; Nosek & Banaji 2001). In the IRAP, however, the target and evaluation/attribute stimuli are presented on-screen at the same time and thus can be scored to produce four different effects for each trial type (i.e., men-masculine, men-feminine, women-masculine, women-feminine). As the IRAP is scored in a similar way to other measures (i.e., the effect size of the response latency differential across task types), the individual trial-type scores would be as follows: men are/are not masculine, men are/are not feminine, women are/are not masculine, and women are/are not feminine.

The IRAP thus allows for an assessment of several theoretically interesting features of gender binarism. First, it provides an index of how both sex and gender categories are related to one another; that is, whether they constitute distinct or oppositional categories. This would be demonstrated by the magnitude and significance of the IRAP effect, as there would be no notable difference across block/task types if the categories were not meaningful opposites (see Rothermund & Wentura, 2004 for a review of the evidence of IATs using non-distinct social categories). Second, it provides a unique way to examine the structure and make up of these sex/gender relations. Because the IRAP can be scored to produce individual trial-type effects, it is able to index the strength of both role-congruity (i.e., men-masculine and women-feminine) *and* role-incongruity effects (i.e., men-feminine and women-masculine). This means it can assess whether the strength and magnitude of biases are comparable across gender categories (i.e., if the significance of the

men-masculine effects are symmetrical to the women-feminine effects), and also if there is a comparable amount of resistance to forming role-incongruent relations (i.e., if the significance of the men-*not*-feminine effects are symmetrical with the women-*not*-masculine effects). Put simply, the IRAP allows for a novel, quantitative and individual-level metric of an important feature of the gender binary: that men are masculine *and not feminine*, and women are feminine *and not masculine.*

In addition to its procedural advantages, the IRAP was considered an appropriate measure given its substantial evidence base. To date, the IRAP has been used to examine biases towards racial and ethnic minorities (Barnes-Holmes et al., 2010; Barnes-Holmes, Murphy, Barnes-Holmes & Stewart, 2010; Drake et al., 2010), people who are overweight (Nolan, Murphy & Barnes-Holmes, 2013; Jurascio et al., 2010), and gay or bisexual individuals (Cullen & Barnes-Holmes, 2008). It has also been used to examine sexual attraction (Timmins, Barnes-Holmes & Cullen, 2016) and has employed extensively in clinical settings to explore self-concept (Timko et al., 2010; Vahey et al., 2009), perspective-taking abilities (Barbero-Rubio; López-López; Luciano Eisenbeck, 2016; Kavanagh, Barnes-Holmes, Barnes-Holmes, McEnteggart & Finn, 2018), and behavioural avoidance and disgust (Nicholson et al., 2013, 2014). While it has not been applied as frequently in the context of gender, a small number of studies have used the IRAP to assess sexist or anti-women bias (e.g., Farrell & McHugh, 2017; Scanlon, McEnteggart, Barnes-Holmes & Barnes-Holmes, 2014). To date, however, no IRAP has examined core gender binary biases (i.e., the extent to which participants can fluently relate or juxtapose feminine and masculine attributes to both women and men).

1.6 Current Thesis

The current thesis an investigation of the automaticity of the gender binary. It has three broad aims: first, to use the IRAP to carry out a technical, quantitative analysis of binary biases, and specifically biases towards associating men (but not women) with stereotypically masculine attributes, and women (but not men) with stereotypically feminine attributes. In so doing, it is hoped this analysis will build on and inform feminist theorising in this area, and also add to the growing empirical literature on binarist beliefs. A second aim of this research is to assess the relationship between binary biases and other forms of gender-based discrimination and prejudice. As reviewed at length previously, a key assumption made in the feminist and gender literature is that conceptualising gender in terms of binaries is problematic. Therefore, throughout this thesis, IRAP scores will be correlated and/or compared with other assessments of gender bias. A final aim is to gather a large enough IRAP data set to be able to conduct an in-depth analysis of the effects produced. IRAP studies have, on average, a sample size of 30-50 (Vahey et al., 2013). While effect sizes are sufficiently large enough to have confidence in the effects produced and to use them for bivariate or correlational analyses (again see Vahey e al., 2013), the IRAP literature would benefit from a larger data set. For the purposes of the current thesis, a large dataset would be particularly beneficial for analysing the strength and significance of trial type effects (and a comparison across role-congruent and incongruent biases), as well as for conducting a deeper analysis of gender differences in these biases (if any).

1.6.1 Overview of Studies

The first empirical chapter in this thesis will primarily focus on stimulus selection for the core gender binary IRAP. Once stimuli have been selected and the

gender binary IRAP compiled, the remaining chapters will focus on its application and replication. Specifically, these chapters will assess how readily participants associate women and men with stereotypically feminine and masculine attributes, in addition to whether these biases are associated with some of the harmful behaviours reviewed previously: sexism, gender-based hiring discrimination, androcentric bias, and a proclivity towards gender-based violence. In some of these studies, participants will complete another IRAPs assessing gender biases around competency, humanity, and scientific ability. One study will additionally include a measure of gender identity. All studies will be conducted on samples of young Irish adults, who will predominantly be recruited from the Maynooth University student population. The final empirical chapter in this thesis will involve a pooled analysis of the binary IRAP and its effects, and a more detailed analysis of gender differences. The final chapter will be a general discussion and review of the findings and broader implications.

Chapter 2

Stimulus Selection and Preliminary Analyses

2.1 Introduction

The primary aim of the current chapter is to identify appropriate stimuli for use in the gender binary IRAP. Representative stimuli for the categories "masculinity" and "femininity" will be selected from a set of 60 personality traits, which participants will rate in terms of their gender and desirability. Preliminary categories will be selected in a pilot study and then validated using a larger sample. A secondary aim of this chapter is to develop an occupational gender preference task for use in Chapter Three. This task will be created using the same gender stimulus categories identified in the pilot study and will be tested using the larger sample. The final aim of this chapter is to generate a dataset of explicit trait ratings for later comparison with the pooled binary IRAP data. As the IRAP is a relatively novel way to examine how gender categories are related to one another, it will be useful to have a dataset of explicit (i.e., self-reported) ratings from a comparable sample of young Irish adults.

Stimulus Selection

Stimulus selection is an important stage in the development or adaptation of any psychological measure or task. Selection procedures and criteria vary across paradigms but generally this process involves a stage-like process of stimulus selection and validation. In questionnaire development, for example, a range of techniques may be used to ensure the stimuli or items are appropriately representative of the construct/s under investigation. According to Gilham (2000), best practice would typically involve an initial stage of stimulus/item generation based either on focus groups or existing literature or theory, followed by some psychometric (e.g., factor or measurement invariance analysis) and/or participant-led validation (e.g., further focus groups or interviews). Given the number of studies that

have been conducted using implicit measures in recent years (see Gawronski & De Houwer, 2014) and indeed the acknowledged impact of individual exemplars on the produced effects (see Gast & Rothermund, 2010 and Wosiefer, Westfall & Judd, 2017 for reviews), it is somewhat surprising that there are currently no universal or gold-standard protocols for stimulus selection. Instead, stimuli tend to be selected using a variety of methods, such as theory-led selection or selection from databases of normed images or words (De Houwer, 2001; Mitchell, Nosek & Banaji, 2003). The vast majority, however, do not report on the procedures or criteria used and instead select stimuli based on their familiarity and unambiguous "classifiability" with the overarching category (see Steffens et al., 2008). Before selecting stimuli for the current thesis, it would thus be useful to provide a review of the current conventions around stimulus selection, as well as some of the research into confounding factors.

To look first at the research around the number or type of exemplars that should be selected for stimulus categories, there are currently no established criteria within the broader literature. For the number, a methodological review conducted by Nosek (2007) recommends using as many as would be needed to represent the overarching category without compromising on relevance. As a rule of thumb, Nosek suggests using more than one (in case participants inadvertently respond to the formal properties of the individual exemplar and not the overarching category), but not so many that each stimulus would only be presented a very small number of times. Previous studies have varied in the number used, however the average number would be between 3-6 (see Xu, Nosek & Greenwald, 2014). With regards to the type, pictorial and word stimuli have both been widely used in previous studies (Bluemke & Friese, 2004). Again, there are no established criteria around when to use pictures

or words, but Foroni and Bel-Bahar (2010) recommend using whichever most clearly evokes the broader category. Previous implicit measures of gender-related issues have tended to use words, particularly when relying on stereotypical gender attributes or competencies (see Carnes et al., 2015; Rudman & Mescher, 2012; Rudman & Kilianski, 2000; Reuben, Sapienza & Zingales, 2014 for some reviews of this literature). Images and symbols have been used in this context before (e.g., images of male and female faces, Venus or Mars symbols, or images of different professions: e.g., Brochu & Morrison, 2010; Parker, Larker & Cockburn, 2018), but generally speaking words appear to be more routinely used in this context.

When selecting stimuli, it is also important to be conscious of any confounds. Stimulus clarity, familiarity and valence have all been shown to influence effects and outcomes on implicit measures (Govan & Williams, 2004; Steffens & Plewe, 2001), and should thus be factored into the selection process for the gender stimuli. As gender is now acknowledged to be a multi-faceted construct encompassing various different roles, attributes, skills and interests, for clarity purposes it would be important to restrict the focus to a specific aspect of gender. Many different aspects of gender have been examined using implicit measures before, including career stereotypes (e.g., science versus art, etc.: Nosek et al., 2009), competence-warmth stereotypes (e.g., Rudman & Kilianski, 2000) and objectification biases (e.g., Rudman et al. 2001). However, for assessments of more general masculinity/femininity gender roles – as in the current thesis – it has been most common to use familiar personality traits (i.e., dominant or nurturing: examples of communality/agency IATs).

With regards to valence, a number of studies suggest effects can change depending on the social desirability of the exemplars employed (e.g. De Houwer,

2003; Eder & Rothermund, 2008; Stahl et al., 2009). These effects seem to be particularly pronounced for evaluative or attitudinal implicit measures, which require participants to associate the target categories with positive and negative attributes (see Steffans et al., 2006). However, valence may also influence effects in indirect ways. Rudman, Greenwald and McGhee (2001), for example, found that female participants were less likely to associate women with stereotypical traits if those traits were negatively valenced (e.g., "weak") than when they were positively valenced (e.g., "delicate"). Similar results have been found in the context of race (e.g., Amodio & Devine, 2006), suggesting that desirability may confound evaluations by activating own-group biases or defences. In a more general sense, it is theoretically important to control for valence in the context of gender given that a key feature of the binary is the differential distribution of power across gender categories (e.g., Ridgeway, 2011).

The above review highlights several factors to consider when selecting stimuli for use in the gender binary IRAP. First, despite the absence of any standardized selection procedures in the existing literature, researchers should be transparent about the processes and criteria employed. Doing so should strengthen the validity of the measure and possible conclusions regarding its effects, and also allow the selection procedures to be critiqued or replicated by other researchers. Second, it is necessary to ensure the stimulus categories are clear and that they represent a coherent, well-defined construct. Because gender is multifaceted and broad, this means restricting the focus to one distinct component (e.g., traits). Finally, to ensure the effects are as clean as possible, it is important to properly control for the influence of potentially confounding variables such as stimulus valence and familiarity.

Hiring Task Development

The second aim of this Chapter is to develop and pilot a measure of occupational gender preference. As mentioned in the introduction, one of the core goals of this thesis is to examine the implicit societal relationship between men, masculinity and power. Chapter Three will explore this in detail by using the IRAP to measure sex-gender role relations and then assess whether participants express a preference for the masculine traits on a separate hiring simulation task. This hiring task will need to have the same masculinity/femininity categories as the IRAP and so will also need to be developed and piloted in advance. The rationale for focusing on occupational preference (as opposed to any other context) will be expanded on in the next chapter, but in brief it is because gender-power relations are known to be context-specific, and the workplace is argued to be one of the primary contexts where these relations play out in contemporary society (Eagly & Karau, 2002; Ferguson, 2003; Koenig, 2018).

The task developed here will be based on existing designs used to explore the impact of gender on hiring choices. Typically, these designs simulate hiring practices by presenting an application, personal description, or CV to a panel of participants or HR managers. The sex of the applicant would usually be manipulated (with some receiving an application from a male applicant and others from a female) and the hirability/suitability ratings of hirability for the male or female candidate compared across groups. Overwhelmingly, these men are more likely to be hired, promoted, and progressed for interview relative to women (Biernat & Fuegen, 2001; Brescoll & Uhlmann, 2008; Cann et al., 1981; Fuegen et al., 2004; Glick et al., 1988; Heilman & Okimoto, 2008; Kawakami et al., 2007; Rudman, 1998; Rudman & Glick, 2001; Smith et al., 2005). A meta-analysis of this literature by Isaac, Lee and Cames (2009)

showed these hiring biases are influenced by a range of factors (e.g., the applicant's physical attractiveness or marital status, the number of women in the application pool, awareness of equality initiatives, etc.), but one of the most significant moderators of bias is the type of position, with men more likely to be hired for stereotypically male roles (e.g., physician, CEO, or engineer) and women for stereotypically female roles (e.g., nurse, teacher, or assistant). As with other measures of gender-related beliefs or practices, preferences also tend to be affected by the gender of the participant or rater, with men more likely to hire men and women to hire women (see Isaac et al., 2009).

While most of the above studies explored preferences for men relative to women, comparably few have directly examined preferences for gendered traits (i.e., for masculine or feminine traits). The small number that have tended to examine the influence of gender attributes on perceptions of a candidate's hirability or suitability, and specifically the influence of gender-typical or atypical behaviour (e.g., Glick, Zion & Nelson, 1988; Rudman & Glick, 2001; Heilman, 1984; Futoran & Wyers, 1986; Twenge & Campbell, 2008). Together, these studies suggest that gender traits can influence hiring preferences, although not always symmetrically across gender or occupational categories. Generally, stereotypically masculine men are evaluated more positively than stereotypically feminine women, across a range of roles and levels (see Isaac et al., 2009). However, while gender-incongruent traits have been shown to increase men's suitability for female-typical jobs (e.g., a teacher or a nurse: Moss-Racusin, Phelan & Rudman, 2010), the same does not appear to be true for women. Several experimental and real-world studies demonstrate that while agentic and/or masculine women tend to be evaluated as more competent than feminine women but less socially capable, and thus less hireable or promotable (e.g., Phelan,

Moss-Racusin & Rudman, 2008; Rosette & Tost, 2010; Rudman & Glick, 2001). Again, these effects will be discussed in more detail in the next chapter, but given that these tasks seem to elicit a complex array of context and role-specific evaluations, the current task will not assess how gender traits increase or decrease men and women's likelihood of being hired for specific jobs. Rather, it will directly assess preferences for socially desirable masculine or feminine traits for a relatively gender-neutral occupation and then separately assess preference for sex (men or women). It will also include an exploratory item around who participants would prefer to let go from an office job between undesirable masculine and feminine candidates. No similar measures could be identified from existing research, but it is possible "firing" preference could be another useful indicator of gender bias.

Explicit Gender Trait Ratings

The final aim of this chapter is to preliminarily analyse how young Irish adults "gender" personality traits on a self-report measure. That is, the extent to which they explicitly evaluate traits as masculine or feminine, or as differentially desirable in women and men. Doing so would allow for a comparison of explicit and implicit effects towards the end of this thesis, and thus provide some insight into the utility of the IRAP methodology relative to more traditional methods. A number of gender norm or role measures exist in the literature already, including the Bem Sex Role Inventory (Bem, 1974), the Sex Role Questionnaire (Broverman, Vogel, Broverman, Clarkson & Rosenkrantz, 1972) and the more recent Traditional Masculinity-Femininity Scale (Kachel, Steffens & Niedlich, 2016). However, with the exception of the Traditional Masculinity-Femininity Scale, the majority of gender role or norm studies were conducted in the 1970s, and none could be found which used Irish samples. Recent research into gender stereotypes more broadly

would suggest that they remain prevalent in Ireland (see Fine-Davis, 2013, for a comprehensive longitudinal review), but it would be useful to examine trait ratings specifically among a similar demographic as the remainder of the studies in the current thesis (i.e., young Irish adults or University students aged between 18 and 25).

Overview of Studies

In sum, the aim of this chapter is to select stimuli for the IRAP and hiring task using a transparent, stage-like process with agreed-upon evaluation criteria. The first stage of this process involves the selection of stimuli from existing gender role measures and theory. In order to ensure the categories are clear, familiar, and representative of a coherent construct, the focus will be restricted to well-known personality traits. Once the list of traits has been selected, a small sample of participants will rate them in terms of their gender (i.e., how masculine or feminine they are) and desirability in both men and women. Because it is important to control for stimulus valence (especially in the context of gender), stimuli will be selected for four distinct categories: "Desirable Feminine"; "Desirable Masculine"; "Undesirable Feminine"; and "Undesirable Masculine". These will be selected by a small team of researchers based on combined gender and desirability ratings. The number of stimuli selected will depend on rating patterns, but it is expected that between 3-6 will be chosen for each category. These stimuli will then be used to develop the occupational preference task, with the desirable categories used for a "hiring" preference item and the undesirable for a "firing" item. The final stage of this process involves administering the 60-item scale and hiring task to a larger sample of young Irish adults. The stimulus categories and hiring task will then be re-reviewed

by the same team of researchers and amended if needed (e.g., if ratings differ notably across the pilot and larger samples).

2.2 Study One

2.2.1 Method

Participants

Forty students recruited from the Maynooth University undergraduate population participated in the pilot study. A sample of convenience was used and participants were recruited through word-of-mouth. Inclusion criteria included being aged between 18-29, living in Ireland at the time of the study, and having fluent English. Nineteen self-identified as female and 21 as male and all participants were aged between 18 and 25 (M = 20.98). Participants were provided with an open-ended response format for gender but all participants identified as either female or male. Race and ethnicity information was not collected as part of this study; however, all participants were currently living in Ireland and enrolled on an undergraduate degree. Participation was voluntary and no remuneration was offered.

Materials

Trait Rating Scale

Traits for the rating scales were selected from a series of existing androgyny and sex stereotype questionnaires: the Bem Sex Role Inventory (BSRI: Bem, 1974; 1985), the Sex Role Questionnaire (SRQ: Broverman, Vogel, Broverman, Clarkson & Rosenkrantz, 1972), and the Personal Attributes Questionnaire (PAQ: Spence, Helmreich & Stapp, 1973), in addition to a few novel traits included by the researchers. A total of 60 traits (20 stereotypically feminine, 20 stereotypically masculine, and 20 comparatively gender neutral) were included in the final scale (Appendix 2). The gender-neutral traits were included to ensure the scale ranged

appropriately from masculine to feminine and to give participants a meaningful neutral point. Traits were pre-screened for familiarity and desirability by a team of three researchers (the author, research supervisor, and another PhD student), and also by the team of undergraduate final year project students who assisted with data collection. Three different rating dimensions were included in the scale: (1) Very feminine to very masculine (mid-point: neither masculine nor feminine), (2) Very desirable in men to very undesirable in men (mid-point: neither undesirable nor desirable), and (3) Very desirable in women to very undesirable in women. Ratings were done for each on a 5-point Likert scale.

Procedure

Participants completed the pencil-and-paper scale individually in the Maynooth University experimental cubicles. Informed consent was obtained prior to participation (see Appendix 1 for a copy of the consent form) and the entire scale took approximately 20 minutes to complete. Once the scale was finished, participants were debriefed and thanked for their time.

Ethical Issues

This research was approved by the Maynooth University Ethics Committee and was in compliance with both the Psychological Society of Ireland's Code of Professional Ethics (PSI, 2011) and the Maynooth University Policy on Research Ethics. Participants were informed of the study's structure and broad aims before signing the consent form. The content and procedures posed no particular ethical risk, but participants were advised they could cease participation at any time. As no personal data were collected and responses were anonymous, they were informed they could not withdraw after completing the study.

Data Analysis

The goal of this analysis was to identify appropriate traits for the four categories of interest: (1) desirable feminine; (2) desirable masculine; (3) undesirable feminine; and (4) undesirable masculine. Traits were thus first ranked in terms of their mean rating on each dimension. To produce the desirable femininity categories, gender/desirability ratings were produced by adding the 'masculinity-femininity' and 'desirability in women' scores and dividing them by two. For the masculinity categories, the same was done but the gender rating was first inverted (given that lower scores indicated a more masculine rating). For the undesirable categories, the same was done but this time the desirability scores were inverted.

2.2.2 Results

All analyses were conducted using JASP (version 0.7.5 Beta 2, University of Amsterdam, Netherlands). All tests were two-tailed with alpha set at .05. Unless otherwise specified, data were normally distributed. Parametric tests were used throughout.

Trait Ratings

Gender Ratings

Complete gender and desirability ratings for all 60 traits are in Appendix 3. For space constraints, only the top 10 most highly-ranked for each category are included in-text (Table 2.1). Ratings were not divided by participant gender at this stage due to the small sample size. Traits clustered generally as expected, with the pre-selected feminine traits being evaluated as feminine and the masculine traits as masculine. Three interesting trends emerged in terms of the ratings: first, there were more feminine than masculine traits in the data set (that is, more traits with a mean value between 4-5 than 1-2). Second, desirability ratings were gendered as expected, with stereotypically masculine traits being less desirable in women, and vice versa.

Third, an interaction emerged between trait gender and valence, such that the most feminine traits were generally more positively valenced than the most masculine traits. These trends will all be explored in more detail in Study Two.

Table 2.1

Top 10 most highly-ranked for the masculinity-femininity subscale

Feminine Traits		**Masculine Traits**	
Trait	Mean rating	Trait	Mean rating
Polite	3.65	Unemotional	1.88
Bossy	3.75	Aggressive	2
Compassionate	3.75	Competitive	2
Empathetic	3.9	Dominant	2.05
Affectionate	3.97	Athletic	2.08
Sensitive	3.98	Arrogant	2.18
Bitchy	4.08	Insensitive	2.2
Gentle	4.13	Violent	2.2
Emotional	4.17	Forceful	2.38
Nurturing	4.25	Tactless	2.53

Note. Traits were rated on a five-point Likert-Style scale from very masculine (1) to very feminine (5).

Desirability Ratings

Desirability ratings were generally gendered (Table 2.2) with traditionally feminine traits evaluated as more desirable in women and traditionally masculine traits more desirable in men.

Table 2.2

Top 10 most highly-ranked traits for the "desirability in women" and "desirability in men" subscales

Traits Desirable in Women		**Traits Desirable in Men**	
Trait	Mean rating	Trait	Mean rating
Sincere	4.40	Loyal	4.48
Loyal	4.35	Witty	4.30
Nurturing	4.35	Driven	4.28
Affectionate	4.25	Independent	4.25
Communicative	4.25	Charismatic	4.23
Compassionate	4.23	Sociable	4.18
Optimistic	4.13	Communicative	4.13

Polite	4.13	Polite	4.13
Sociable	4.13	Sincere	4.13
Gentle	4.13	Helpful	4.10

Note. Traits were rated on a five-point Likert-Style scale from very undesirable (1) to very desirable (5).

Stimulus Selection

Composite scores were generated for each trait, which combined the gender and desirability in women/men ratings. These scores were ranked from highest to lowest and the top eight of each were selected for a team review (Table 2.3). As mentioned, there was no pre-set criterion/score minimum for trait selection, given the lack of similar or comparable research in this area. Instead, the top eight were screened by three researchers for familiarity, representativeness, face validity and redundancy due to repetition. Four traits were selected for each category and are highlighted in the table in bold.

Table 2.3

Composite gender-desirability scores for stimulus categories

Desirable Feminine		**Desirable Masculine**		**Undesirable Feminine**		**Undesirable Masculine**	
Trait	Rating	Trait	Rating	Trait	Rating	Trait	Rating
Nurturing	4.30	Athletic	3.91	**Bitchy**	4.21	Violent	4.24
Affectionate	4.11	**Witty**	3.85	**Insecure**	3.94	**Aggressive**	4.10
Gentle	4.11	**Charismatic**	3.81	**Bossy**	3.88	**Arrogant**	4.09
Sensitive	4.03	Loyal	3.74	Jealous	3.84	**Unemotional**	4.09
Compassionate	3.99	**Competitive**	3.70	**Helpless**	3.71	**Insensitive**	4.04
Empathetic	3.95	**Decisive**	3.63	Secretive	3.71	Callous	3.95
Communicative	3.90	Independent	3.63	Dishonest	3.70	Tactless	3.90
Polite	3.89	Practical	3.63	Vain	3.68	Dishonest	3.90

Note: For the desirable subscales, higher scores represent more desirable traits (in women and men). For the undesirable subscales, higher scores represent more undesirable traits. The final selected stimuli are in bold.

2.3 Study Two

Study Two will build on the pilot study by validating the chosen stimulus categories. The same gender traits scale will be administered to a larger sample of young Irish adults, along with a new hiring preference task. Ratings will be compared across the two studies to ensure the categories are valid and represent the four constructs. Responses will also be compared across male and female participants, now that the sample is large enough for meaningful comparisons.

2.3.1 Method

Participants

Completed questionnaires were gathered from a sample of 228 respondents recruited from the Irish undergraduate population. Participants were again recruited using convenience sampling and the same inclusion criteria as in the previous study were used here. All individuals were aged between 18 and 25 (M = 20.7, SD = 1.60). Participants were again provided with an open-ended response format for providing gender information, with 111 identifying as female and 117 as male. All participants were currently living in Ireland and enrolled on an undergraduate degree. Participation was voluntary and no remuneration was offered.

Materials

Trait Rating Scale

The trait rating scale was identical to the pilot study.

Hiring Task

Hiring preference was assessed using a brief task in which participants were presented with two identical questions about their hiring preferences for an office job (Table 2.4). The generic title "office job" was selected due to its non-specific nature and absence of any salient gender connotations (see Reuben et al., 2014 for a recent

list of gender employment stereotypes). Using the traits selected in the pilot study, the first question in this task ascertained hiring preferences for a stereotypically masculine over a stereotypically feminine person, while the second item asked who they would prefer to let go again between a masculine or feminine person. The category labels (masculine or feminine) were not explicitly stated in the questions. The last item more explicitly asked for their preference for a man relative to a woman. For all items, participants were presented with a third response option: "I prefer not to answer." This was employed as a catchall for non-responses that may be due to any number of preferences (e.g., neutral/neither/both/disagree with the premise of the question, etc.) and to eliminate the possibility of inaccurate data produced by forced-choice responding.

Table 2.4

Questions and response options in the hiring task

If you were an employer hiring for an office job, which of the following two categories of people would you be more likely to hire?
Someone who is nurturing, gentle, affectionate, and sensitive
Someone who is witty, charismatic, competitive, and decisive
I prefer not to answer
If you were an employer cutting staff from an office job, which of the following two categories of people would you be more likely to let go?
Someone who is aggressive, unemotional, insensitive, and arrogant
Someone who is bitchy, insecure, bossy, and helpless
I prefer not to answer
If you were an employer hiring staff for an office job, which of the following two categories of people would you be more likely to hire?
A man
A woman
I prefer not to answer

Procedure

All study sessions were conducted one-to-one in individual experimental cubicles. Participants were briefed on the general nature and structure of the study

and were given a short overview of the study's subject matter (i.e., contemporary beliefs about gender) prior to participation. Written informed consent was provided by all individuals prior to participation. Participants completed the hiring task first, followed by the trait rating scales. Upon completion of the tasks, participants were fully debriefed and thanked for their time.

Data Analysis

The same analytic approach will be used here as in the pilot study, with traits being ranked from highest to lowest for each rating dimension. Combined gender/desirability scores will be produced for each trait and the selected stimuli will again be screened by a team of researchers. Because of the larger sample size, ratings will also be split by participant gender. To test whether the chosen stimulus categories are representative of the four constructs (i.e., desirable/undesirable masculinity and femininity), single-sample t-tests will be run on each to assess whether they significantly skew to the left and right of the scale. A MANOVA will then be run to examine any gender differences in average ratings for the four categories. For the hiring task, the distribution of candidate preferences (i.e., for the stereotypically feminine or masculine person) will be presented, again split by participant gender.

2.3.2 Results

Gender Traits Scale

Male and female participant ratings for each dimension are presented below. Again, due to space constraints, only the top 10 most highly-ranked traits for each category will be included in the tables. Complete data can be found in Appendix 4.

Masculinity-Femininity Ratings

As evidenced by the data in Tables 2.5 and 2.6, very few traits were rated as highly masculine or feminine. Participants generally had more "extreme" (or more gendered) ratings for the traits pertaining to their own gender; that is, female participants had higher ratings for the more stereotypically feminine traits, while and male participants had lower ratings for the more stereotypically masculine traits. As in the pilot study, the feminine traits overall seem more positively valenced than the masculine traits. Of the top 10 feminine traits, only two are overtly negative ("bitchy" and "insecure"), in comparison to at least six of the masculine traits ("Arrogant", "Violent", "Forceful", "Blunt" and "Aggressive", "Insensitive"). "Unemotional", "Competitive" and "Dominant" have more mixed social connotations and could be desirable or undesirable depending on the context (e.g., occupational versus social settings).

Table 2.5

Top 10 most highly-ranked masculine traits

Trait	Mean rating	Women (*n* = 111)	Men (*n* = 117)
Aggressive	2.16	2.25	2.08
Violent	2.21	2.31	2.12
Arrogant	2.21	2.19	2.24
Unemotional	2.21	2.22	2.21
Forceful	2.29	2.35	2.24
Dominant	2.32	2.42	2.21
Insensitive	2.41	2.44	2.38
Athletic	2.46	2.47	2.45
Competitive	2.50	2.68	2.34
Blunt	2.59	2.46	2.72

Note. Traits were rated on a five-point Likert-Style scale from very masculine (1) to very feminine (5).

Table 2.6

Top 10 most highly-ranked feminine traits

Trait	Mean rating	Women (*n* = 111)	Men (*n* = 117)
Bitchy	4.02	4.02	4.03

Nurturing	4.02	4.01	4.04
Emotional	3.89	3.90	3.89
Gentle	3.86	3.86	3.86
Sensitive	3.80	3.79	3.81
Affectionate	3.75	3.75	3.75
Compassionate	3.71	3.57	3.86
Empathetic	3.65	3.52	3.78
Communicative	3.61	3.54	3.68
Insecure	3.57	3.55	3.59

Note. Traits were rated on a five-point Likert-Style scale from very masculine (1) to very feminine (5).

Desirability Ratings

The same pattern as in the pilot dataset was observed here, with desirability ratings generally being quite gendered (see Tables 2.7 and 2.8). As with the masculinity-femininity ratings, male participants tended to give higher (i.e., more extreme) desirability ratings than female participants.

Table 2.7

Top 10 most highly-ranked traits for the "Desirability in Men" subscale

Trait	Mean rating	Women (*n* = 111)	Men (*n* = 117)
Loyal	4.65	4.58	4.73
Witty	4.40	4.41	4.39
Sociable	4.37	4.34	4.41
Driven	4.33	4.27	4.40
Capable	4.32	4.29	4.35
Communicative	4.32	4.19	4.45
Charismatic	4.32	4.20	4.44
Sincere	4.30	4.17	4.44
Helpful	4.29	4.21	4.36
Polite	4.28	4.18	4.39

Note. Traits were rated on a five-point Likert-Style scale from very undesirable (1) to very desirable (5).

Table 2.8

Top 10 most highly-ranked traits for the "Desirability in Women" subscale

Trait	Mean rating	Women (*n* = 111)	Men (*n* = 117)
Loyal	4.68	4.64	4.73
Affectionate	4.49	4.42	4.56

Polite	4.46	4.39	4.53
Compassionate	4.45	4.38	4.52
Sincere	4.44	4.39	4.49
Communicative	4.41	4.32	4.50
Sociable	4.40	4.32	4.49
Helpful	4.35	4.29	4.41
Nurturing	4.32	4.20	4.44
Optimistic	4.31	4.22	4.41

Note. Traits were rated on a five-point Likert-Style scale from very undesirable (1) to very desirable (5).

Stimulus Selection Review

Ratings for the four stimulus categories are presented in Table 2.9. Overall, the Masculine and Feminine categories tended to be skewed in the expected direction (i.e., feminine traits fell to the right of neutral and masculine to the left). Ratings did not change considerably from Study One, with all the selected traits again in the top eight most highly-ranked in each category.

Table 2.9

Composite gender-desirability scores for stimulus categories

Desirable Feminine		**Desirable Masculine**		**Undesirable Feminine**		**Undesirable Masculine**	
Trait	Rating	Trait	Rating	Trait	Rating	Trait	Rating
Nurturing	4.17	**Witty**	3.87	**Bitchy**	4.19	Violent	4.25
Affectionate	4.12	**Charismatic**	3.82	**Insecure**	3.83	**Aggressive**	4.08
Compassionate	4.08	Loyal	3.80	Dishonest	3.80	**Arrogant**	4.08
Gentle	4.04	Athletic	3.73	Selfish	3.79	Unreliable	4.05
Communicative	4.01	**Competitive**	3.68	**Bossy**	3.79	**Unemotional**	3.96
Polite	3.93	Capable	3.65	**Helpless**	3.75	Forceful	3.94
Empathetic	3.88	**Decisive**	3.63	Jealous	3.73	**Insensitive**	3.93
Sensitive	3.87	Sociable	3.62	Secretive	3.70	Dishonest	3.91

Note: For the desirable subscales, higher scores represent more desirable traits (in women and men). For the undesirable subscales, higher scores represent more undesirable traits. The stimuli that were selected for use in the measures are highlighted in bold.

One-sample t-tests were conducted to explore whether the masculinity/femininity ratings for these categories were significantly skewed in the expected left/right direction. Average ratings were generated for each category for use in the analyses. Assuming a test value of 3 (a rating of "neither masculine nor feminine"), results revealed ratings for both the Masculine and Feminine categories were on average significantly to the left and right of neutral, for both male and female participants (see Table 2.10). Table 2.10

One-sample t-test results for the average Masculine and Feminine category ratings.

	Women (*n* = 111)			**Men (*n* = 117)**		
	t	*df*	*p*	*t*	*df*	*p*
Desirable Fem	15.17	110	<.001***	18.85	116	<.001***
Desirable Masc	-6.20	110	<.001***	-10.30	116	<.001***
Undesirable Fem	11.37	110	<.001***	13.97	116	<.001***
Undesirable Masc	-14.73	110	<.001***	-15.99	116	<.001***

*Note:*** denotes significance after a Bonferroni-correction ($p < .00625$).

Gender Differences

Masculinity-Femininity Ratings

As the four stimulus categories (desirable/undesirable feminine/masculine traits) will be used in the IRAP for future experiments, it was important to analyse gender differences in how they were rated. A MANOVA comparing the average masculinity-femininity ratings for these categories revealed a significant main effect for gender, Wilks $\lambda = .95$, $F(4, 223) = 2.77$, $p = .028$. Follow-up univariate tests found significant differences on the Desirable Masculine category only, $F(1, 226) = 7.66$, $p = .006$ (all other p-values > .5), with ratings lower (i.e. more masculine) for male relative to female participants.

Table 2.11

Mean Masculinity-Femininity ratings for the stimulus categories by gender.

	Group	Mean	SD	df	F	*p*
Desirable Feminine	Males	3.86	0.49	1	0.02	0.866
	Females	3.86	0.6			
Desirable Masculine	Males	2.57	0.45	1	7.66	0.006
	Females	2.73	0.45			
Undesirable Feminine	Males	3.57	0.44	1	0.35	0.552
	Females	3.61	0.56			
Undesirable Masculine	Males	2.23	0.52	1	0.45	0.504
	Females	2.27	0.52			

Note. Traits were rated on a five-point Likert-Style scale from very masculine (1) to very feminine (5). Means were calculated for each stimulus categories by averaging the ratings for the four traits.

Desirability Ratings

Two additional MANOVAs were run to compare male and female participants' desirability ratings. For the 'Desirability in Men' subscale, analyses found a significant main effect for gender, Wilks $\lambda = .825$, $F\ (4, 223) = 11.827$, $p < .001$. Univariate analyses found significant differences on all of the subscales except the 'Undesirable Feminine' scale ($p = .438$: see Table 2.12).

Table 2.12

Descriptive statistics and ANOVA results for 'Desirability in Men' ratings for the four stimulus categories by gender

	Group	Mean	SD	df	F	*p*
Desirable Feminine	Males	3.64	0.66	1	19.11	<.001
	Females	4.00	0.61			
Desirable Masculine	Males	4.09	0.52	1	4.138	0.043
	Females	3.95	0.51			
Undesirable Feminine	Males	1.77	0.54	1	0.60	0.438
	Females	1.71	0.49			
Undesirable Masculine	Males	1.97	0.70	1	38.42	<.001
	Females	1.46	0.55			

Note. Traits were rated on a five-point Likert-Style scale from very undesirable (1) to very desirable (5). Means were calculated for each stimulus categories by averaging the ratings for the four traits.

A MANOVA run on the 'Desirability in Women' subscale found no significant main effect for gender, Wilks $\lambda = .975$, $F(4,223) = 1.45$, $p = .218$. Univariate analyses were therefore not run on these data, though descriptive statistics for male and female participants can be found in Table 2.13.

Table 2.13

Mean 'Desirability in Women' ratings for the four stimulus categories by gender.

	Group	Mean	SD
Desirable Feminine	Males	4.18	0.46
	Females	4.23	0.53
Desirable Masculine	Males	3.93	0.59
	Females	3.94	0.55
Undesirable Feminine	Males	1.84	0.62
	Females	1.78	0.63
Undesirable Masculine	Males	1.69	0.61
	Females	1.53	0.50

Note. Traits were rated on a five-point Likert-Style scale from very undesirable (1) to very desirable (5). Means were calculated for each stimulus categories by averaging the ratings for the four traits.

Hiring Task

Response distributions for the hiring task items varied across gender groups and questions. For the item ascertaining hiring preference for the "feminine" over the "masculine" individual, responses were heavily skewed in the stereotypically masculine direction (94% of the responses for men and 88% for women). Responses for the firing item were more evenly split, with approximately half of both samples selecting the "masculine" person and half the "feminine". For the more explicit gender preference question, approximately two thirds of each group selected "I prefer not to answer". For female participants, 33% elected to choose the woman and 4% the man. Male responses on this item were more evenly divided between the two, with 14% selecting the woman and 15% the man.

Gender differences in responses were assessed using Chi-square tests for independence. Given that we were only interested in those who selected either the man/woman or the "masculine"/"feminine" person, analyses on each item excluded those who selected "I prefer not to answer". These analyses revealed no significant difference between male and female responses on the first two items (hiring and letting go of the "masculine" and "feminine" individuals), though a significant association was found between participant gender and hiring preference for the explicit gender question, χ^2 (1, 75) = 10.52, $p < .001$, phi = .4. These results should be interpreted somewhat tentatively, however, given that data from only a third of the sample was analysed for this item.

2.2.3 Summary

In terms of trait ratings, there were a number of interesting trends in the data. First, traits were generally gendered as expected, with participants evaluating certain traits as masculine and others as feminine. Desirability ratings were similarly gendered, such that traditional femininity seems to be valued in women and masculinity in men. Second, participants generally rated their own gender categories more strongly, with women rating feminine traits as more feminine and men rating masculine traits as more masculine. Gender differences were found in how the four stimulus categories were rated, though follow up tests found this to be driven by differences on the "Desirable Masculine" stimulus category. Desirability ratings differed across male and female participants, though only on the "Desirability in Men" subscale. Third, a gender-valence interaction emerged across the two studies, with the most masculine traits generally being negatively valenced and the most feminine traits positively valenced. For the hiring task, responses were overwhelmingly skewed towards the stereotypically masculine individual, for both

male and female participants. There were significant gender differences in terms of the masculinity ratings for these traits (with male participants evaluating them as more masculine than female participants), but single-sample t-tests found both groups evaluated the traits as significantly more masculine than feminine.

2.4 Discussion

Trait Ratings

Across two studies, a total of 268 young Irish adults rated 60 personality traits in terms of their gender (i.e., masculinity-femininity), desirability in women, and desirability in men. Traits were generally evaluated as expected and clustered into three broad groups: traditionally masculine, gender-neutral, and traditionally feminine. Desirability scores were similarly gendered, with stereotypically feminine traits more valued in women than men and stereotypically masculine traits more valued in men than women. While ratings were not particularly "extreme" for either the masculine or feminine categories (that is, there were very few traits with mean ratings close to 1 or 5), the distinction suggests that gender trait stereotypes and biases remain prevalent among young Irish adults.

The masculine and feminine categories identified here map largely onto the communality-agency distinction noted in previous research (Eagly & Steffen, 1982; Conway, Pizzamiglio & Mount, 1996; Ridgeway, 2001). Communality (or communion) broadly refers to traits related to interpersonal care (e.g., compassion, empathy, or selflessness) and agency to traits related to individual freedom and success (e.g., independence, competitiveness, decisiveness). The differential ascription of communal and agentic traits is considered problematic given the inherent differences in status across the two categories (Conway et al., 1996; Eagly & Karau, 2002; Glick & Fiske, 2002; Koenig, 2018). As argued by Ridgeway

(2001), social evaluations of status by and large rest on assumptions of competency, and research suggests a considerable overlap between agency and competency stereotypes (see Jost, Kivetz, Rubini, Guermandi, & Mosso, 2005 and Koenig et al., 2010 for empirical reviews). As such, it is socially relevant that the current sample considered agentic traits to be more masculine than feminine.

While agentic traits are considered to be more socially valuable (in terms of exerting social influence or accessing resources), it is important to note that the feminine traits were more positively valenced on average than the masculine traits. As mentioned in the introduction, research shows that femininity and womanhood are often evaluated positively, especially in the abstract. Occasionally referred to as the "women-are-wonderful" effect, studies suggest both men and women tend to hold more favourable feminine than masculine stereotypes (e.g., Eagly & Mladinic, 1989; 1994; Eagly, Mladinic & Otto, 1991; Glick & Fiske, 2001), and that positive attributes may be more readily attributed to women over men (Rudman & Goodwin, 2004). It is worth noting, however, that the positively valenced feminine traits identified here (and in previous research) are all consistent with the traditional feminine gender role. That is, they are contextually specific and cast women as the selfless, compassionate care-giver. These data therefore support arguments made elsewhere that the "women-are-wonderful" effect may be better described as the "women-are-wonderful-*when*" effect (Dovidio, Glick & Rudman, 2008; Rudman & Glick, 2014), with the "when" referring exclusively to contexts where women display role-appropriate behaviour. As discussed in the previous chapter, theorists argue that it is not necessarily true that femininity is generally negative and masculinity is generally positive under patriarchy; rather, it is that women are denied the agentic traits which are more readily associated with status and power. The IRAP

methodology will be a useful measure to investigate this sex-gender-valence-power relationship in more detail, as it provides an index of how readily both masculine and feminine traits are attributed and denied to men and women. As such, it will allow for an investigation of the above claims in more detail.

The relative lack of desirable masculine traits in the data also supports theoretical arguments around the androcentric positioning of masculinity within the gender binary. As discussed previously, a key argument made in feminist literature is that masculinity is centralized and prioritized within binary systems, while femininity is cast as the "other" (e.g., Bem, 1993). From this perspective, it could be that the agentic traits – though still stereotypically masculine and considered desirable in men – are comparably less likely to be explicitly gendered, particularly among contemporary samples. A related explanation is that responses reflect a growing awareness of women's changing roles in society. Identity research over the past two decades does suggest women are developing more agentic self-gender stereotypes (Diekman and Eagly, 2000; Ebert et al., 2014; Fine-Davis, 2013; Spence & Buckner, 2000; Twenge, 1997; Wilde and Diekman, 2005), potentially as a result of observing more women enter more traditionally masculine spheres (Eagly & Karau, 2002). The gender differences on this category support this assumption, as it would make sense that women in particular would be less inclined to evaluate these traits as masculine. Again, this speaks to the broader complex relationship between sex, gender and power, which will be explored in more detail in later chapters.

Stimulus Selection

Using a transparent and stage-like selection process, the author and a small team of researchers identified four exemplars for each of the four stimulus categories from the overall set of 60 traits. Exemplars were selected on the basis of their

composite gender-desirability rating, familiarity, and recognisability. Based on the recommendations from Nosek (2007), no more or less stimuli were chosen than were required to represent the overarching category on average. As the ratings did not change significantly across the pilot and main studies, the team of researchers decided to retain the same exemplars identified in the pilot study. Doing so will allow us to use an identical hiring task in the next Chapter, and therefore add to the dataset for that task. It will also allow for a direct comparison between implicit/explicit trait gendering and occupational gender preference towards the end of the thesis.

The comparable weakness of the "Desirable Masculine" category should be noted at this stage as it may have relevance in later chapters. As mentioned above, this set of traits was not as strongly gendered as the other three categories, and particularly for female participants. Given that analyses still found these traits to be significantly more masculine than feminine for all participants, it does not cause immediate concern around the category's potential effectiveness in the IRAP or hiring task. However, it is an important caveat that should be taken into account during the interpretation of results and investigated again towards the end of the thesis.

Hiring Preferences

Responses on the hiring task cohere with existing feminist theories of gender order and the idea that greater cultural value is placed on masculine traits. When asked which sort of person they would rather hire, participants expressed an overwhelming preference (91% overall) for the stereotypically masculine traits. While this question did not specify the sex of the candidate directly, analyses suggest that the hirable traits were considerably more male than female. These data therefore

support previous research identifying a link between masculine traits and occupational competency (Eagly & Carli, 2007; Koenig, Eagly, Mitchell, & Ristikari, 2011) and, again, the argument that Western societies continue to implicitly prioritize masculinity (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Johnson, 2005; Serano, 2009).

Somewhat surprisingly, responses on the firing/letting go item did not differ significantly by either participant or target gender. These results are difficult to contextualize within existing literature, given that most to date have looked at evaluations of hirability or advancement potential. However, one possible explanation is that women (or stereotypically feminine individuals) are not actually more likely to be let go from an office job than male/masculine individuals. Theories of occupational gender discrimination often use the "glass ceiling" metaphor to describe the insidiousness of gender barriers in the workplace (Koenig et al., 2011). Rather than being at an increased risk of being let go from their jobs, such theories suggest that women may instead be less likely to be hired for valued positions or indeed put forward for leadership or more senior roles (see Eagly & Karau, 2002). This explanation would of course require further testing and substantiation; however, for the purposes of the current thesis, the firing item does not appear to be a useful measure of occupational gender bias and there is no strong theoretical rationale for its inclusion. It will thus be removed for future studies.

It is interesting to note that, contrary to existing work on hiring bias (see Isaac et al. 2015 for a meta-analysis), responses on the direct gender-preference question (i.e., between a man and a woman) were not in a pro-male direction. A large proportion (two-thirds) of the current sample elected not to express an explicit preference between male and female candidates (i.e., they selected the option "I

prefer not to answer"); of those who did, however, the majority selected the female in favor of the male. Though several factors may have contributed to this response pattern, the widespread unwillingness to express an explicit preference – at least a male one – is worth noting. For instance, it may reflect either self-presentational distortions (i.e., social desirability biases), which could attest to a growing awareness on behalf of participants of the issues women face in occupational contexts, or to a willingness to prioritize women, at least in theory.

Conclusion

This chapter identified stimuli for use in the gender binary IRAPs and occupational preference task. Four stimulus categories were selected, each containing four exemplars. Analyses suggested these categories were significantly gendered, with the masculine categories rated as more masculine than feminine and vice versa. The gender categories overlapped considerably with the agency-communality divide within the stereotype literature; as such, these terms (masculinity/agency and femininity/communality) will be used interchangeably for the remained of the thesis. This chapter also added to the scant literature on gender trait stereotypes in Ireland, and provided preliminary evidence that Irish society continues to implicitly associate men, masculinity and occupational success.

Chapter 3

Binary Biases and Occupational Discrimination

3.1 Introduction

Chapter Two investigated contemporary gender trait stereotypes in a young Irish sample. Data generated from two surveys provided preliminary evidence that, first, masculinity and femininity remain distinct social categories (with separate sets of attributes and traits) and, second, stereotypes are generally framed around traditional agency and communion. This Chapter will expand on these findings by using the IRAP to measure not only automatic role-congruent trait stereotypes (i.e., men-agentic and women-communal biases), but additionally participants' resistance to ascribing role-*incongruent* traits to women and men (i.e., men-*not*-communal and women-*not*-agentic biases). Doing so will provide an experimental assessment of the relational and oppositional structure of gender, which is considered to be a key facet of gender binarist ideologies (e.g. Bem, 1993; Butler, 2002; Hegarty, Ansara & Barker, 2018).

A secondary aim of this chapter is to further assess the feminist claim that masculinity is more socially valued than femininity. Across three separate experiments, participants' will be required to express a preference for the masculine-agentic and feminine-communal traits identified in Chapter Two (and used here in the IRAP). Because gender evaluations are known to be context-specific (Eagly & Karau, 2002; Ridgeway, 2011), we will examine preferences for sex and gender in an occupational context only. The workplace was selected as a useful starting point given that (a) there is a large body of existing theoretical and empirical research in this area to draw from (see Koenig et al., 2011 for a recent meta-analysis); (b) simulated hiring provides a relatively straightforward and uncontroversial way to directly assess gender preferences (Rudman & Glick, 2001); and (c) employment is

considered one of the primary processes through which gender dynamics play out in contemporary society (e.g., Ferguson, 2013).

3.2 Study Three

Gender Bias in the Workplace

Before introducing how the IRAP may be used in this context, it would be useful to provide a brief review of the literature around workplace gender bias. As noted in the General Introduction, women continue to experience considerable discrimination and prejudice in occupational settings. This discrimination includes the well-researched gender pay gap (EIGI, 2018), significant under-representation in lucrative and traditionally male-dominated spheres (e.g., finance, politics, science and technology: see CSO, 2016 for national and UN, 2018 for global statistics), and a reduced likelihood of occupying or being considered for leadership roles (UNESCO, 2017). Women have also been found to experience many indirect forms of workplace discrimination, ranging from gender-biased or male-centric policies (e.g., around maternity or health coverage) to an increased risk of experiencing bullying, exclusion, and harassment (including sexual or gender-based harassment: see Stamarski & Son Hing, 2015 for a comprehensive review).

Historically, gender discrimination in the workplace was understood as the consequence of sexism or misogyny. As with many forms of inequality, it was believed that women's lower participation in the workforce resulted from negative attitudes about their inherent incompetence or ineffectiveness (Koenig et al., 2011). While explicit misogyny was indeed a significant barrier to women's progress in previous decades, more recent attitude surveys would suggest sexism (or at least hostile or open forms of sexism) have lowered significantly across many Western societies (see Eagly & Mladinic, 1994; Haines et al., 2016; Langford & MacKinnon,

2000) including Ireland (Fine-Davis, 2013). As such, anti-women prejudice is unlikely to be an accurate or complete explanation of contemporary workplace inequality on the whole (Eagly & Karau, 2002; Heilman, 2001; Hogue & Lord, 2007; Lyness & Heilman, 2006). Another historic explanation for the lack of women in lucrative roles was the so-called "pipeline problem", believed to be created by the shortage of appropriately skilled, motivated, or experienced female candidates (Browne, 1999; Goldberg, 1993; Greenhaus & Parasuraman, 1999). As discussed in the introduction, however, this explanation is also unlikely to account for modern gender inequality given the broadly equal numbers of women and men who now graduate University and enter the skilled workforce (e.g., CSO, 2016), and express ambition to pursue more senior or lucrative roles (e.g., Diekman & Eagly, 2008; Gino, Wilmuth & Brooks, 2015).

Social Role Theory and the Role Congruity Hypothesis

Instead, workplace inequality is increasingly understood to be a feature of our broader gender role system. To return briefly to the Social Role Theory (SRT) discussed in Chapter One, Eagly and colleagues propose that all gender inequality results from society's generalized preference and expectation that men and women adhere to their traditional gender role (SRT: Eagly, 1987; Eagly, Wood, & Diekman, 2000; Eagly & Wood, 2016). Within the SRT, gender roles encompass the comprehensive set of both *descriptive* and *injunctive* norms around male and female behaviour, with descriptive norms referring to assumptions about men and women's actual behaviour, and injunctive norms (also called prescriptive norms) the assumptions around what men and women ought to be (Eagly et al., 2000). Because people tend to infer a correspondence between a group's ascribed social norms and their natural dispositions (see Eagly & Koenig, 2008), gender roles also include

beliefs around men and women's inherent interests and competencies across a range of areas (e.g., caregiving, leadership, etc.). According to the SRT these assumptions underpin inequality because it means that, for example, it is not just that women are not typically agentic, but that they both *cannot* and *ought not to* be (Burgess & Borgida, 1999; Eagly & Mitchell, 2004; Fiske & Stevens, 1993; Gill, 2003; Prentice & Carranaza, 2002).

Eagly and Karau's Role Congruity Theory (RCT: 2002) is an extension of the SRT specifically to occupational settings. Relying on the same core principle that gender inequality arises because of the expectation and assumption of differentiated social roles, the RCH elaborates on the processes through which women are hypothesized to experience discrimination in the workplace. Similar to other systemic feminist accounts of prejudice, the RCH proposes women's subordinate status in the workplace results from the implicit societal overlap between masculinity and leadership. Because gender is relationally structured (i.e., that which is male is importantly *not-female*, and vice-versa), there is thus a lack of fit between femininity and occupational success (Cejka & Eagly, 1999; Heilman, 2001). A range of classic studies support the "think manager-think male" assumption, suggesting masculine stereotypes overlap considerably with those required for a successful leader (Arkkelin & Simmons, 1985; Carli, LaFleur, & Loeber, 1995; Heilman, Block, Martell & Simon, 1989; Duehr & Bono, 2006; Jackson & Engstrom, 2007; Powell & Butterfield, 1989; Schein, 1975), and also that men are considered more naturally suitable for leadership roles (Dodge, Gilroy & Fenzel, 1995; Glick, Zion &Nelson, 1988; Heilman et al., 1995). Feminine stereotypes on the other hand overlap as expected with traits considered desirable in more communal or supportive

occupational roles (e.g., nursing, working with the poor, peace efforts, etc.; Carnes et al., 2015; Mueller, 1986; Sapiro, 1983).

In addition to the research into leadership stereotype overlaps (and mismatches), the RCH is further evidenced by research showing the social punishment experienced by women when they violate (or are perceived as violating) traditional feminine norms. Often referred to as the "backlash effect", a large body of research suggests women are subject to a range of organisational and social penalties when they display agentic behaviour or leadership styles (Rudman, 1998; Phelan & Rudman, 2010). In the context of hiring, for example, studies show that although agentic women tend to be evaluated as competent, they are often less likely to be hired on the grounds of being socially incompetent, hostile, or abrasive (Phelan et al., 2008; Rudman, 1998; Rudman & Glick, 1999, 2001; Rudman et al., 2009; Uhlmann & Cohen, 2007). Similar effects have been found across studies simulating promotion potential (Heilman et al., 2004; Rudman et al., 2009), salary negotiation (Babcock & Laschever, 2003; Bowles et al., 2007), or leadership evaluations (Ayman, Korabik & Morris, 2009; Butler & Geis, 1990). A meta-analysis by Eagly, Karau and Makhijani (1995), for instance, showed that women who adopt more directive or assertive management styles are generally evaluated as less effective or competent than either similar men or indeed women with more inclusive or communal styles. Similarly, research suggests women are penalized significantly more harshly than men if they discipline or criticize subordinate members of staff (Atwater et al., 2001; Sinclair & Kunda, 2000). Other studies suggest women (but not men) are required to strike a balance between agentic and communal traits to be successful. Johnson, Murphy, Zewdie and Reichard (2008), for instance, found that female leaders were only evaluated as competent if they managed to display both

"strong" and "sensitive" traits; by contrast, for male leaders only needed to be rated highly on strength.

Backlash for gender role deviations can also be experienced by women in subtler ways. Research into workplace "emotion politics", for instance, revealed that while emotional displays of anger can significantly lower perceptions of a woman's status or competence, they do not have any influence on a man's (Brescoll & Uhlmann, 2008; Rudman, 1998). Studies suggest a similar gender bias in evaluations of men and women's volubility, or the amount of time they spend talking or contributing in a group conversation. For example, two studies showed that when presented with written or audio-visual descriptions of a man and woman contributing equally to a group meeting, people rate the woman as having spoken more (Cutler & Scott, 1990; Brescoll, 2011). Interestingly, effects in both studies were strongest when the woman was talking in a group with both women and men (as opposed to just women), suggesting that women are especially expected to adopt a more role-congruent, subordinate position when men are present. Taken together, it's worth noting that the body of evidence challenges the neoliberal feminist argument that women simply need to better adhere to the agentic stereotype in order to be successful (c.f. Sandberg, 2015). Tangentially, these findings also cohere with Butler's argument around discrimination being one of the processes through which the gender binary is maintained (Butler, 2002). In socially punishing women who behave in an agentic or traditionally masculine manner, women are discouraged and/or prevented from entering powerful male-dominated spaces, and thus the patriarchal order is maintained.

Interestingly, research suggests that gender role violations are not dealt with in a symmetrical way across genders or occupational categories. As discussed

previously, studies show that overall feminine men may experience greater societal backlash than masculine women (Blashill & Powlishta, 2009; Cherry & Deaux, 1978; Cahill & Adams, 1997; Kite & Deaux, 1987). They also report higher degrees of "gender role stress" around being perceived as non-masculine than women do about being non-feminine (Levant et al., 1992; 2003; Smiler, 2004, 2006). In work settings, however, role-incongruent behaviour in men does not seem to elicit the same negative social or career consequences for men as it does for women (Eagly & Koening, 2008; Eagly, Makhijani, & Klonsky, 1992; Rudman & Glick, 1999). Using a hypothetical hiring task, for example, one study found that while women with stereotypically masculine traits were less likely to get hired for a gender-atypical position (e.g., finance manager), the same was not true for men (e.g., for the position of a nurse). Communal men do experience degrees of backlash when they occupy leadership positions (Tepper et al., 1993; Moss-Racusin et al., 2010; Rudman & Fairchild, 2004); however, because agentic traits are generally more valued in the context of leadership, the career consequences are likely to be more severe for women than for men (e.g., through lower income and reduced advancement potential; Moss-Racusin et al., 2010).

Explicit and Implicit Stereotyping and Workplace Discrimination

In addition to research looking at the overlap between masculine/leadership stereotypes and the incongruity between femininity and competence, a number of studies have examined the relationship between gender stereotyping and workplace discrimination. As expected, participants who report higher levels of gender stereotyping or who more strongly endorse gender differences and/or traditional gender roles are less likely to positively evaluate, hire, or promote a female candidate (Davidson & Burke, 2000; Eagly & Carli, 2007; Levinson, 1982;

Neumark, Bank, & Van Nort, 1996; Skewes et al., 2018). While the majority of these studies have used explicit or self-report questionnaires, some have used the IAT or other implicit measures. Rudman and Glick (2001), for example, found that communality-agency biases on an IAT predicted an increased likelihood of discriminating against an agentic female candidate on a hypothetical hiring task. Interestingly, the agentic female was less likely than an agentic male or androgynous female to be hired for either a masculinized or feminized managerial position, despite being evaluated as more competent. Rudman and Kilianski (2000) similarly found that communality-agency biases were associated with higher scores on a separate gender-authority IAT (wherein participants associated women and men with high and low-authority roles). Another study found that men were more readily associated with traits describing a successful manager (e.g., productive, knowledgeable, skilled) and women with an unsuccessful manager (e.g., lacklustre, boring) on an IAT, and that these scores predicted increased workplace rewards assigned to male managers (Latu, Stewart, Myers, Lisco, Estes & Donohue, 2011). Across all of the studies mentioned, men were found to both stereotype more, "essentialise" more, and exhibit more pro-male or pro-agentic hiring preferences.

In sum, the above review highlights two important ideological barriers to gender equality in the workplace: (1) the implicit overlap between masculinity, agency, and occupational success; and (2) the denial of these same traits to women. While previous research (and indeed the previous chapter) provided evidence for the first barrier, the current research will expand by examining the role of both role-congruent and incongruent biases in occupational inequality. Specifically, using the IRAP, these studies will quantify the extent to which both men and women are associated with both communal and agentic traits. While a range of explicit and

implicit attitude measures have been used to examine components of this relationship (i.e., agency-communion gender stereotypes, the impact of stereotypes on hirability, backlash effects, etc.), the IRAP methodology is presented as a unique way to assess both role-congruent and incongruent biases simultaneously. Moreover, given that the previous chapter already assessed the extent to which agentic and communal traits are gendered as masculine and feminine respectively, and also developed a measure of occupational preference that uses the same masculine-agentic and feminine-communal categories as the IRAP, this Chapter outlines a direct assessment of the relationship between binarist beliefs and gender preference in an occupational context.

Current Study

Study Three will utilize the IRAP to assess binarist men-masculine and women-feminine biases; that is, the ascription of certain attributes to women *and not men* and others to men *and not women*. To control for and assess any effects based on stimulus valence, two separate IRAPs will be employed: one for socially desirable masculine and feminine traits and another for socially undesirable traits. To explore whether the traits related to the stimulus category "man/male" are indeed more occupationally valuable than those related to "woman/female" (i.e., to examine the relationship between sex, masculinity, agency and power), this study will also employ the hypothetical hiring task from Chapter Two. In this, participants will be asked to express their hiring preference for a gender-neutral occupation between a man and a woman in one item, and a stereotypically masculine or a stereotypically feminine person in another. The same traits that are employed in the IRAP will be used in this task. A small number of self-report measures will also be included to assess explicit anti-women and gender-normative beliefs. It is hypothesised that

participants will gender the stimulus categories on the IRAP in a binary-consistent manner and that there will be an overall pro-masculine preference on the hiring task. Based on the explicit ratings in Chapter Two and other published studies, it is expected that effects will be higher/more pronounced for male participants on all measures.

3.2.1 Method

Participants

Forty-seven undergraduate students (26 self-identified as female, 21 as male) aged between 18 and 32 years participated in this study (M_{age} = 23.84, SD = 5.49). A sample of convenience was used. The sample comprised White Irish participants, with the exception of two White Western European individuals. Participation was voluntary and no remuneration was offered. Inclusion criteria included fluent English, normal or corrected-to-normal vision, and full use of both hands. Participants were provided with an open-ended response format for gender.

Procedure

All experimental sessions were conducted one-to-one in individual experimental cubicles. Participants were briefed on the general nature and structure of the study and were given a short overview of the study's subject matter (i.e., contemporary beliefs about gender) prior to participation. Written informed consent was provided by the participant, followed by a verbal assessment by the researcher for all inclusion criteria. The general experimental sequence was as follows: hiring task, measures ascertaining self-reported beliefs towards women and gender, and two IRAPs. The IRAP was programmed and presented in JavaScript and the self-report tasks were presented in Google Forms. The order of the implicit measures and the presentation order of the blocks within them were both counterbalanced across

participants (see Appendix 4), in keeping with IRAP methodological convention (see Hussey et al., 2015). Upon completion of all tasks, participants were fully debriefed and thanked for their time.

Materials

Self-Report Measures

Self-reported sexism was assessed using the Modern Sexism Scale (MS: Swim et al., 1995: Appendix 5) and more general gender-normative beliefs using the Heteronormativity Attitudes and Beliefs Scale (HABS: Habarth, 2015: Appendix 6).

Modern Sexism Scale (MS). The MS is a 10-item scale, comprised of two five-item subscales, assesses beliefs about women and gender. The first subscale is intended to assess more traditional anti-women sentiments (e.g., "It is more important to encourage boys than to encourage girls to participate in athletics.") and the second more subtle or contemporary sexist attitudes (e.g., "Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences."). Items are scored on a Likert-scale from 1 (*strongly disagree*) to 5 (*strongly agree*), with possible scores ranging from 10-50 (for overall scores). Higher scores indicate greater sexism. Scoring information can be found in Appendix 5. The MS has been shown to be an acceptable measure of sexist attitudes, both in terms of its internal consistency (Cronbach's alpha = .75 in Swim et al., 1995 and .82 in Swim & Cohen, 1997) and ability to predict scores on other measures of gender bias (Campbell, Schellenberg & Senn, 1997).

Heteronormativity Attitudes and Beliefs Scale (HABS). The HABS is a 16-item questionnaire assessing heteronormative beliefs and assumptions. As mentioned in Chapter One, heteronormativity can be broadly defined as the belief that people

fall into one of two distinct gender categories (male and female), which form a natural heterosexual dyad. The HABS consists of two eight-item subscales assessing, first, gender-as-binary beliefs (e.g., "All people are either male or female") and, second, attitudes around natural or normative sexual behaviour, such as the assumption of heterosexuality in men and women (e.g., "There are particular ways that men should act and particular ways that women should act in relationships"). Items are scored on a seven-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*), allowing a scoring range for the entire scale of 16-112. Higher scores indicate more pronounced heteronormative beliefs. Scoring information can be found in Appendix 6. The HABS has been shown to have good internal consistency (Cronbach's alpha = .85 in Habarth, 2015) and to correlate positively with other measures of gender bias (Habarth, 2015).

IRAPs

Stimuli. Participants completed two gender binary IRAPs: the first contained positively valenced masculine and feminine traits, and the second contained negatively valenced traits. Stimuli categories (Table 3.1) were identical to the ones developed and used in the previous chapter.

Table 3.1.

Stimuli used in the IRAPs

Desirable Traits IRAP					
Label 1: Men	**Label 2: Women**	**Target 1: Masculine traits**	**Target 2: Feminine traits**	**Rule A**	**Rule B**
Men	Women	Witty Charismatic Competitive Decisive	Nurturing Gentle Affectionate Sensitive	Please respond as if men have more stereotypically masculine traits and women more	Please respond as if women have more stereotypically masculine traits and men more stereotypically feminine traits

				stereotypically feminine traits	

Undesirable Traits IRAP					
Label 1: Men	**Label 2: Women**	**Target 1: Masculine traits**	**Target 2: Feminine traits**	**Rule A**	**Rule B**
Men	Women	Aggressive Unemotional Insensitive Arrogant	Bitchy Insecure Bossy Helpless	Please respond as if men have more stereotypically masculine traits and women more stereotypically feminine traits	Please respond as if women have more stereotypically masculine traits and men more stereotypically feminine traits

Task structure. Prior to commencing the task, participants were provided with verbal instructions on how to complete the IRAP. These instructions broadly outlined the task structure (i.e., that they would be presented with blocks consisting of multiple word pairings and they would need to respond in accordance with a response "rule" presented before each block). The instructions also emphasized the importance of maintaining speed and accuracy throughout the task. Once participants were comfortable with these instructions, they began the "practice" phase of the IRAP, which was designed to train participants to a certain level of response fluency (78% accuracy and a median response latency of >2000ms; see Hussey et al., 2015). In keeping with convention (Hussey et al., 2015; Vahey et al., 2015), participants were presented with up to four pairs of practice blocks (i.e., four iterations of paired Rule A and Rule B blocks) until they reached the desired level of fluency, after which point they moved to the "test" portion of the IRAP. Those who did not meet the practice criteria did not complete the test blocks.

The practice and test phases of the IRAP were identical in terms of their stimuli and block structure. Both involved the presentation of a pre-block rule screen, 32 individual trials and a post-block feedback screen outlining the participant's accuracy and latency scores for that block. The pre-block rule screen presented Rule A or B (e.g., "Please respond as if men have more stereotypically masculine traits and women more stereotypically feminine traits") and outlined the task instructions:

This task will determine what makes "intuitive sense" to you by seeing what rules you find easy and hard to follow. You'll pair words or images according to a rule. You'll be told the rule and when it changes. If you make a mistake, you'll see a red "X." Provide the correct response to continue. Learn to respond accurately according to the rule. When you've learned to be accurate you'll naturally speed up too. Going quickly without being accurate will not provide meaningful data.

As is typical in an IRAP, two stimuli were presented together on the screen per trial (one label stimulus and one target stimulus, e.g., "women" and "nurturing"). The two response options (*true* and *false*) remained static across all trials at the bottom left and right of the screen. Each stimulus remained on the screen until the correct response was emitted. If participants responded incorrectly, corrective feedback in the form of a red "X" appeared in the centre of the screen. Each block pair consisted of one "Rule A" block and one "Rule B" block. In the "Rule A" block, response contingencies reinforced choices of the on-screen word *true* when men-masculine and women-feminine stimulus pairings were present and *false* for men-feminine and women-masculine word pairings. In the "Rule B" block, the inverse response options were reinforced. The order of the rule blocks was always counterbalanced between participants.

The category labels (masculine and feminine) were clearly specified to participants in the task. This was due to existing research suggesting category labels as well as features of the individual stimuli influence effects on implicit measures (Steffens & Plewe, 2001). Bluemke and Friese (2006) and Govan and Williams (2004), for instance, found that effects on an IAT were considerably decreased or even reversed when the category exemplars were inconsistent with the valence of the category labels. To ensure the categories were salient and coherent, the labels were therefore included and specified during the pre-block rule screen. It is important to note that this was not expected to confound effects (e.g., due to the conceptual overlap in society between sex and gender categories). Implicit measures are sensitive to category labels, but the individual exemplars still need to form a coherent class independent of the label (see De Houwer, 2001).

Hiring task.

Hiring preference was assessed using the same hiring task as Chapter Two.

Ethical Issues

Ethical issues and procedures were identical to previous studies. The content and procedures posed no particular ethical risk other than participant fatigue; to mitigate this, participants were advised they could take a break (between blocks or tasks) when needed and could cease participation at any time. As no personal data were collected and responses were anonymous, they were informed they could not withdraw after completing the study.

Data Processing and Analysis

Following routine practices, latency differentials across Rule A and Rule B blocks were quantified using the D_{IRAP} scoring algorithm, a scoring metric based on an adaptation of Cohen's *d*. As previously mentioned, IRAP scores are analysed at

the trial-type level so as to provide an assessment of effect size for each individual trial-type (i.e., men-masculine, men-feminine, women-masculine, women-feminine). Thus, four separate D_{IRAP} scores were produced for each IRAP trial type, all of which have a potential range of +2 to -2. In keeping with convention, the scores for the third and fourth trial types were inverted after calculation (see Hussey et al., 2015). Inverting trial-type scores does not change the effect produced in any way, it merely makes it slightly easier to interpret effects on a graph. After inverting, a positive D_{IRAP} score on any trial-type represents a "masculine" or "not-feminine" effect and a negative score a "feminine" or "not-masculine" effect. Overall D_{IRAP} scores were also calculated and used for any analyses involving multiple comparisons (e.g., correlations between IRAPs and other measures). This is commonly done in IRAP research as a way to minimise the risk of a Type 1 error (i.e., because of multiple comparisons on trial type scores). Overall scores are calculated by simply averaging the four un-inverted trial-type scores.

Practice-block data were not included in the analysis, and thus IRAP data were only collected from participants who progressed to the test phase. In this study, this was 45 participants for the positive traits IRAP and 44 for the negative traits IRAP. Using the exclusion method outlined in Nicholson and Barnes-Holmes (2012), IRAP data were removed for participants who failed to meet accuracy and/or latency criteria in more than one of the three test-block pairs. Three participants were removed from the negative traits IRAP on this basis. Participants were not excluded from the analysis if they failed to meet criteria in only one of their test-block pairs; however, the final D_{IRAP} scores for these individuals were calculated by averaging the D_{IRAP} scores across the remaining two (rather than three) pairs of test blocks. D_{IRAP} scores for three participants were calculated in this manner.

3.2.2 Results

Analyses were again conducted using JASP (version Beta 2).

Self-Report Measures

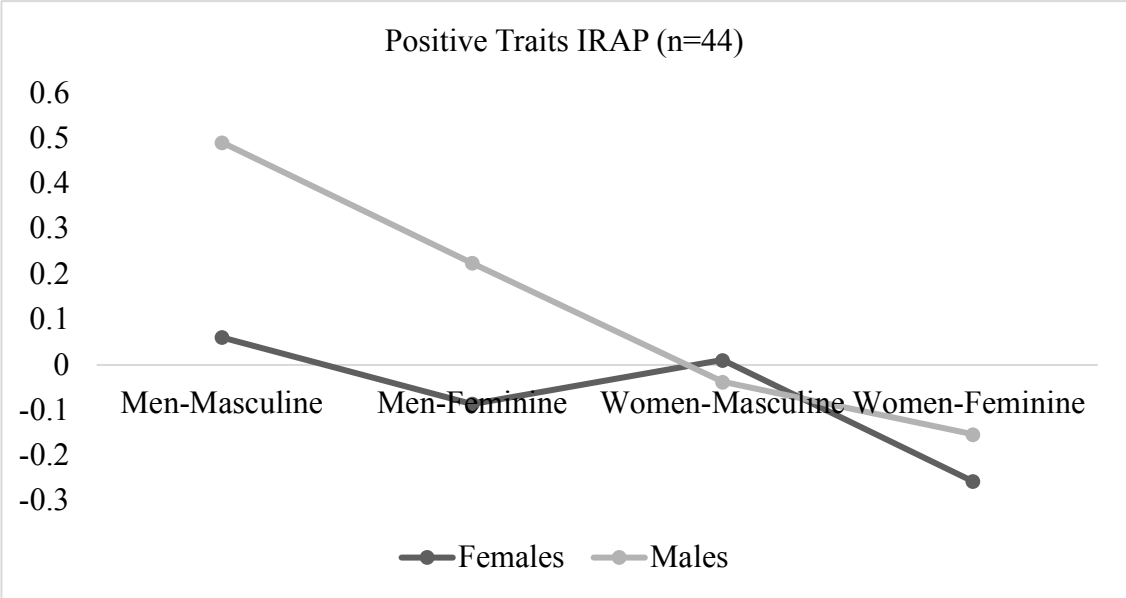
Scores on the Modern Sexism Scale (MSS) were higher in male participants than females (females: $M = 24.42$, $SD = 5.87$; males: $M = 28.9$, $SD = 7.43$). Heteronormativity was also higher in male participants (females: $M = 41.88$, $SD = 12.12$; males: $M = 50.4$, $SD = 18.11$). Independent samples t-tests identified significant gender differences on both the HABS, $t(46) = -1.9$, $p = .031$, and the MS, $t(46) = -2.26$, $p = .014$.. Effect sizes for both were large (Cohen's $d > .8$).

IRAP Performance

A visual inspection of the graph indicates that effects on all trial types reflected a binary-consistent pattern of responding (see Figure 3.1). That is, men were associated with masculine but not feminine traits, and women were associated with feminine but not masculine traits. Bias scores were typically larger for male participants, with the exception of the women-feminine trial-type. Larger resistance to forming role-incongruent relations was found for the men-feminine relative to the women-masculine trial type for both genders, with males demonstrating notably stronger "men-*not*-feminine" effects. Means and standard deviations are in Table 3.2.

Figure 3.1. Trial Type Effects for Male and Female Participants for the Positive and Negative IRAPs

Positive Traits IRAP



Negative Traits IRAP

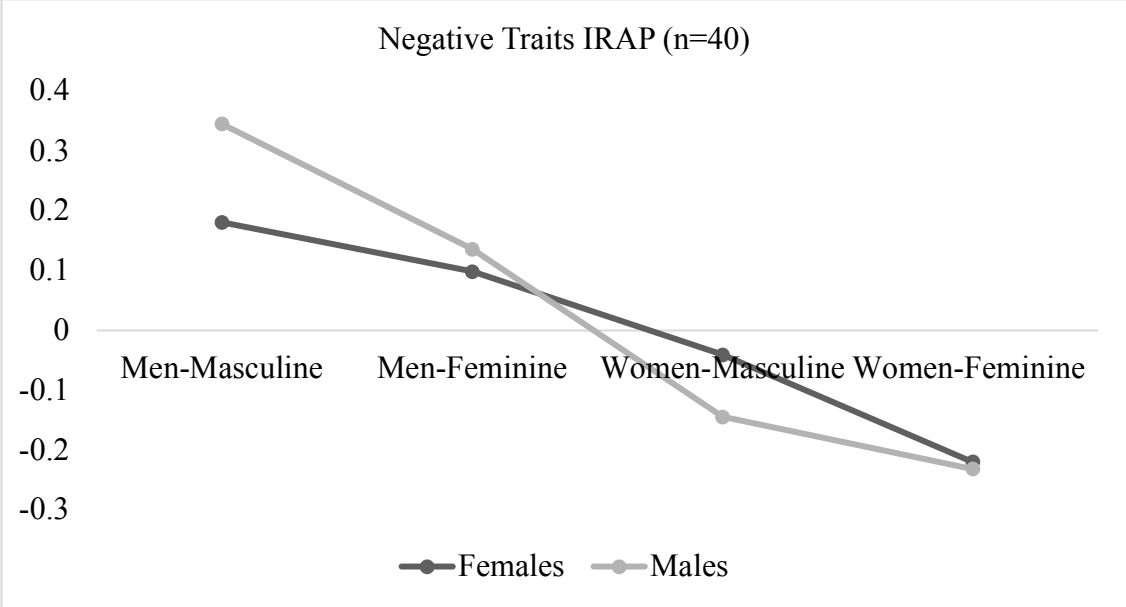


Figure 3.2. Trial-type level D_{IRAP} scores for the IRAPs. Note. D_{IRAP} scores have a possible range of -2 to +2. A positive score can be interpreted as a "masculine" or "not-feminine" effect, and a negative score a "feminine" or "not-masculine" effect.

Table 3.2

Trial-type means and standard deviations for the positive and negative IRAPs

IRAP	Trial-Type	Gender	Mean	SD	N
Positive IRAP	Men-masculine	Female	0.06	0.36	26
		Male	0.49	0.37	18
	Men-feminine	Female	-0.09	0.26	26
		Male	0.23	0.33	18
	Women-masculine	Female	0.01	0.41	26
		Male	-0.04	0.36	18
	Women-feminine	Female	-0.26	0.41	26
		Male	-0.15	0.39	18
Negative IRAP	Men-masculine	Female	0.18	0.28	23
		Male	0.35	0.35	17
	Men-feminine	Female	0.10	0.36	23
		Male	0.14	0.37	17
	Women-masculine	Female	-0.04	0.34	23
		Male	-0.15	0.31	17
	Women-feminine	Female	-0.22	0.36	23
		Male	-0.23	0.42	17

One-sample t-tests were run on the trial-type scores for the entire sample (i.e., male and female participants combined: Table 3.3). Analyses assumed a test value of 0. Scores were collapsed across genders for these analyses to minimize the number of comparisons, and a Bonferroni correction was applied to minimize the risk of Type I errors. Effects were broadly comparable across the two IRAPs, with significant men-masculine and women-feminine effects on both. The only notable difference is on the men-feminine trial-type; there is a stronger men-*not*-feminine effect on the negative but not the positive traits IRAP, though this was non-significant after a Bonferroni correction.

Table 3.3

Bonferroni-corrected one sample t-test results

IRAP	Trial-Type	t	df	p
Positive IRAP	Men-masculine	3.769	43	<.001***
	Men-feminine	0.828	43	0.412
	Women-masculine	-0.149	43	0.882
	Women-feminine	-3.568	43	<.001***
Negative IRAP	Men-masculine	4.974	39	<.001***
	Men-feminine	2.029	39	0.049*
	Women-masculine	-1.636	39	0.110
	Women-feminine	-3.736	39	<.001***

*Note:*** denotes significance after a Bonferroni-correction ($p < .00625$).

Gender Differences

A 2×2×4 repeated measures ANOVA was conducted to assess the impact of participant gender and IRAP type (i.e., positive or negative traits) on trial-type scores. Only participants with data from both IRAPs were included in this analysis ($n = 39$). While there was a significant two-way interaction between trial type and gender, $F(3, 35) = 6.82$, $p = .047$, $\eta_p^2 = .045$, no significant three-way interaction was found between trial-type, gender, and IRAP type, $F(3, 35) = 1.77$, $p = .820$, $\eta_p^2 = .008$. That is, although male and female participants performed significantly differently based on trial type within the IRAP, these effects were not related to the valence of the traits. There was a significant main effect for gender, with men showing more binary-consistent biases than women on both IRAPs, $F(1, 37) = 6.4$, $p = .016$, $\eta_p^2 = .148$. There was also a significant interaction effect between gender and IRAP type, $F(1, 37) = 6.8$, $p = .013$, $\eta_p^2 = .155$.

Follow-up Bonferroni-corrected independent-samples t-tests were conducted to explore the above two-way interaction effect between participant gender and IRAP trial-type (see Figure 3.2 for a graphical representation of trial-type scores). As performance on the IRAP did not differ according to trait valence, participant scores

for both IRAPs were collapsed (at the trial-type level) for these analyses. Thus, only participants with data for both IRAPs were included here ($n = 40$). Differences were found only on the men-masculine trial-type: male participants demonstrated more positive D_{IRAP} scores ($M = 0.50$) than females ($M = 0.05$), $t(37) = 3.71$, $p < 0.001$. Men also demonstrated stronger men-*not*-feminine effects than women, though this was not signifiant ($M_{women} = .018$; $M_{men} = .159$), $t(37) = 1.95$, $p < 0.059$. This suggests that, across both of the male trial-types, men demonstrated stronger gender binary-consistent biases than women on the trial-types pertaining to their own gender.

Hiring Preferences

When asked about their preference for a particular gender (i.e., male or female), responses in the current sample were varied: 11% selected the man, 44% selected the woman, and 45% selected "I prefer not to answer"). A chi-square goodness-of-fit test revealed this distribution to be significantly unequal, χ^2 (2, $n = 47$) = 11.40, $p = 0.003$. For the stereotypical feminine/masculine-preference item, however, participants again demonstrated an overwhelming preference for the masculine person (83% selected masculine traits, 13% selected feminine traits, and 4% selected "I prefer not to answer"). A chi-square goodness-of-fit test again revealed this to be a significantly unequal distribution χ^2 (2, $n = 47$) = 52.64, $p < 0.001$. Chi-square tests for independence revealed no significant differences between male and female responses for either item (all ps > .5).

Measure Comparisons

Pearson's r correlations were conducted to explore the direction and significance of the relationship between explicit binarist or anti-women attitudes and IRAP performance. Overall D_{IRAP} scores were used for these analyses as mentioned in the method section. For male and female participants, no significant correlations

were found between scores on the positive or negative IRAP and either the HABS or MS (*p*s > .15).

3.2.3 Summary

This study revealed significant binarist gender biases in a young adult sample. Across two IRAPs, participants demonstrated effects in the expected role-congruent direction (i.e., men are masculine *and not feminine*, women are feminine *and not masculine*). Gender differences were identified in IRAP performances, with males demonstrating larger response biases across both IRAPs. Follow-up tests revealed this to be driven predominantly by differential performance on the "men-masculine" trial-type, with males demonstrating significantly stronger effects on this trial-type. Hiring preferences were in the expected direction, with a significant proportion of participants preferring the masculine over the feminine candidate. As in Chapter Two, when asked to express a preference for a male over a female candidate, nearly half of the participants responded that they would prefer not to answer, while the majority of the remaining sample elected the female. No significant correlations were found between scores on either IRAP and self-reported attitudes towards women and gender. Significant gender differences were found on the two self-report scales, with males demonstrating larger sexism and heteronormativity scores than females.

3.3 Study Four

Study Three evidenced the utility of the IRAP in assessing gender binary biases. It also demonstrated the indirect cultural association between men, masculinity, and occupational success. Study Four will now expand on these findings by first refining some procedural properties of the IRAP, and second

modifying the hiring task to assess a broader range of occupational gender biases and stereotypes.

Changes to the IRAP

The previous study assessed gender binary biases by presenting word pairs on screen (e.g. "men" and "dominant") and having participants respond either *true* or *false*. One advantage of the IRAP, however, is that it allows for full statements to be included on screen (e.g., "Men *are* dominant"), and indeed some research suggests this may have certain procedural advantages. Two studies, for example, found that changing the presentation of stimuli on self-esteem IRAPs from pairs to statements (i.e.., from self/me-good to "I am good" or "I want to be good") resulted in better predictive and discriminant validity in the context of depression (Remue, De Houwer, Barnes-Holmes, Vanderhasselt & De Raedt, 2013; Remue, Hughes, De Houwer & De Raedt, 2014). It is possible that similar results would be found in this context, given the way gender stereotypes are believed to be structured. As mentioned, a distinction is often drawn in the literature between the descriptive and prescriptive components of gender norms; that is, between assumptions around how men and women *do* behave versus how they *should* behave (see Eagly & Karau, 2002; Prentice & Carranaza, 2002). This study will explore this distinction by including one IRAP for descriptive binary biases and another for prescriptive biases. It should be noted that Prentice and Carranaza (2002) and Ridgeway (2011) further suggest gender stereotypes have a significant *pro*scriptive component, in that they also set out clear rules for how men and women should *not* behave. However, this aspect will not be directly assessed here because it would require to participants to respond in a potentially confusing manner (e.g., responding 'false' to "Men should *not* be sensitive"), which may confound results (see Hussey et al., 2015).

Study Four will also remove the negatively valenced IRAP. Both positive and negative IRAPs were originally included in order to control for and measure the effects of trait desirability. However, an IRAP takes a relatively long time to complete (between 10-30 minutes); it is thus advisable to restrict the number of IRAPs per study to two, in order to keep the total length of the study to under one hour and therefore minimise participant fatigue and complex order effects. Furthermore, as responses did not differ considerably across the two IRAPs, it is likely unnecessary to use both in future studies. The positive over the negative IRAP was retained primarily because the hiring task now only uses the positive traits (the "firing" item using the negative trait was dropped after Chapter One) and thus allowed for a direct comparison. In addition, previous research on implicit gender bias suggested valence was particularly likely to influence results when the traits were negative, possibly because men and women may resist pairing negative traits with their own gender (see Rudman, Greenwald & McGhee, 2001).

Changes to the Hiring Task

The original task ascertained preferences between a feminine and a masculine person for a single gender-neutral occupation ("office job"). Responses across the two previous studies skewed very heavily in the pro-masculine/agentic direction (approx. 90% of 290 participants overall) and demonstrated that masculinity-agency is considerably more socially valuable than communality. While this is a significant and meaningful finding, it is not generally recommended to do group comparisons with data this skewed (see Delucchi & Bostrom, 2004). The current study will therefore expand on the task and look at gender preferences for a broader range of occupations: some neutral, some stereotypically feminine and some stereotypically masculine. In so doing, we can examine not only gender preference for a range of

occupations (and thus the degree of overall pro-feminine/masculine bias) but also participants' degree of occupational gender stereotyping, which may also be relevant.

Additional Changes

A small number of other procedural changes will also be made before Study Four. First, the age range moving forward will be restricted to 18-29 to ensure the samples could be meaningfully described as young adults. Second, the HABS measure will be removed from future studies. While this was a useful explicit measure of heteronormative and gender-as-binary beliefs, there was a lot of variance in responses and it also pushed the number of measures participants were asked to complete up to five. Some participants also verbally reported finding parts of the questionnaire confusing. The Modern Sexism Scale is a more stable and widely used measure of explicit gender beliefs so this will be retained for use in the remaining studies.

Current Study

The current study (Study Four) will assess descriptive and prescriptive gender biases and explore their relationship to explicit hiring preferences. Based on the results of the previous study, it is hypothesised that participants will more readily attribute masculine traits to men (but not women) and feminine traits to women (but not men). Given the lack of research to date on the descriptive/prescriptive distinction on implicit measures, no specific hypotheses are put forward in terms of the relationship between the two IRAPs, or between the IRAPs and the hiring task. For the hiring task, it is expected that participants will hire in a stereotype-consistent manner (i.e., hire the masculine individual for the masculinized jobs and the feminine person for the feminized jobs). However, it is also expected that

participants will be more likely to hire the agentic individual for the more gender-neutral positions, and thus that there will be an overall pro-masculine bias. As in the previous studies, it is expected that bias scores will be higher/more pronounced for male participants across all measures.

3.3.1 Method

Participants

Sixty-one White Irish undergraduate students aged between 18 and 23 participated in this study (M_{age}=20.7). Of the sample, 30 identified as female and 30 as male. Participation was voluntary and no remuneration was offered. Inclusion criteria included fluent English, normal or corrected-to-normal vision, and full use of both hands. Participants were provided with an open-ended response format for gender.

Procedure

Experimental sequence. The experimental sequence and context were identical to Study Three.

Materials

Modern Sexism Scale.

The version of the MS in this study was identical to Study Three.

IRAPs.

Two separate IRAPs were employed in this study: one for descriptive biases and another for prescriptive biases. Both contained the same stimulus categories but differed in the relational terms used (see Table 3.4).

Task structure. The format and procedure of the IRAPs were identical to Study Three.

Table 3.4

Stimuli used in the descriptive and prescriptive IRAPs.

Descriptive IRAP					
Label 1: Men	Label 2: Women	Target 1: Masculine traits	Target 2: Feminine traits	Rule A	Rule B
Men are	Women are	Witty Charismatic Competitive Decisive	Nurturing Gentle Affectionate Sensitive	Please respond as if men have more stereotypically masculine traits and women more stereotypically feminine traits	Please respond as if women have more stereotypically masculine traits and men more stereotypically feminine traits
Prescriptive IRAP					
Label 1: Men	Label 2: Women	Target 1: Masculine traits	Target 2: Feminine traits	Rule A	Rule B
Men should be	Women should be	Witty Charismatic Competitive Decisive	Nurturing Gentle Affectionate Sensitive	Please respond as if men should have more stereotypically masculine traits and women should have stereotypically feminine traits	Please respond as if women should have more stereotypically masculine traits and men should have stereotypically feminine traits

Hiring task. Hiring preference was assessed using a modified version of the task from Chapter Two and Study Three of this Chapter. Similar to previous studies of occupational gender preference (e.g., Rudman & Glick, 2001; Rudman & Kilianski, 2000), participants were required to choose candidates for a range of "feminized", "masculinized" and gender-neutral occupations (see Table 3.5). These occupations were extracted from a series of studies into occupational gender stereotypes (Beggs & Doolittle, 1993; Shinar, 1975; Elsaid & Elsaid, 2012). The task instructions were identical to the previous task, as were the traits ascribed to "Person A" and "Person B" (i.e., the desirable masculine-agentic and feminine-communal traits employed in the IRAP). The occupations were presented in a fixed randomized order.

Table 3.5

Occupations included in the hiring task.

Feminine Occupations	**Masculine Occupations**	**Gender-Neutral Occupations**
Secretary	Electrician	Salesperson
Nurse	Computer scientist	Journalist
Psychotherapist	Financial advisor	Pharmacist
Hairdresser	Police officer	
Primary school teacher	Lawyer	

Ethical Issues

Ethical issues, approval, and procedures were identical to Study Three.

Data Processing and Analysis

IRAP

Data were processed in an identical manner to the previous study and participants were excluded using the same criteria. One participant was removed

from the descriptive IRAP (final $n = 60$) and three from the prescriptive on this basis (final $n = 57$).

Hiring Task

The hiring task was processed to produce separate gender preference and gender stereotype scores. That is, to produce indices of (i) participants' overall preference for hiring the masculine relative to the feminine person, and (ii) participants' tendency to hire the masculine individual for the stereotypically male positions and feminine for the stereotypically female positions. The gender preference score was calculated by recoding a masculine preference as +1 and a feminine preference as 0 and then totalling the responses. Possible scores range from 0 to 13, with higher scores reflective of an overall pro-masculine bias.

The gender stereotype score was calculated by recoding responses as either stereotypical or non-stereotypical, depending on whether the occupation was stereotypically male or female. Stereotypical responses (i.e., selecting the feminine person for the stereotypically feminine occupation) were coded as +1 and non-stereotypical responses were coded as -1. The score was calculated from responses on the 5 masculine and 5 feminine occupations combined, so possible scores ranged from -10 (strongly non-stereotypical responses) to +10 (strongly stereotypical responses).

3.3.2 Results

Self-Report Measures

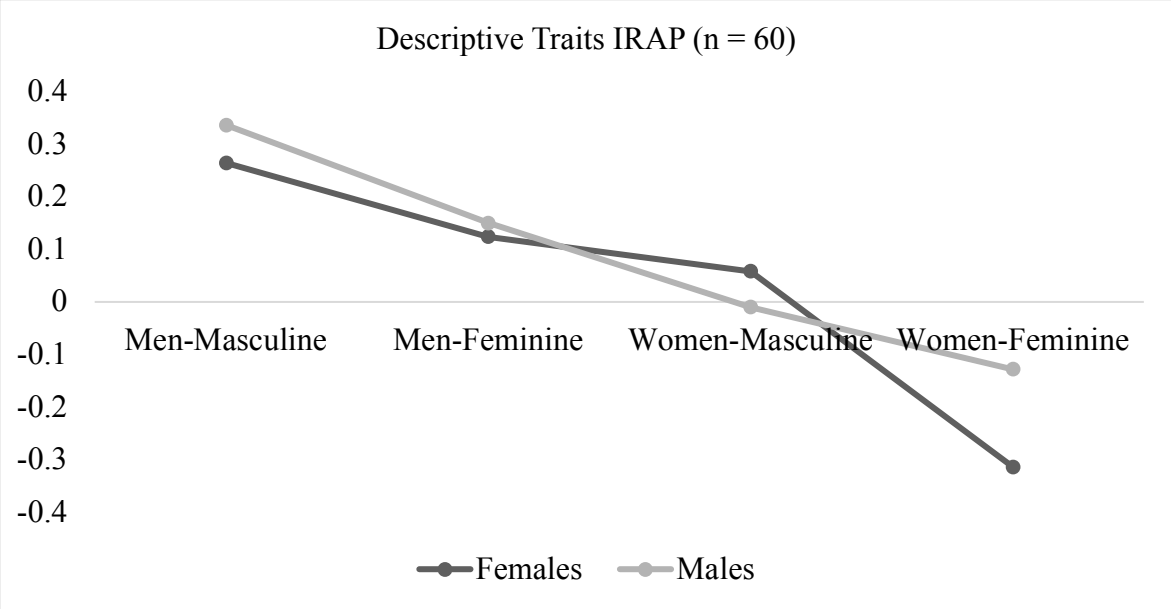
MS scores were lower again in the current sample (females: $M = 19.41$, $SD = 5.35$; males: $M = 19.67$, $SD = 4.75$), potentially due to the restricted age range. An independent samples t-tests identified no significant gender differences, $t(58) = -0.393$, $p = .696$.

IRAP Performance

Effects across both IRAPs reflected a binary-consistent pattern of responding (see Figure 3.2 and Table 3.6). That is, men are/should be masculine and not-feminine, and women are/should be feminine and not-masculine. An inspection of the data would suggest no notable differences in performance between the genders or across the two IRAPs (i.e., prescriptive or descriptive), however, a 2x2x4 repeated measures ANOVA was conducted to explore this in more detail. Again, only participants with data from both IRAPs were included in this analysis (n = 58). Analyses revealed no significant main effect for either gender, $F(1, 55) = 1.89$, p = .179, or IRAP type, $F(1, 55) = .781$, and no significant two- or three-way interaction effects (all ps > .15).

Figure 3.2 Trial Type Effects for Male and Female Participants for the Descriptive and Prescriptive IRAPs

Descriptive IRAP



Prescriptive IRAP

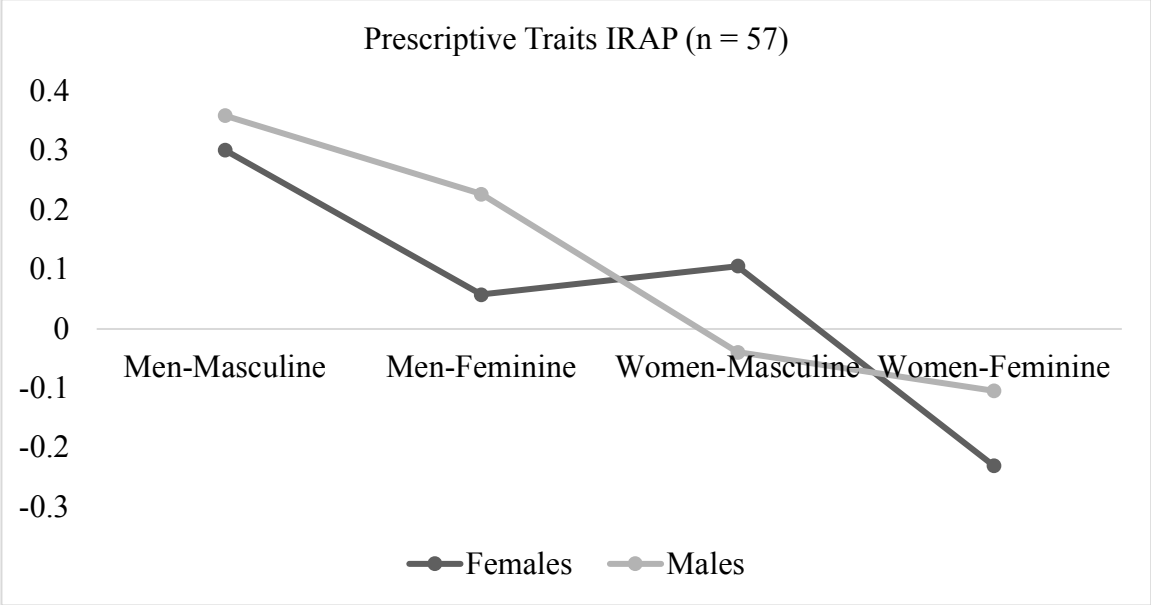


Figure 3.2. Trial-type level D_{IRAP} scores for the Descriptive and Prescriptive IRAPs. Note. D_{IRAP} scores have a possible range of -2 to +2. A positive score can be interpreted as a "masculine" or "not-feminine" effect, and a negative score a "feminine" or "not-masculine" effect.

Table 3.6

Trial-type means and standard deviations for both IRAPs

IRAP	Trial-Type	Gender	Mean	SD	N
Descriptive IRAP	Men-masculine	Female	0.26	0.36	31
		Male	0.34	0.43	29
	Men-feminine	Female	0.12	0.39	31
		Male	0.15	0.39	29
	Women-masculine	Female	0.06	0.26	31
		Male	-0.01	0.35	29
	Women-feminine	Female	-0.31	0.34	31
		Male	-0.13	0.36	29
Prescriptive IRAP	Men-masculine	Female	0.30	0.34	29
		Male	0.36	0.38	29
	Men-feminine	Female	0.06	0.33	29
		Male	0.23	0.42	29
	Women-masculine	Female	0.11	0.31	29
		Male	-0.04	0.38	29
	Women-feminine	Female	-0.23	0.35	29
		Male	-0.10	0.43	29

Though the ANOVA found no difference in responding based on gender or IRAP type, one-sample t-tests were run to assess the significance of the effects at the trial-type level for the sample overall, again assuming a test value of 0. As in Study Three, significant men-masculine and women-feminine effects were found across both IRAPs (see Table 3.7). It is interesting to again note the near-significant men-*not*-feminine effects, suggesting resistance to ascribing feminine traits to men (i.e., participants were faster to select the response option "false" rather than "true"), in addition to the absence of comparable effects on the women-*not*-masculine trial-types.

Table 3.7

Bonferroni-corrected one sample t-test results

IRAP	**Trial-Type**	**t**	**df**	**p**
Descriptive IRAP	Men-masculine	5.861	59	< .001
	Men-feminine	2.725	59	0.008
	Women-masculine	0.637	59	0.527
	Women-feminine	-4.787	59	< .001
Prescriptive IRAP	Men-masculine	7.006	56	< .001
	Men-feminine	2.793	56	0.007
	Women-masculine	0.751	56	0.456
	Women-feminine	-3.253	56	0.002

*Note:**** denotes significance after a Bonferroni-correction ($p < .00625$).

Hiring Preferences

Occupational gender preferences for female and male participants can be found in Figure 3.3. A visual inspection of the data surprisingly indicates that female participants had a stronger overall masculine preference than the males, in addition to more noticeable stereotyping (i.e., responses are more skewed for both the pro-feminine and pro-masculine positions). Independent samples t-tests revealed these differences to be significant (Table 3.8).

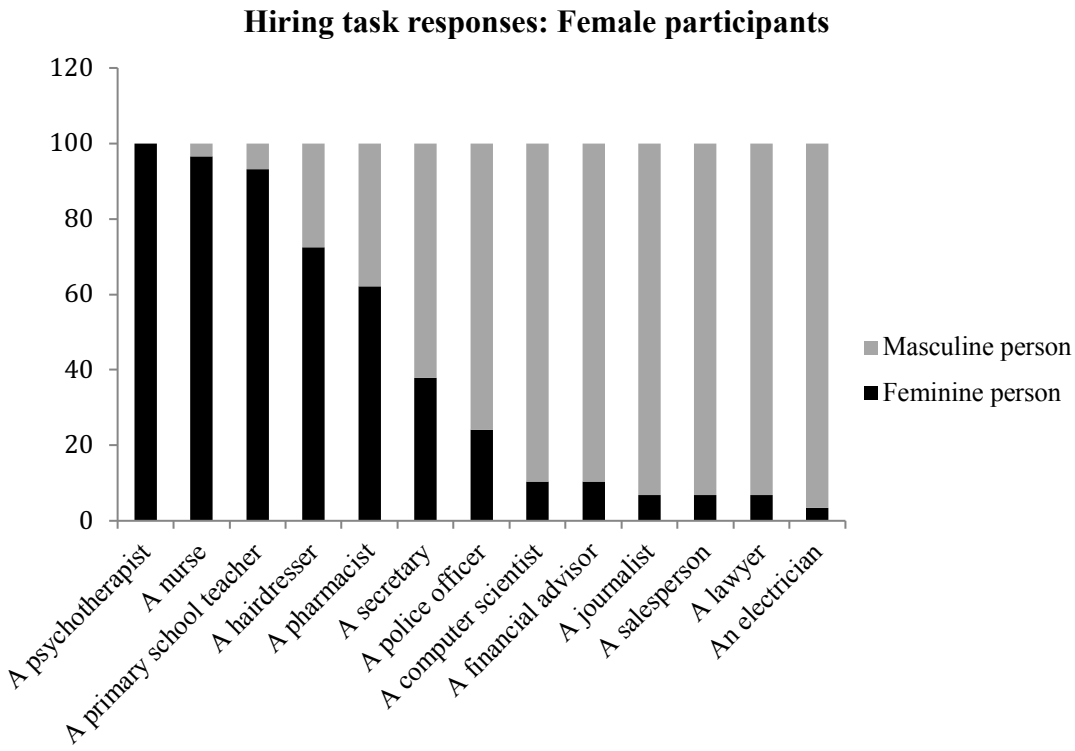
Table 3.8

Descriptive statistics and independent-samples t-test results for the hiring task

	Females Mean (SD)	**Males** Mean (SD)	***t***	**df**	***p***
Gender Preference	9.7 (8.5)	7.9 (1.5)	5.46	57	<.001***
Gender Stereotyping	6.9 (2.7)	5 (3.6)	2.29	57	.026*

Note. Gender preference scores range from 0-13 with higher scores indicating greater pro-male bias. Gender stereotype scores range from -10 to +10 with negative scores representing non-stereotypical responding and positive scores representing stereotypical responding.
* $p < .05$ *** $p < .001$

Figure 3.3. Hiring Task Responses



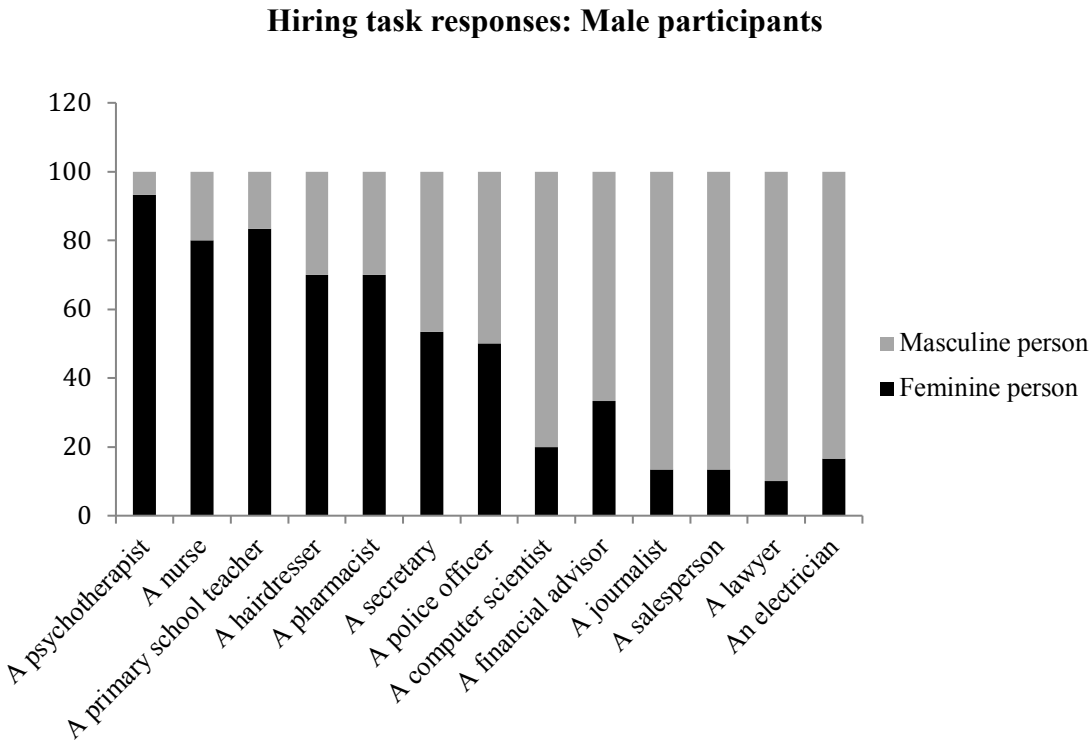


Figure 3.3. Hiring preferences per occupation for male and female participants..
Note. Histograms were generated to examine levels of preference and stereotyping for the entire sample. In terms of overall gender preference, there was a strong pro-masculine trend with nearly all preference scores falling over the halfway 6.5 mark. Stereotype scores were also skewed to the right, suggesting a general tendency toward more stereotype-consistent than inconsistent responding.

Figure 3.4 Distribution Plots for Hiring Task Responses

Figure 3.4. Frequency distributions for overall masculine and feminine preference and for levels of occupational stereotyping. Note: Gender preference scores range from 0-13 with higher scores indicating greater pro-male bias. Gender stereotype scores range from -10 to +10 with negative scores representing non-stereotypical responding and positive scores representing stereotypical responding.

In terms of other notable trends, it is interesting that the three "neutral" occupations – salesperson, journalist, and pharmacist – were not evaluated by the sample as gender-neutral. Approximately 90-95% of the sample had a masculine-agentic preference for both the salesperson and the journalist, while 60-65% had a feminine-communal preference for the pharmacist.

Measure Comparisons

Pearson's *r* correlations were conducted to examine the relationships between IRAP performance, hiring preferences, and self-reported sexism (Table 3.8). There were significant correlations between the two hiring task scores and between the two IRAP scores, as would be expected. There was also a significant correlation between scores on the descriptive IRAP and the gender preference score. No other significant correlations were found.

Table 3.8

Measure comparisons.

		Descriptive IRAP	**Prescriptive IRAP**	**Stereotype score**	**Preference score**	**MS**
Descriptive IRAP	Pearson's *r*	—	0.376***	0.167	0.321***	0.073
	p		0.002	0.104	0.007	0.295
Prescriptive IRAP	Pearson's *r*		—	0.029	0.262	0.132
	p			0.583	0.974	0.169
Stereotype score	Pearson's *r*			—	0.238*	0.004
	p				0.035	0.489
Preference score	Pearson's *r*				—	0.046
	p					0.366

MS	Pearson's *r* *p*	—

3.3.3 Results Summary

Similar to Study Three, significant binarist biases were found across both IRAPs. While effects were stronger for male participants, there was no significant main effect for gender in the current study. Response patters across both IRAPs did not differ significantly to one another, and the same asymmetrical response pattern on the role-incongruent trial-types was observed on both as in the previous study (i.e., a significant men-*not*-feminine effect but no comparable women-*not*-masculine effect). The absence of differences across the descriptive and prescriptive IRAPs could be interpreted in two ways: first, as is consistent with stereotype theory (Prentice & Carranaza, 2002), that gender stereotypes are simultaneously prescriptive and descriptive, or second, that the IRAP is not as sensitive as hoped to relational terms such as "should" and "are", as has been suggested in other domains (Remue et al., 2013; 2014). The paucity of research in this area makes it difficult to contextualise these findings, but the absence of any difference is interesting nonetheless.

While participant responses were not significantly different across the two IRAPs, they did differ in terms of their relationship to hiring preferences. A relationship was found between the descriptive IRAP and the overall gender-preference score only. No relationship was found between gender preference and effects on the prescriptive IRAP, and neither measure predicted the gender stereotyping score. Results are thus consistent with mainstream feminist theories of binary gender order (e.g., Butler, 2002; Connell, 2005). However, the absence of a relationship between agency/communion biases and occupational stereotyping

contradicts previous questionnaire research (Eagly & Carli, 2007) and studies using implicit measures (Rudman & Glick, 2001; Rudman & Kilianski, 2000).

For the hiring task, effects were generally in the expected direction: participants had an overall tendency towards hiring the masculine-agentic person and also to stereotype as expected for the masculinized and feminized roles. It is worth noting at this stage that gender stereotype and pro-masculine preferences were significantly higher for female relative to male participants. One explanation for this unexpected finding is that the women in the current sample are more sexist or gender-conservative than the previous study; however, their Modern Sexism Scale scores were not notably higher relative to the males or indeed the female participants in Study Three. A more likely explanation is that women do not consider the masculine-agentic traits to be uniquely masculine, as indicated by the explicit ratings in Chapter Two and supported by the absence of women-*not*-masculine biases on the IRAP. This will require expansion and replication in Study Five and future studies.

3.4 Study Five

The previous two studies demonstrated the utility of the IRAP in measuring binarist gender biases. They also indicated a relationship between these biases and explicit hiring preferences. The current study will now assess whether the IRAP is a useful methodology for assessing gender evaluations, and specifically evaluations of competency. As discussed in the introduction, there is a well-established overlap between masculine agency and occupational competency or success (see Eagly & Carli, 2008). The previous studies supported this assumption, revealing that (a) agentic traits are both more readily ascribed to men than women, and (b) more valued in an occupational context than communal traits. It is therefore possible that traits related to competency or occupational success will be gendered on an IRAP.

The current study will thus include two IRAPs: one assessing descriptive gender binary biases (as in the previous studies), and the other associations between gender roles (i.e., the masculine and feminine stimulus categories) and either competent or incompetent traits.

For the gender binary IRAP, we expect a repeat of the same binary-consistent response pattern. No specific pattern of results is expected for the gender-competency IRAP, as this will be the first study to examine competency evaluations of gender roles (as opposed to men and women). However, based on previous research showing the explicit overlap between masculinity and competency stereotypes (Koenig et al., 2011) and the implicit associations between competency and maleness (Rudman & Kilianski, 2000), it is generally expected that competent traits will be more readily attributed to masculinity over femininity. The same hiring task as in Study Four will be used here, so we again anticipate a relationship between binarist biases and occupational preferences. No specific hypothesis is proposed regarding the relationship between the gender-competency evaluations and hiring preferences however, as it is the first time it is being used. As gender differences in IRAP and hiring preference scores have varied across the previous two studies, no directional hypotheses are proposed here.

Given that this study includes an additional IRAP for competency evaluation (and the maximum number of IRAPs it is feasible to include is two), the prescriptive stereotype IRAP will be dropped from this study. This was done because it was not sufficiently different to the descriptive/gender binary IRAP and also because it was not associated with hiring preferences.

3.4.1 Method

Participants

Forty White Irish undergraduate students aged between 18 and 23 participated in this study (M_{age}=21.9). Of the sample, 20 identified as female and 20 as male.

Procedure

Experimental sequence. The experimental sequence including instructions and debriefing information were identical to previous studies.

Materials

Modern Sexism Scale.

The version of the MS in this study was identical to previous.

IRAPs.

Two separate IRAPs were employed in this study: one for binarist gender biases and one for gender-competency evaluations. Stimuli were the same as previous, though competent and incompetent traits were pulled from Rudman and Kilianksi (2000) for the competency IRAP (see Table 3.9).

Task structure. The format and procedure of the IRAPs were identical to previous, with one exception. Because we wanted to stimulus categories to be salient for both IRAPs, we elected to provide participants with a reference sheet outlining the "masculine" and "feminine" stimulus categories (Appendix 7). Typical IRAPs use relatively simple label categories (e.g., "men", "women", "White faces", "Black faces", etc.) that pair with more complex target categories (e.g. positive and negative words: see Hussey et al. 2015 for a description of a typical IRAP). The reference sheet was used in this study as the masculine and feminine categories might not have formed strong categories outside of the context of the study, and may therefore have

made the practice phase of the IRAP more difficult than normal. They were asked to refrain from commencing the study until they were comfortable with the two categories.

Table 3.9.

Stimuli used in the gender binary and gender competency IRAPs

Gender binary IRAP					
Label 1: Men	Label 2: Women	Target 1: Masculine traits	Target 2: Feminine traits	Rule A	Rule B
Men are	Women are	Witty Charismatic Competitive Decisive	Nurturing Gentle Affectionate Sensitive	Please respond as if men have more stereotypically masculine traits and women more stereotypically feminine traits	Please respond as if women have more stereotypically masculine traits and men more stereotypically feminine traits
Gender-competency IRAP					
Label 1: Masculine traits	Label 2: Feminine traits	Target 1: Competent words	Target 2: Incompetent words	Rule A	Rule B
Witty Charismatic Competitive Decisive	Nurturing Gentle Affectionate Sensitive	Competent Efficient Capable Employable	Incompetent Inefficient Incapable Unemployable	Please respond as if masculine traits are competent and feminine traits are incompetent	Please respond as if feminine traits are competent and masculine traits are incompetent

Hiring task. The hiring task was identical to the task used in Study Four.

Ethical Issues

Ethical issues, approval, and procedures were identical to previous.

Data processing and analysis

Data were processed in an identical manner to studies one and two and participants were excluded using the same criteria. Three participants were excluded from the binary IRAP (final $n = 37$) and five from the competency IRAP (final $n = 35$) on this basis.

3.4.2 Results

Self-Report Measures

Sexism scores were again low (females: $M = 18.82$, $SD = 3.9$; males: $M = 20$, $SD = 5.59$). An independent samples t-test found no significant gender differences, $t(38) = 0.453$.

IRAP Performance

Gender Binary IRAP. A visual inspection of the graph indicates a binary-consistent pattern of responding across all four trial-types (see Figure 3.5). That is, as in previous studies, men are masculine and not-feminine and women are feminine and not-masculine (Table 3.10 and Table 3.11). A 2x4 repeated measures ANOVA run on the eligible sample ($n = 37$) found a significant main effect for gender, $F(1, 35) = 7.58$, $p < .001$ with effects typically larger for males relative to females, but no significant interaction effect between trial-type and gender $F(3, 105) = .33$. In other words, although effects on the IRAP were moderated by participant gender, these differences were not driven by performance on one or more specific trial-types.

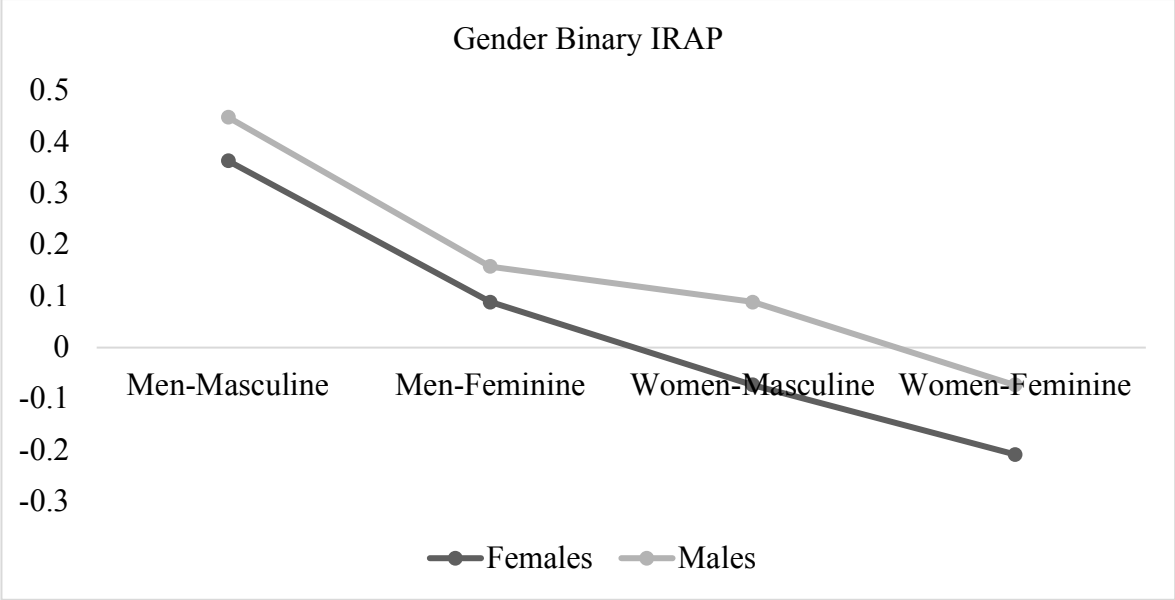


Figure 3.5. Note: positive scores reflect a "masculine" or "not-feminine" response pattern and negative a "feminine" or "not-masculine" response pattern.

Gender-Competency IRAP. Effects on the gender-competency IRAP were not all in the expected direction (Table 3.10 and Table 3.11). Significant masculine-competent and masculine-*not*-incompetent effects were found as anticipated; however, there was also a significant feminine-competent effect for both male or female participants, and significant feminine-*not*-competent effects were found. Though effects were slightly larger for the trial-types corresponding to the participant's own gender (i.e., effects on the feminine trial-types were larger for females than males and vice versa), a 2x4 repeated measures ANOVA on the eligible sample (n = 35) found no significant gender differences, $F(1, 33) = 1.02$, p =.321 and no significant trial-type–gender interaction effect, $F(3, 99) = 1.03$.

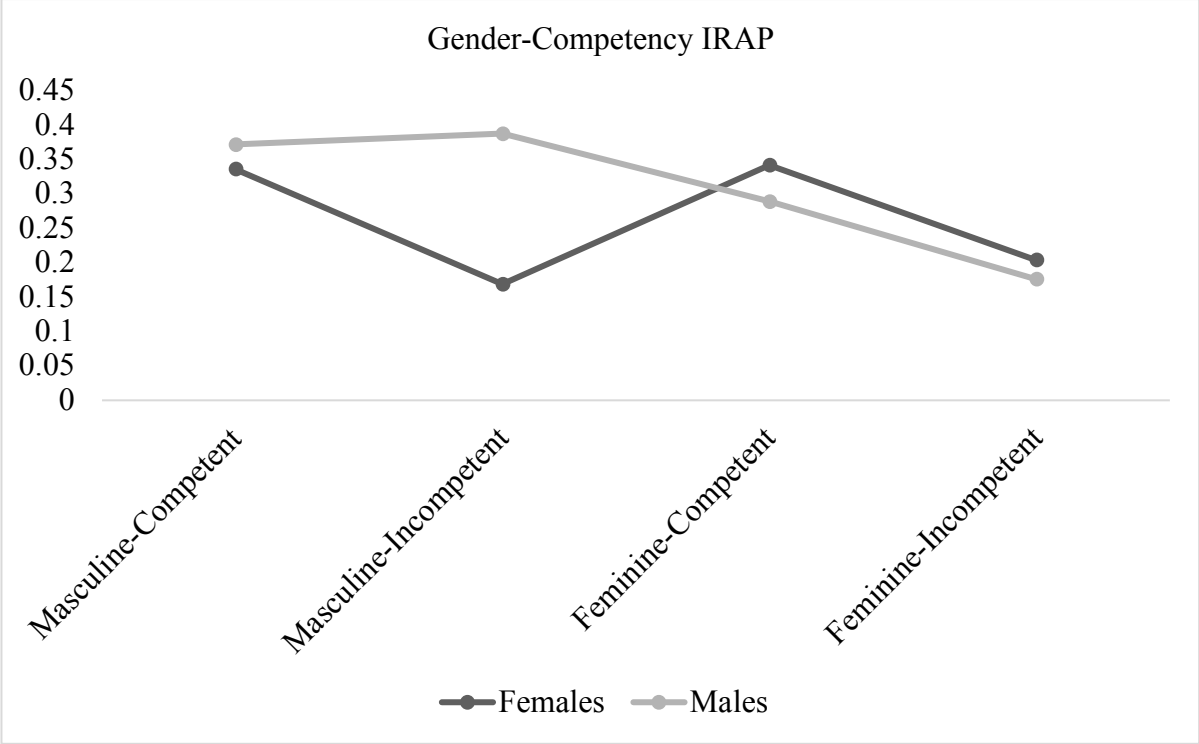


Figure 3.5. Note: positive scores reflect a "competent" or "not-incompetent" response pattern and negative an "incompetent" or "not-competent" response pattern.

Table 3.10

Means and Standard Deviations for the Gender Binary and Gender-Competency IRAPs

Table 3.6

Means and standard deviations for the gender Binary and Competency IRAPs

IRAP	Trial-Type	Gender	Mean	SD	N
Gender Binary IRAP	Men-masculine	Female	0.36	0.32	20
		Male	0.45	0.35	17
	Men-feminine	Female	0.09	0.29	20
		Male	0.16	0.36	17
	Women-masculine	Female	-0.07	0.34	20
		Male	0.06	0.29	17
	Women-feminine	Female	-0.21	0.30	20
		Male	0.00	0.25	17
Gender-Competency IRAP	Masculine-Competent	Female	0.34	0.34	19
		Male	0.37	0.44	16
	Masculine-Incompetent	Female	0.17	0.26	19
		Male	0.39	0.18	16
	Feminine-Competent	Female	0.34	0.31	19

		Male	0.29	0.35	16
	Feminine-Incompetent	Female	0.20	0.22	19
		Male	0.18	0.49	16

Table 3.11

Bonferroni-corrected one sample t-test results.

IRAP	**Trial Type**	***t***	**df**	***p***
Gender Binary	Men-Masculine	7.472	36	< .001***
	Men-Feminine	2.299	36	0.027
	Women-Masculine	-0.233	36	0.817
	Women-Feminine	-2.381	36	0.023
Gender-Competency	Masculine-Competent	5.501	34	< .001***
	Masculine-Incompetent	6.411	34	< .001***
	Feminine-Competent	5.737	34	< .001***
	Feminine-Incompetent	3.088	34	0.004

Note. * $p < .05$. **$p < .01$*** $p < .001$. Acceptable *p*-values after Bonferroni correction </= .00625.

Hiring Preferences

Hiring preferences for female and male participants can be found in Figure 3.6. A visual inspection of the data reveals no notable gender differences in terms of preferences for individual jobs or the overall level of occupational stereotyping. Independent samples t-tests revealed no significant differences between males and females on either score (see Table 3.12).

Table 3.12

Descriptive statistics and independent-samples t-test results for the hiring task

	Females	Males	*t*	df	*p*
Gender preference	*M* = 7.4 *SD* = 1.5	*M* = 7.2 *SD* = 1.4	-.62	38	.541
Gender stereotyping	*M* = 5.7 *SD* = 2.5	*M* = 6.4 *SD* = 4.1	-.49	38	.631

Note. * $p < .05$. **$p < .01$*** $p < .001$. Acceptable *p*-values after Bonferroni correction </= .01. Gender preference scores range from 0-13 with higher scores indicating greater pro-male bias. Gender stereotype scores range from -10 to +10

with negative scores representing non-stereotypical responding and positive scores representing stereotypical responding.

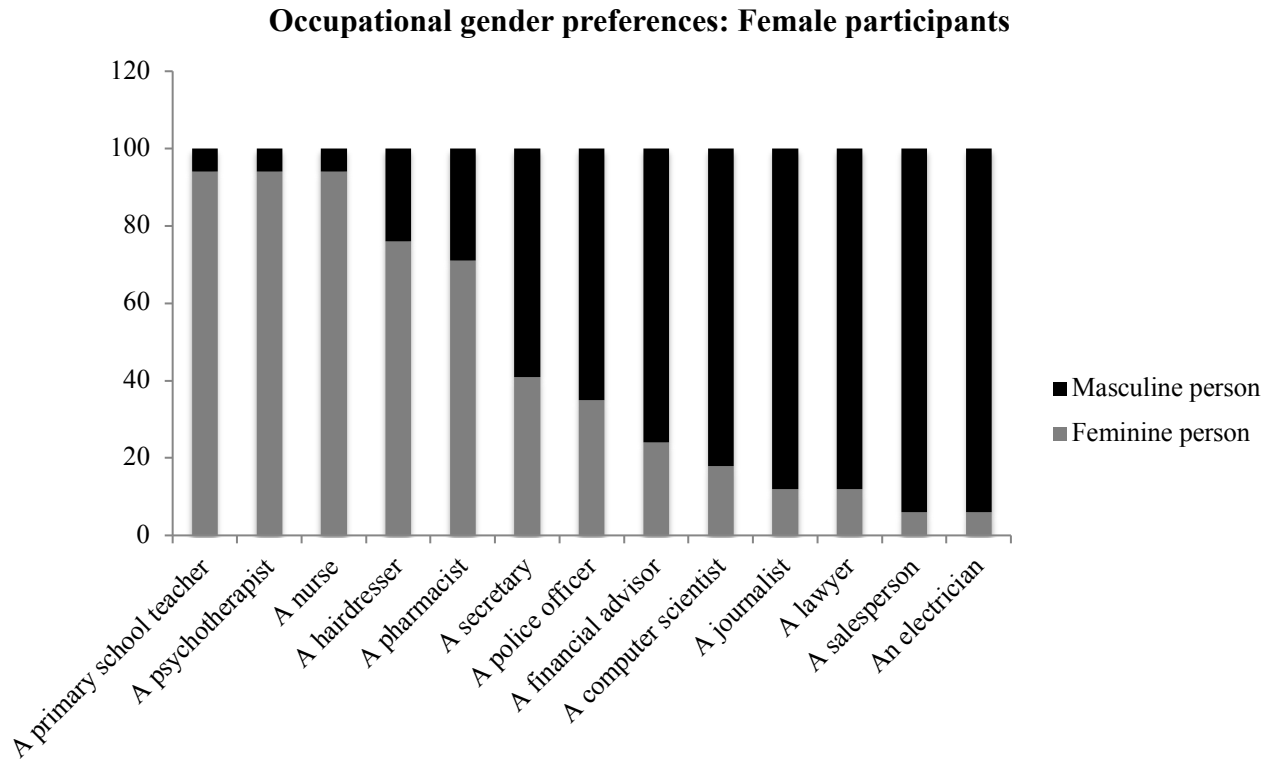


Figure3.6. Masculine and feminine occupational preferences (expressed in percent) for female participants

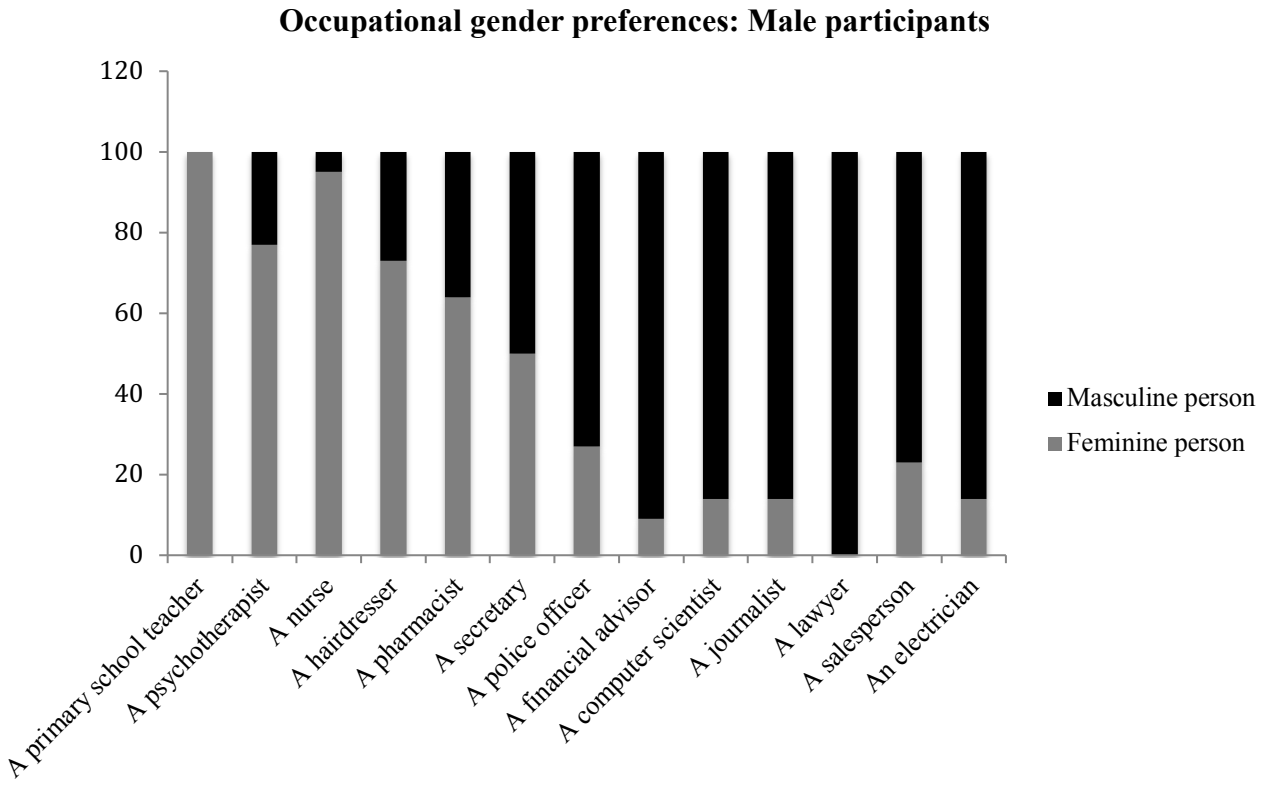


Figure 3.6. Masculine and feminine occupational preferences (expressed in percent) for male participants.

Histograms were again generated to examine levels of preference and stereotyping for the entire sample (Figure 3.7). There was a slight pro-masculine bias in the sample overall, with a larger proportion of preference scores falling over the halfway 6.5 mark. Stereotype scores were also skewed to the right, suggesting a general tendency toward more stereotype-consistent than inconsistent responding.

Figure 3.7. Frequency distributions for overall masculine and feminine preference and for levels of occupational stereotyping. *Note*: Scores range from 0-13 for the preference score (higher scores reflect pro-masculine bias) and -10 to +10 for the stereotype score (positive scores reflect stereotypical responses and negative scores non-stereotypical responses). Histograms were not split by gender given the relatively small sample size and that independent-samples t-tests identified no significant differences between male and female participants.

Measure Comparisons

Bonferroni-corrected Pearson's correlations were conducted to examine potential relationships between IRAP scores, self-reported sexism, and hiring preferences (Table 3.12). Significant positive correlations were found between scores on the gender binary IRAP and the overall hiring preference score, and between the two hiring task scores (though this is perhaps unsurprising). No other significant correlations were found but all relationships were in the expected positive direction.

Table 3.12

Measure Comparisons

		Gender binary IRAP	**Gender-competency IRAP**	**Stereotype Score**	**Preference Score**	**MS**
Gender binary IRAP	Pearson's r	—	0.220	0.160	0.414***	0.095
	p-value	—	0.106	0.175	0.006	0.288

Gender-competency IRAP	Pearson's r	—	0.128	0.082	0.193
	p-value	—	0.235	0.323	0.133
Stereotype Score	Pearson's r		—	0.421***	0.163
	p-value		—	0.004	0.160
Preference Score	Pearson's r			—	0.199
	p-value			—	0.112
MS	Pearson's r				—
	p-value				—

Note. All tests one-tailed, for positive correlation. * $p < .05$, ** $p < .01$, *** $p < .001$, one tailed. Acceptable *p*-values after Bonferroni correction </= .01

3.4.3 Results Summary

Responses on the gender binary IRAP were broadly consistent with previous studies. There were significant gender differences (as in Study Three but not Study Four), with bias scores higher for males relative to females, but no significant gender–trial-type interaction effect. For the gender-competency IRAP, scores were in the expected direction on the masculine trial-types, with significant masculine-competent and masculine-*not*-incompetent effects found for the overall sample. Contrary to what was expected, however, the data also showed a significant feminine-competent effect and no significant feminine-incompetent effect. Responses also did not differ by participant gender. While these data were somewhat surprising, the absence of anti-feminine bias on this IRAP does cohere with the hiring task responses in Chapter Two. One of the items in the earlier version of the task measured the extent to which participants would rather fire a stereotypically masculine or feminine person. Responses were equally split across the two options (i.e., there was no particular anti-feminine bias); this suggests, as proposed by the Role Congruity Hypothesis, that discrimination occurs because of the implicit men-masculinity-agency overlap rather than anti-feminine beliefs or beliefs about

women's incompetence. This will be explored further in the Chapter discussion below.

A significant relationship was found between the binary IRAP and an overall pro-masculine preference, suggesting again that binary biases play a role in workplace discrimination. No significant correlation was found between the two IRAPs in this study. These effects therefore do not cohere with the Rudman and Kilianski (2000) study, which found a correlation between communality-agency biases and gender-competency evaluations on an IAT. Again, this will be expanded on below in the discussion.

3.5 Discussion

This chapter had three aims: first, to measure and quantify gender binary biases using the IRAP; second, to examine the relative extent to which masculine and feminine traits are considered hireable; and third, to explore the relationship between IRAP scores and occupational hiring preferences. Across three studies, gender binary IRAPs were completed by a total of 147 young Irish adults. The relative impact of trait valence (i.e. positive or negative traits) was assessed in Study Three, and prescriptive/descriptive relational terms (i.e., *are* relative to *should be*) in Study Four. Study Five included an additional IRAP to examine gender-competency evaluations, or the extent to which competency and incompetency were differentially attributed to masculine or feminine traits. Gender preferences were assessed using a simulated hiring task that required participants to hire either a stereotypically masculine or feminine individual. Study Three used the single-item preference task developed and used in Chapter 2, and Studies Four and Five used a modified version that assessed preferences for a range of occupations. Gender differences were assessed throughout the Chapter, as were explicit or self-reported sexist beliefs.

IRAP Effects

In the current Chapter, effects on all binary IRAPs were consistent with a binarist gender ideology. That is, participants readily (i.e., speedily) coordinated men *but not women* with stereotypically masculine traits, and women *but not men* with stereotypically feminine traits. While the role-incongruent effects are relatively subtler than the strong "men-masculine" and "women-feminine" biases, both response patterns are theoretically important in suggesting that male and female traits may not merely be distinct, but also mutually exclusive. The ability to separate out specific biases (e.g., using the IRAP) therefore distinguishes the current work from previous studies (e.g., Rudman & Glick, 2001), and allows for stronger theoretical conclusions. For instance, the current study provides a starting point for investigating the potentially asymmetrical ways in which we "gender" men relative to women. Several researchers have argued that masculinity is potentially a more rigid social construct than femininity with more well-defined boundaries (Bem, 1993; Leaper & Friedman, 2007; Thorne, 1993; Smiler, 2006), with studies showing that displays of gender-nonconforming behaviour tend to be more actively punished in men relative to women (Adams & Coltrane, 2004; Bem, 1993; Cherry & Deaux, 1978; Cahill & Adams, 1997; Kimmel & Messner, 2009; Leaper, 2002). As discussed in Chapter One, the typical explanation for this effect is that many of our patriarchal or male-dominated social spheres (such as politics or business) place more value on masculine traits, meaning that gender-role deviations are more problematic for men than for women (see Coltrane & Adams, 2008). In this way, there may be more of a resistance to ascribing feminine traits to men than masculine traits to women, as observed here.

Another related explanation for the asymmetry is that patriarchal societies encourage and reward agentic behaviour in women, at least to a certain degree. Wetherell (1986) argued that because masculinity is positioned as the ideal norm within the binary, "progress" for women often manifests as them assuming or aspiring to more masculine roles. In contrast, because femininity is cast as the weaker, less desirable "other", there is less of a tendency or expectation for men to assume these roles. Recent surveys suggest attitudes such as these are changing across Western society (e.g., Bolzendahl and Myers 2004; Brewster and Padavic 2000; Cotter et al. 2011; Fine-Davis, 2013), potentially due to the increased exposure of women in more traditionally male-dominated spheres (Eagly & Carli, 2007; Miller et al., 2018). Responses on the gender-competency IRAP cohere with this, given that both masculine-competent *and* feminine-competent effects were observed. This could suggest the feminine gender role is changing, or at least that femininity is not as under-valued as it was historically. Indeed, the absence of anti-feminine bias on the "firing" preference item in Chapter Two further supports this explanation. The overlap between masculinity, agency and hirability is still significant and does point to the indirect way women may experience workplace discrimination; however, these effects overall suggest the feminine role is more flexible and potentially varied relative to masculinity. This will be assessed in more detail in the pooled analyses planned for the end of this thesis.

The absence of significant differences based on either trait valence or relational terms is also interesting to note. As mentioned above, this pattern of results could be interpreted as evidence that gender stereotypes are broad constructs that encompass a range of positive, negative, descriptive, and proscriptive norms (Eagly & Karau, 2002; Prentice & Carranaza, 2002). This may explain why participant

responses did not change across different binary IRAPS, or why similarly strong biases were found on both. However, these results could also be due to the IRAP's potential insensitivity to such subtle stimulus or procedural modifications. Aside from the Remue et al. papers in the context of depression (2013, 2014), no research to date has examined the impact of valence or relational qualifier changes on IRAP effects. It is therefore difficult to know whether responses should be expected to vary here or not. The IAT literature also cannot inform these findings, as the IAT is a categorization task and does not allow for the inclusion of relational terms (e.g., such as *are* or *should*). Some other more recent implicit measures of relational beliefs suggest effects do change depending on the relational terms used (e.g., Heider, Spruyt & De Houwer, 2018; Tobboel, De Houwer, Dirix & Spruyt, 2017); however, these measures are quite procedurally different to the IRAP. A full investigation of this explanation would require a more detailed (and indeed more basic) methodological research than is possible in the current thesis, but future research should consider exploring this.

With regard to gender differences, effects were stronger for males on nearly all trial-types, as expected. Effects were also strongest for all participants on the trial-type that was congruent with their own gender (i.e., the men-masculine trial type for males and women-feminine trial type for females) in all studies. Interestingly, these effects were not symmetrical across males and females, with men demonstrating more pronounced men-masculine and men-*not*-feminine biases than women on the comparable women-feminine and women-masculine trial types. It is important to note at this stage that participant gender did not exert the same impact across all studies, and main effects were only found in Studies Three and Five. As gender differences will be examined in detail using a measurement invariance

analysis on the pooled IRAP dataset, a detailed analysis of these differences will not be done until the General Discussion of the thesis.

Hiring Preferences

Hiring preferences across all studies supported feminist theories of patriarchal, male-dominated gender order. As in Chapter Two, while most people elected not to express a direct gender preference (i.e., between a man and a woman), a significant percentage of participants in Study Three elected to hire the agentic over the communal individual. While the individuals were not explicitly described as masculine or feminine, the effects on the IRAP and trait rating scale suggest these categories are meaningful proxies for stereotypical masculinity and femininity. The findings in this Chapter thus cohere with existing research showing an overlap between stereotypes of masculinity and occupational success (Arkkelin & Simmons, 1985; Carli, LaFleur, & Loeber, 1995; Heilman et al, 1989; Duehr & Bono, 2006; Jackson & Engstrom, 2007; Powell & Butterfield, 1979, 1984, 1989; Schein, 1973, 1975), and demonstrate that traditional masculinity remains a more valued and desirable construct in workplace settings.

As with the IRAP, the absence of consistent gender differences in hiring preferences is worth noting at this stage. Generally speaking, research suggests men tend to be more conservative and display more pro-male scores on measures of gender beliefs (see Russell & Trigg, 2004 for a review). However, significant differences were only found in one of the three studies (Study Four). This could be due to differences in the gender beliefs across participants in these studies, however explicit sexism scores did not vary in a similar way across them. A more likely explanation is the previously mentioned issue with the masculine-agentic stimulus category. If these traits are not considered by women to be strongly and uniquely

male (at least relative to the men in these studies), then the effects overall may be weaker for women compared to men. Indeed, studies by Jackson and Engstrom (2007) and Duehr and Bono (2006) both found that while male participants were more likely to assume an agentic leader was a man than a woman, female participants were not. Moreover, a study of leadership style by Schein (2001) suggested women (but not men) are increasingly adopting androgynous styles in the workplace, while a review by Eagly (2004) suggests women's self-concept has become considerably more agentic in recent decades. Again, these claims will be informed by the more detailed gender difference analyses planned for later in the thesis.

In addition to demonstrating an overall pro-masculine/agentic bias, Studies Four and Five found significant gender stereotyping in hiring preferences. Both male and female participants tended towards hiring the feminine-communal person for jobs traditionally associated with women and the masculine-agentic person for jobs traditionally associated with men. These data are consistent with previous research into the sex differentiation of different careers (i.e., the jobs considered typically female and male: e.g., Rueben et al., 2017); however, the current study is one of the first to examine gender *role* stereotypes in this context. That is, the extent to which agentic and communal traits are more hireable for different occupational roles. Very little research has examined occupational trait stereotypes in an Irish context, so these studies make a unique contribution to the contemporary literature on workplace gender discrimination in Ireland.

As the current set of studies focused exclusively on hiring potential, future research should explore whether binary biases are associated with other forms of occupational discrimination. For instance, experiments could examine biases towards

hiring role-congruent over incongruent individuals (i.e., agentic versus communal women, etc.), as has been done previously (see Isaac et al., 2009 for a meta-analysis). Drawing from the research into leadership discrimination and the "glass ceiling"/"labyrinth" models of inequality (Eagly & Karau, 2002), research may also look at the relationship between binary biases and a willingness to promote or consider a woman for a leadership role. System justification in occupational settings could be another important area for future investigation, given that previous studies have found a relationship between sexist beliefs and a tendency to justify or naturalize systems of inequality (Glick & Fiske, 2001; Jost et al., 2005; Skewes et al., 2018). Organizational analyses suggest much of the prejudice experienced by women in the workplace is systemic and indirect (e.g., gender-biased policies, etc.: see Stamarski & Son Hing, 2015), so it is possible that binary biases predict the acceptance or endorsement of such practices.

Relationship of IRAP Scores to Hiring Preferences

While the IRAP did not correlate with explicit sexism in any study, the descriptive gender binary IRAP was associated with a pro-masculine hiring preference in Studies Four and Five (note that measure comparisons could not be done in Study Three due to the skewed data). These results are similar to the small number of studies that have been conducted in this area (Rudman & Glick, 2001; Latu et al., 2011) and supports the mainstream feminist argument that the binary plays a role in discrimination (Butler, 2002; Hegarty et al., 2018). Interestingly, there was no relationship between gender-competency evaluations or indeed prescriptive IRAP scores; there was also no relationship between any IRAP scores and gender stereotyping biases. The paucity of research directly assessing the role of implicit bias in hiring discrimination makes it difficult to interpret either the significant or

insignificant effects, and these data would require substantiation in order to draw any conclusions. That said, however, there does appear to be some relationship between the extent to which gender categories are binarized on the IRAP and a tendency to value masculinity in occupational settings. The absence of any correlations between self-reported sexism and either hiring task or IRAP scores should be noted at this stage; however, this will be discussed in more detail after the pooled analysis at the end of the thesis.

Conclusions

These data provided preliminary evidence for the IRAP's utility as a measure of binary biases. While the effects produced by the task are novel, they do cohere with existing social psychological, feminist, and social constructionist models of gender. Studies in this chapter also provided support for theories of patriarchal or male-dominated gender order, at least in occupational settings. Specifically, in demonstrating that agency is more both hireable and more readily ascribed to men, the data obtained here provides direct evidence of the implicit societal overlap between masculinity and occupational success. Future experiments will now assess the relationship of IRAP scores to other socially-relevant beliefs and behaviours. They will also explore in more detail the gender differences, which varied across studies here.

Chapter 4

Binary Biases and Androcentrism

4.1 Introduction

As noted in the General Introduction, male dominance under patriarchy generally manifests in two ways: male *supremacy* and male *centrality* (e.g., Butler, 2002; Ridgeway, 2011). The previous two chapters focused primarily on male supremacy, demonstrating a link between participants' tendencies to binarize gender and prioritize maleness/masculinity in a socially-valued setting (the workplace). The current chapter will expand on these findings by examining whether gender binary biases are also associated with a tendency to centralize and/or normalize maleness. That is, a tendency to assume that men are more prototypically human than women. Across two studies, samples of undergraduate students will complete IRAPs measuring gender binary and other human-as-male biases. In addition, participants in both studies will complete a measure of "androcentric bias", which may broadly be defined as the practice of equating humanity with maleness (Bem, 1996).

In this chapter, participants' bias towards assuming or attributing greater humanity to men will be measured using both implicit and explicit measures. Study Four will assess the relationship between scores on the gender binary IRAP, participant gender, and androcentric bias. As no measure of androcentric bias existed that suited the needs of the current studies, a measure was developed that required participants to choose either "male" or "female" when presented with a gender-ambiguous stimulus. Study Five will be identical except for the IRAPs; instead of gender binary biases, this study will examine the extent to which participants differentially relate men and women with complex or "uniquely human" characteristics. Two IRAPs will be used for this purpose, with the first measuring associations between gender and uniquely human emotions, and the second gender and scientific ability.

4.2 Study Six

Measuring Androcentric Bias

As mentioned in Chapter One, feminists have long acknowledged the male-centric nature of Western societal practices (Gillman, 1911; De Beauvoir, 1979). These range from the androcentric biases in language (e.g., masculine generics or universals) to the various societal practices which normalize and centralize masculinity (e.g., the centrality of men or male stories in media, history, scientific practice, etc.: Bem, 1993: Hegarty, 2006). While there are several ways to analyse androcentric practices at the cultural level, androcentrism has received comparably little attention in the laboratory (Bailey, LaFrance & Dovidio, 2018). Of the research that has been conducted, most has explored how exposure to androcentric language or information (e.g., sex-biased terms or masculine generics) may influence gender cognition or behaviour. Hamilton (1991) and Ng (1990), for example, both revealed that presenting participants with masculine generic terms (e.g., *mankind*) resulted in significantly more male-biased mental imagery than gender-neutral terms (e.g., *they)*. Other studies similarly found that masculine generics or sex-biased language increased the likelihood of attributing a male gender to a gender-neutral or ambiguous character (Bailey & LaFrance, 2016; DeLoache, Cassidy, & Carpenter, 1987; Gastil, 1990; Miller, Taylor, & Buck, 1991; Moyer, 1997; van Berkel, Molina, & Mukherjee, 2017). Eagly and Kite (1987) used a different approach, instead assessing the relative extent to which men and women are considered typical humans. In their study, they asked participants to rate stereotypically male and female traits in terms of how typically human they are, revealing a significant effect between typically male and human traits. Two studies by Bailey and colleagues also found participants more readily paired gender-inclusive words (e.g., *people*) with

male faces than with female faces (Bailey & LaFrance, 2018), and selected more male faces when asked to identify the typical human (Bailey et al., 2018). Interestingly, the second Bailey et al. study also explored the relationship between "gender polarization" (defined as the extent to which participants believe men and women are opposites) and androcentric bias, though it found no significant effect. With regards to gender differences in androcentric behaviour or bias, results are varied; some suggest men are more likely to produce male imagery and women more likely to remain neutral (Gastil, 1990; Switzer, 1990). However, other studies have found no difference in androcentric behaviour between male and female participants (Eagly & Kite, 1987; Lambdin et al., 2003).

While the above humanity-attribution or human-gender matching paradigms provide useful insights into androcentric thinking, no measure currently exists that assess participants' likelihood of assuming maleness in the abstract; that is, their androcentric preference independently of specific cues (e.g., masculine generics). Such assumptions may be theoretically important as they could represent a generalized propensity to centralize men or maleness relative to women, which has not been examined to date. The current chapter will therefore employ a novel task developed by the researchers for the specific purposes of these studies. This task will require participants to attribute gender to a neutral stimulus, and will have a binary, forced-choice response format (e.g., male or female). Similar to other recent research in this area (Bailey & LaFrance, 2018; Bailey et al., 2018), the task will use facial stimuli and will require participants to attribute a gender to an ambiguous, composite facial stimulus.

Current Study

Study Four will assess the relationship between gender binary and human-as-male (i.e., androcentric) biases. The same gender binary IRAP as in Chapter Two will be used here, and androcentrism will be measured using a novel task in which participants will evaluate a gender-ambiguous stimulus as either male or female. Based on the theoretical and empirical literature in this area, it is hypothesised that a greater number of participants will rate the gender-ambiguous stimulus as "male" rather than "female". Similarly, given the existing literature and the response pattern observed thus far in this thesis, it is expected that men may be more likely than women to select the male response option. As no research to date has looked at the relationship between implicit gender-binary biases and any form of androcentrism, no specific hypotheses are proposed around the relationship between IRAP scores and responses on the androcentrism task. However, it is tentatively expected based on feminist theorizing that more pronounced IRAP scores will be associated with a tendency to attribute a male gender to an ambiguous stimulus, and that this may interact with participant gender.

4.2.1 Method

Participants

Forty-four White Irish undergraduate students aged between 18 and 27 participated in this study (25 female, 19 male; $M_{age} = 20.8$). Sampling, recruitment methods and inclusion criteria were the same as the previous study. **Procedure**

The experimental sequence including instructions and debriefing information were identical to previous studies. The order of the tasks was as follows: androcentrism task, IRAP, Modern Sexism Scale.

Materials

The gender binary IRAP and MS scale were identical to previous studies.

Androcentrism Task

While androcentrism is a broad ideological construct that can manifest in multiple ways (Bailey et al., 2018), it was operationalized here as a tendency to assume maleness when presented with gender ambiguity. As no validated measure of (abstract) androcentric bias existed in the literature, a brief empirical assessment was developed in which participants had to assign a gender to an image of a gender-neutral face. In this task, participants were presented with a series of seven faces, one of which was the "target" composite gender-neutral image. The additional six faces ranged from very female to very male and were included both to provide a range and to conceal the purpose of the study from participants. The gender-neutral image was created using composite image software. All images were obtained from the following two datasets and permission was granted from the lead author for their reuse: Rhodes, Hickford and Jeffrey (2000) and Rhodes, Jacquet, Jeffrey, Evangelista, Keane and Calder (2011). Participants were required to select either "male" or "female" for all seven faces, which were presented one at a time and in a fixed random order (Appendix 8). The following instructions were provided prior to the task:

"Below you will be presented with a series of faces. For each image, please select whether you think the person is male or female. There is no need to spend too long on each item, but do try and be as accurate as possible."

Response options were presented on-screen below each face and the order of the male/female responses was randomized across items. As we are only interested in the responses for the gender-neutral image, the androcentrism task produced a single categorical variable: female or male choice.

Ethical Issues

Ethical issues and procedures were identical to the previous studies.

Data processing and analysis

IRAP data were processed in an identical manner to studies one and two and participants were excluded using the same criteria. Five participants were excluded on this basis (final $n = 39$)

4.2.2 Results

Androcentrism Task

Responses for the androcentrism task are in Table 4.1. For the ambiguous facial stimulus, responses were nearly evenly split across the sample (i.e., broadly equal numbers of participants attributed a female as a male gender). A chi-square goodness of fit test found no significant deviation towards either the male or female response option, $p = .89$, and a chi-square test of independence found no significant relation between participant gender and the male/female response option, X^2 (1, $N = 44$) = 1.39, $p = .24$.

Table 4.1

Reponses on the Androcentrism task by Gender

	Participant gender		
Androcentrism Task	**Female**	**Male**	**Total**
Selected "Female"	15	8	23
Selected "Male"	10	11	21
Total	25	19	44

Modern Sexism Scale

MS scores for this sample were again somewhat low and skewed to the left of the scale ($M = 18.6$, $SD = 5.7$; possible range 10-50). An independent samples t-test found no difference between male ($M = 19.05$, $SD = 4.99$) and female ($M =$

18.24, *SD* = 6.2) participants, $t(42) = -.47$, $p = .643$, though male scores were slightly higher.

IRAP Performance

Effects on the gender binary IRAP were similar to previous studies (see Figure 4.1 for a graphical depiction and Table 4.2 for means and standard deviations). Male and female response patterns did not differ notably, though D_{IRAP} scores were slightly more binary-consistent for men. That is, males demonstrated more positive scores on the men trial-types and more negative scores on the women trial-types.

Figure 4.1 Mean binary IRAP trial-type scores for female and male participants

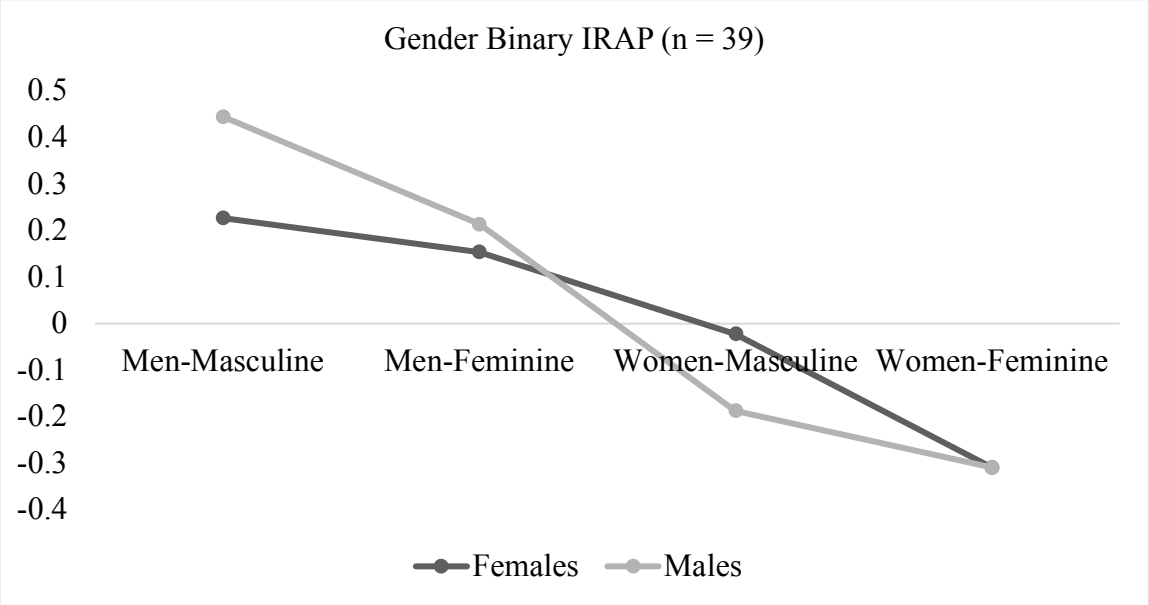


Figure 4.1. Mean trial-type level D_{IRAP} scores for male and female participants. Note: D_{IRAP} scores can range from -2 to +2. Positive scores reflect a "masculine" or "not-feminine" response pattern and negative a "feminine" or "not-masculine" response pattern.

Table 4.2

Means and standard deviations for the Binary IRAP trial-types

Trial-Type	Gender	N	Mean	SD
Men-masculine	Female	24	0.23	0.41
	Male	15	0.45	0.48
Men-feminine	Female	24	0.15	0.31
	Male	15	0.21	0.41
Women-masculine	Female	24	-0.02	0.30
	Male	15	-0.19	0.41
Women-feminine	Female	24	-0.31	0.34
	Male	15	-0.31	0.40

One sample t-tests were run to assess the significance of the trial-type effects against zero (Table 4.3). These were run on the entire sample to minimize the number of comparisons, and a Bonferroni correction was applied (reducing the acceptable *p*-value to .00125). As in previous studies, significant men-masculine, women-feminine and men-*not*-feminine effects were found.

Table 4.3

One-sample t-test results for IRAP trial-types

Trial-Type	t	df	p
Men-masculine	4.397	38	< .001***
Men-feminine	3.207	38	0.003***
Women-masculine	-1.537	38	0.133
Women-feminine	-5.433	38	< .001***

Note. *$p < .05$, **$p < .01$Acceptable *p*-values after Bonferroni correction </= .0125

Measure and Group Comparisons

Gender, Androcentric Bias, and IRAP scores

A 2x2x4 repeated measures ANOVA was conducted to examine the relationship between participant gender (2 levels: male or female), androcentric bias (2 levels: male or female assumption), and trial-type level IRAP scores (4 levels).

There was no main effect for gender, $F(1, 35) = .004$, $p = .947$, or androcentric bias, $F(1, 35) = 2.27$, $p = .14$, though there was a significant two-way interaction effect between androcentric bias and trial type, $F(3, 105) = 20.78$, $p = .014$. No other significant interaction effects were found (all ps > .4).

Follow up independent-samples t-tests were run to investigate which trial-type/s were potentially driving the differences in IRAP effects between those who chose the male relative to the female. These analyses revealed a significant difference on the men-masculine trial-type only, $t(37) = -3.5$, $p < .001$, with scores higher for participants who assumed male ($M = .523$, $N = 20$) than female ($M = .088$, $N = 19$). No other significant differences were found on any of the other trial-types (all ps > .1), though participants who chose the male had more binary-consistent effects on all four. Given the relatively small sample size of this study and the number of comparisons being made, however, this effect should be interpreted cautiously.

Gender, Androcentric Bias and Modern Sexism

A 2x2 ANOVA was conducted to examine any relationship between gender, androcentric bias, and scores on the Modern Sexism Scale. There was no main effect for gender, $F(1, 40) = .106$, $p = .75$, or androcentric bias, $F(1, 40) = .89$, $p = .35$.

Modern Sexism and IRAP Scores

A Pearson's r analysis found no significant relationship between overall D_{IRAP} effects and MS scores, $r(35) = .019$, $p = .45$.

4.2.3 Summary

Response patterns on the IRAP were the same as in previous studies, though there were no significant gender differences. This will not be discussed here as gender differences will be explored in more detail towards the end of this thesis in

the chapter devoted to pooled analyses. For the androcentrism task, participants were not more likely to assume maleness in the context of gender ambiguity. This is in slight contrast to existing research and theory in this area. However, the existing research has focused predominantly on attributions of humanity to male/female stimuli, or to the relative impact of androcentric cues on mental imagery (Eagly & Kite, 1987; Hamilton, 1991; Bailey et al., 2018) or andro/gynocentric response options (e.g., Lambdin et al., 2003). As this is the first study to assess androcentric bias in the abstract– that is, the tendency to attribute maleness or femaleness to a gender-neutral stimulus independently of specific cues or additional information – it is difficult to interpret this outcome. However, this task will be replicated in Study Seven, which will further inform the conclusions drawn here.

Concerning the relationship between the measures, it is interesting to note that a male choice was associated with binarist effects on the IRAP (specifically speedier responding to male-masculine-*true*), though this did not interact with participant gender. Effects are thus broadly similar to the results of the Bailey et al. (2018) study, which found no relationship between gender polarization (measured on an explicit rating scale) and androcentric bias, although the specific relationship to the male-masculine trial-type should be followed up in future studies. There was no relationship between explicit sexism and androcentric bias, and this effect also did not vary according to participant gender. Again, this analysis was likely underpowered because the sample size was low and there were multiple pair-wise comparisons, but this issue will be addressed in the following study.

4.3 Study Seven

The previous study tested the theoretical assumption that gender binarization is related to a tendency to assume maleness in the context of gender ambiguity (e.g.,

Bem, 1996). Results on the androcentrism task were not in the expected direction, and the relationship between the IRAP and androcentric bias task were weaker than expected. Study Seven will expand on these data by (1) replicating the androcentrism task to increase the power and confidence in the results obtained, and (2) assessing whether androcentric bias is related to effects on other more attitudinally similar gender IRAPs. Two novel IRAPs will be included which together measure the extent to which participants differentially associate men and women with rationality, complexity and logic. One will examine gender-science stereotypes (and thus the differential ascription of scientific or reasoning abilities to different gender groups) and the other *infrahumanisation* biases. The significance of both to a study of androcentrism – and indeed gender ideology more generally – will be discussed below.

Infrahumanisation

Labelled the "emotional side of prejudice", infrahumanisation refers to the societal or psychological tendency to view members of a particular social group as less or sub-human (Leyens et al., 2000, 2001, 2003). Somewhat distinct from the concept of *de*humanization, which describes the more explicit stereotyping of social groups as animals or objects (e.g., referring to a woman as "it", or the "Negro-ape" metaphor: see Rudman & Mescher, 2012), infrahumanisation describes the subtler denial of uniquely human characteristics, abilities or emotions to one social group over another (see Demoulin et al., 2002). Uniquely human traits refer here to those which distinguish humans from simpler species, and so include higher-order cognitive abilities and emotions (e.g., rationality or compassion). In contrast, non-uniquely human traits would be those shared between humans and other simpler mammals or primates (e.g., rudimentary forms of intelligence or basic emotions).

The tendency to infrahumanize and deny complex humanity to social groups has been demonstrated in numerous contexts. Using basic infrahumanisation paradigms in which participants relate in and outgroups with primary (i.e., simpler) or secondary (i.e., higher order) emotions, several studies demonstrate the pervasiveness of this bias towards outgroups in contexts of race, culture and ethnicity (e.g., Cortes, Demoulin, Rodriguez, Rodriguez, & Leyens, 2005; Demoulin et al., 2009; Gaunt, 2009; Gaunt, Leyens & Demoulin, 2002; Paladino, Vaes, Castano, Demoulin, & Leyens, 2004; Rohmann, Niedenthal, Brauer, Castano, & Leyens, 2009). Studies also suggest these biases are relatively automatic, with similar effects found on implicit measures such as the IAT (e.g., Boccato, Cortes, Demoulin, & Leyens, 2007; Viki et al., 2009). As with androcentrism or other seemingly benevolent beliefs, infrahumanisation biases are considered harmful and have been shown to increase the likelihood of intergroup conflict and discrimination (e.g., Cuddy, Rock, & Norton, 2007; Pereira, Vala, & Leyens, 2009; Vaes, Paladino, & Leyens, 2004).

Gender Infrahumanisation

While infrahumanisation has predominantly been studied in the context of race and ethnicity, there is a rationale for exploring these biases in the context of gender. First, if patriarchal societies are indeed androcentric (Bem, 1996, Pilcher & Whelehan, 2004), then it is likely that men will be afforded greater humanity and/or complexity than women. Second, while not as widely studied as ingroup/outgroup biases, some studies suggest that infrahumanisation does occur across high/low status groups (e.g., between blue and white-collar jobs; Leyens et al., 2001). As men are the higher status group under patriarchy, similar effects may be observed across gender categories. Indeed, Fiske et al. (2001) found that infrahumanisation correlated

with stereotypes about a group's reduced competence or status; as competency stereotypes have been shown to be gendered (e.g., Eagly & Karau, 2002), this may mean women are more likely to be infrahumanised relative to men. Third, there is considerable conceptual overlap between the "uniquely human" characteristics and traditional masculine-agentic stereotypes (i.e., in terms of the shared rationality, logic, and capacity to disregard basic emotions: see Adam, 1995 and Weber, 2005). As such, it is possible that some bias towards association these traits with men will be found here.

Some research to date has explored infrahumanisation in the context of gender, all using self-report measures. Using the same basic primary/secondary emotion attribution paradigm described above, Viki and Abrams (2003) and Gaunt (2013) both explored the relationship between sexist beliefs and infrahumanisation. Interestingly, these studies reported effects in the opposite direction (i.e., secondary emotions were more readily attributed to women than men), though these effects were mediated by both gender beliefs and participant gender. Specifically, across both papers, male participants high in "hostile sexism" (explicitly negative or hateful beliefs about women: see Glick & Fiske, 1995) attributed primary emotions to women, while men high in "benevolent sexism" (well-intentioned but patronizing beliefs about women) had an increased likelihood of attributing secondary emotions to women. No significant effect was found for female participants in either direction, though in both studies women scored significantly lower on hostile and benevolent sexism.

No research to date has examined gender infrahumanisation using an implicit measure. However, given that research in other contexts suggest these biases are relatively automatic (e.g., Viki et al., 2009) combined with the IRAP's utility in

separating out biases at the trial-type level (i.e., its ability to measure men/women-complex/simple biases simultaneously), an IRAP analysis of gender infrahumanisation may prove a useful addition to the literature. Two studies have explored gender *de*humanization biases using an implicit paradigm, both exploring the differential association of men and women with object and animal stimuli. Using the IAT, Rudman and Mescher (2012) found that women were more strongly associated with inanimate objects than were men. Interestingly, this study also revealed a relationship between women-object IAT scores and sexual harassment proclivity. Hussey et al. (2015) similarly found associations of women but not men with inanimate objects on an IRAP. The current study will use a similar approach but instead of animal/object associations, will measure the association of primary/secondary emotions with gender groups.

Considering that studies in this thesis use both male and female samples, it is important that the studies can separate out in/outgroup bias from beliefs about gender categories. That is, it is important that the study does not merely measure male and female participants' in-group biases, given that this would not be the focus of the current thesis. According to Gaunt (2013), while prejudice in many intergroup contexts is determined by group membership (i.e., White vs. Black racial difference, etc.), gender dynamics are fundamentally different. Building on the claims made in Ambivalent Sexism Theory that both women and men may exhibit gender prejudice due to having access to the same mainstream gender ideology and information (Glick & Fiske, 2001), Gaunt suggests it is a person's core gender beliefs and not their group membership *per se* that is more relevant to an analysis of gender infrahumanisation. As such, it is possible that both men and women would attribute less complex humanity to women, and that this may be influenced or mediated by

gender binary biases. Misogyny, sexism and traditional gender beliefs are typically higher in men than women (see Russell & Trigg, 2004), so the interaction between infrahumanisation biases and participant gender would still be important to measure.

Gender-Science Biases

Given the broader focus in this chapter on gendered associations of complex humanity, it may also be interesting to explore gender biases related to scientific ability. According to a review by Leyens et al. (2001), intelligence is one of the three most cited characteristics associated with humanity (along with secondary emotions and language). While intelligence is of course a multi-faceted construct, theoretical and lay definitions of intellectual ability do overlap considerably with abilities associated with scientific inquiry (i.e., the capacity to reason, learn, problem solve and adapt to a changing environment; Neisser et al., 1996; Sternberg, 1981), and STEM careers are generally rated among the careers requiring the highest IQ (Richardson & Norgate, 2015). As such, the differential association of men and women with science careers may represent a useful proxy assessment of their perceived capacity for complex or higher-order reasoning ability.

Gender-science stereotypes are well-established in the literature and have been documented across a series of cultural contexts. A recent study of more than 350,000 participants across 66 nations, for example, found significant "scientist-as-male" stereotypes in all contexts, and particularly those with low representation of women in science (Miller, Eagly, & Linn, 2015). Another study by Smyth and Nosek (2015) found significant male-favouring science stereotypes among a sample of over 176,000 college graduates, with effects again significantly higher in male-dominated and STEM disciplines. Content analyses of scientist depictions in textbooks, advertisements and online image searches similarly note a bias towards a male

representative (Pew Research Centre, 2018; Potter & Rosser, 1992; Barbercheck, 2001), while a recent meta-analysis of studies employing the "draw-a-scientist" paradigm showed both children's and adults' sketches of a prototypical scientist are significantly more likely to be male (Miller, Nolla, Eagly & Uttal, 2018). The Gender-Science IAT available online at *ProjectImplicit*, for example, has been completed over 500,000 times (Zitelny, Shalom & Bar-Anan, 2017). Using common careers associated with STEM and the arts (e.g., Biology vs. English), this IAT has reliably documented a bias towards associating women with arts/humanities and men with science/math (Kessels, Rau, & Hannover, 2006; Lane, Goh, & Driver-Linn, 2012; Nosek, Banaji, & Greenwald, 2002; Nosek et al., 2007; Nosek & Smyth, 2011). While the IRAP has not to date been used in this context, it may have certain procedural advantages to the IAT. Specifically, given that it can assess biases at the trial-type level, it may provide further insight into whether the effects are driven by men-science biases, women-arts biases, or both.

As with other gender beliefs discussed in this thesis (e.g., leadership stereotypes), gender-science stereotypes are believed to originate early in life from our broader gender ideologies (e.g., Bian, Leslie & Cimpian, 2017). Miller, Nolla, Eagly and Uttal (2018), for instance, recently extended Eagly's Social Role Theory to gender-science stereotypes, suggesting that assumptions of greater scientific ability in men are due in large part to the communality-agency gender role distinction. According to these authors, this differentiation means there is a cultural congruity between masculine agency and scientific ability, meaning that men seem more naturally suited to STEM careers. Indeed, a recent study by Carli and colleagues explored the relationship between gender, communality/agency stereotypes and the stereotypes associated with a successful scientist. As would be

expected, this study found that agentic traits were significantly more likely to be ascribed to both men and scientists (but not women), while communal traits were associated with women (but not men or scientists: Carli, Alawa, Lee, Zhao & Kim, 2016). In keeping with the findings thus far in this thesis, however, female participants also perceived more similarity between women and scientists than male participants and they also rated women as significantly more agentic. This study thus points to the relevance of gender-science stereotypes to our broader gender ideology, and the way in which communality-agency and participant gender mediate gender-science stereotypes.

Current Study

The goal of the current study is to further assess the centrality of men within a binarist gender system. Using the same androcentric bias task as Study Six, the current study will examine the relationship of androcentrism to other implicit gender biases. Two IRAPs will be employed: the first will assess gendered associations of simple and complex emotions, and the second gender-science stereotypes. Across both IRAPs, it is hypothesised participants will more readily associate men with more "uniquely human" characteristics (operationalized here as complex/secondary emotions and scientific ability). As in previous studies, weaker effects are expected for the role-incongruent trial-types for women relative to men (i.e., on the women-complex or women-science trial-types). Concerning the relationship between measures, gender theory would suggest that effects on these IRAPs would be related to a tendency to assume maleness when presented with gender ambiguity (i.e., androcentric bias: Bem, 1993; Bailey et al., 2018). However, this is only a tentative and exploratory hypothesis based on the absence of any IRAP-androcentrism relationship in the previous study.

4.3.1 Method

Participants

Fifty-nine White Irish undergraduate students aged between 19 and 26 participated in this study (30 identified as female, 29 as male; M_{age} = 20.51). Sampling, recruitment methods and inclusion criteria were the same as the previous study.

Procedure

The experimental sequence including instructions and debriefing information were identical to previous studies, with one exception. As with all IRAPs, the infrahumanisation IRAP required the presentation of stimulus category labels in the pre-block rule and instruction screens. While there are potentially multiple category labels that could have been used, the current study employed the labels "complex emotions" and "simple emotions". This was because other theoretically meaningful labels (e.g., primary/secondary emotions, more/less human, etc.) were deemed unlikely to form an intuitive or natural stimulus category for participants. However, as the "complex" and "simple" emotions were also potentially weak categories, participants were provided with a stimulus reference sheet prior to the study (Appendix 9), similar to the masculine/feminine trait reference sheet used in Chapter Three. Participants were instructed to learn the stimulus classes and to only start once they felt comfortable, and could consult the sheet between blocks if needed. The task structure (including the response options and block structure) of the IRAP was identical to previous. The order of the tasks was as follows: androcentric bias task, two IRAPs (the order was counterbalanced across participants), and self-report measure of sexism.

Materials

Androcentric Bias Task

The measure of androcentric bias was identical to study four.

IRAPs

Two separate IRAPs were employed in this study: one for infrahumanisation biases (men-complex/women-simple) and one for gender-science biases (men-science/women-arts). Stimuli for the infrahumanisation IRAP were obtained from lists of primary and secondary emotions in Demoulin et al. (2009; see Table 4.4). As in Demoulin et al., stimuli were matched for valence and broad meaning across the two categories (i.e., "Contempt" with "Anger", and so on). For the gender-science IRAP, the arts and science subject categories were identical to those used in the *ProjectImplicit* gender-science IAT. This was accessed via the *ProjectImplicit* website (www.projectimplicit.org) in September 2015.

Table 4.4

Stimulus categories for the infrahumanisation and gender-science IRAP

Infrahumanisation IRAP					
Label 1: Men	Label 2: Women	Target 1: Complex emotions	Target 2: Simple emotions	Rule A	Rule B
Men feel	Women feel	Apprehension Contempt Pride Disappointment	Fear Anger Happiness Sadness	Please respond as if men feel complex emotions and women feel simple emotions	Please respond as if women feel complex emotions and men feel simple emotions
Gender-Science IRAP					
Label 1: Men	**Label 2: Women**	**Target 1: Science subjects**	**Target 2: Arts subjects**	**Rule A**	**Rule B**
Men do	Women do	Biology Physics	Philosophy Arts	Please respond as	Please respond

	Chemistry Maths Engineering	Literature English Music	if men do science subjects and women do arts subjects	as if women do science subjects and men do arts subjects

Modern Sexism Scale.

The version of the MS in this study was identical to previous studies.

Ethical Issues

Ethical issues and procedures were identical to the previous studies.

Data processing and analysis

Data were processed in an identical manner to studies one and two and participants were excluded using the same criteria. Four participants were excluded from the infrahumanisation IRAP (final $n = 56$) and two from the gender-science IRAP (final $n = 58$) on this basis.

4.3.2 Results

Androcentric Bias

As evidenced by the figures in Table 4.5, a noticeably larger proportion of participants attributed a female gender to the gender-ambiguous face. A chi-square goodness-of-fit test assuming a test value of .5 (i.e., a 50/50 response split) indicated the distribution was significantly skewed towards assuming female ($p < .001$). A chi-square test of independence found no significant relation between participant gender and the male/female response option, X^2 (1, $N = 59$) = 1.61, $p = .448$.

Table 4.5

Reponses on the Androcentrism task by Gender

	Participant Gender		
Androcentrism task	**Female**	**Male**	**Total**

Selected “Female”	21	22	43
Selected “Male”	9	7	16

Modern Sexism Scale

Sexism scores were slightly higher than previous studies but were again skewed to the left of the scale (M = 20.51, SD = 6.52). An independent samples t-test found no significant difference between males (M = 21.52, SD = 5.74) and females (M = 19.53, SD = 7.15), $t(57) = -1.17$, $p = .25$.

IRAP Performance

Infrahumanisation IRAP

Effects on the infrahumanisation IRAP do not seem particularly pronounced in either a men-complex/women-simple or women-complex/men-simple direction (see Figure 4.2). Mean D_{IRAP} scores are positive for both the men-complex and women-complex trial-types, indicating that participants were faster to respond “true” relative to “false” for both men-complex and women-complex stimulus pairings. Bonferroni-corrected one-sample t-tests for the entire sample revealed significant effects on the women-complex trial-type only (see Table 4.7). These analyses again assumed a test value of 0. The only notable gender differences in responses seem to be on the men-simple trial-type, with stronger effects for female over male participants.

Figure 4.2 Mean infrahumanisation IRAP trial-type scores for female and male participants

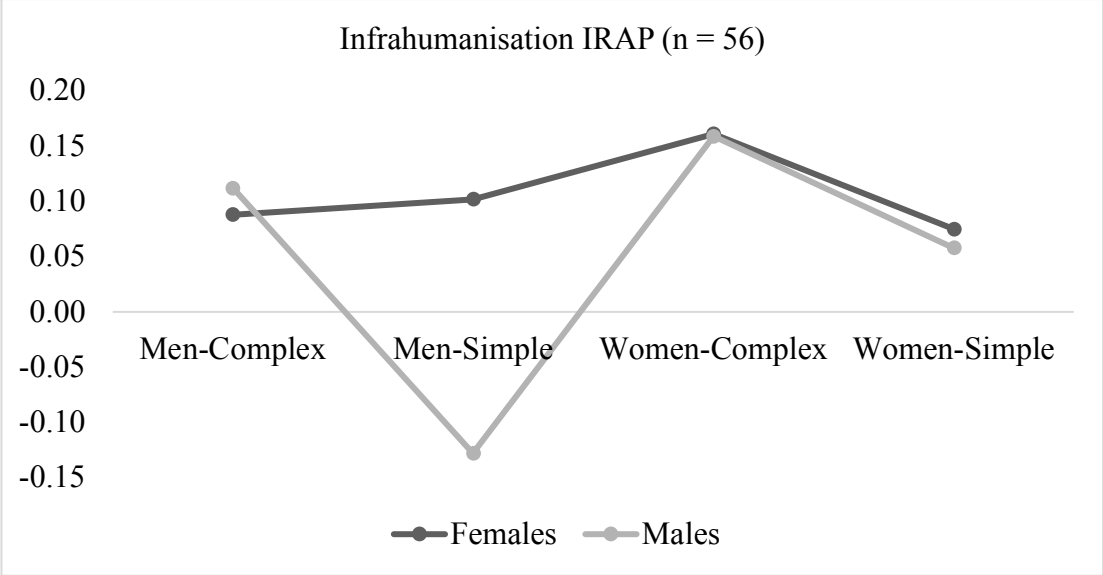


Figure 4.2. Note: positive scores reflect a "complex" or "not-simple" response pattern and negative a "simple" or "not-complex" response pattern.

Table 4.6

Means and standard deviations for the Infrahumanisation IRAP

Trial-Type	Gender	Mean	SD	N
Men-Complex	Female	0.09	0.33	27
	Male	0.11	0.38	29
Men-Simple	Female	0.10	0.27	27
	Male	-0.13	0.29	29
Women-Complex	Female	0.16	0.39	27
	Male	0.16	0.37	29
Women-Simple	Female	0.08	0.32	27
	Male	0.06	0.36	29

Table 4.7

One-sample t-test results for the Infrahumanisation IRAP

	t	df	p
Men-complex	2.118	55	0.039*
Men-simple	-0.425	55	0.673
Women-complex	3.183	55	0.002**
Women-simple	1.469	55	0.148

Note. *$p < .05$, **$p < .01$. Acceptable *p*-values after Bonferroni correction </= .0125

Gender-Science IRAP

The response pattern on the gender-science IRAP was somewhat surprising: there were significant men-science and men-*not*-arts biases as expected; however,

there were also significant women-science effects and no significant women-arts effects (see Table 4.8).

Figure 4.3 Mean gender-science IRAP trial-type scores for female and male participants

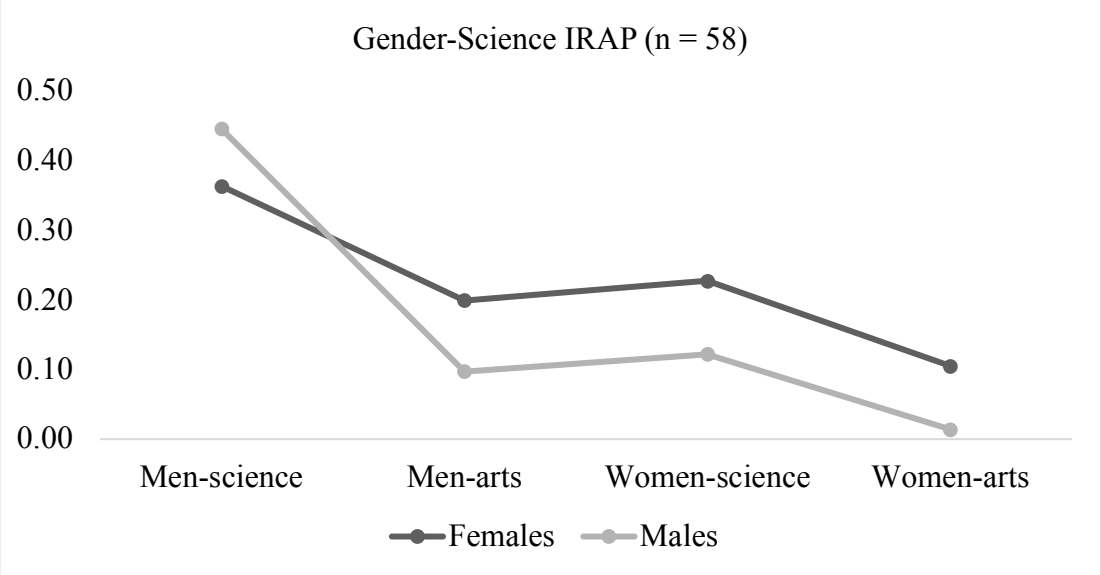


Figure 4.3. Note: positive scores reflect a "science" or "not-arts" response pattern and negative an "arts" or "not-science" response pattern.

Table 4.8

Means and standard deviations for the Gender-Science IRAP

Trial-Type	Gender	Mean	SD	N
Men-Science	Female	0.36	0.34	30
	Male	0.45	0.37	28
Men-Arts	Female	0.20	0.31	30
	Male	0.10	0.35	28
Women-Science	Female	0.23	0.44	30
	Male	0.12	0.34	28
Women-Arts	Female	0.11	0.34	30
	Male	0.01	0.40	28

Table 4.9

One-sample t-test results for the Gender-Science IRAP

	t	df	p
Men-science	8.665	57	< .001***
Men-arts	3.452	57	0.001***
Women-science	3.412	57	0.001***

Women-arts	1.265	57	0.211

Note. *$p < .05$, **$p < .01$. Acceptable p-values after Bonferroni correction </= .0125

Measure/Group Comparisons

Gender, Androcentric Bias and IRAP Performance

The relationship between gender, androcentric bias and IRAP performance was assessed using 2x2x4 repeated measures ANOVAs (one for each IRAP). For the infrahumanisation IRAP ($n = 56$), no significant main effect was found for either gender, $F(1, 52) = .95$, $p = .34$, or androcentric bias, $F(1, 52) = .99$, $p = .33$, and no significant two- or three-way interaction effects were found (all ps > .4). There were also no significant group differences on the gender-science IRAP ($n = 58$) between either males and females, $F(1, 54) = 2.19$, $p = .15$, or participants who assumed maleness relative to femaleness, $F(1, 54) = .19$, $p = .66$, and no significant interaction effects (all ps < .6). A Pearson's r correlation on overall D_{IRAP} scores found no significant relationship between IRAPs ($r = .057$, $p = .678$).

Gender, Androcentric Bias and Modern Sexism

A 2x2 ANOVA was conducted to examine any differences in MS scores by androcentric bias or participant gender. Analyses found no significant main effect for gender, $F(1, 54) = 1.14$, $p = .29$ but a significant effect for androcentric bias $F(2, 54) = 5.15$, $p = .009$ with a medium effect size ($\eta^2 = .149$). A significant interaction effect was also found between androcentric bias and gender $F(1, 54) = 3.95$, $p = .05$ but the effect size was small ($\eta^2 = .057$). A review of the descriptive statistics suggests sexism scores were higher among those who assumed maleness; however, this was particularly strong for female participants (see Table 4.9), indicating that more sexist women were particularly likely to choose the male response option. The small number of participants in this cell (9) and the high variance in responses among this group should be noted however.

Table 4.10

Modern Sexism Means and SDs by Gender and Androcentric Task Preference

Gender	Androcentrism Task	Mean	SD	N
Female	Assumed Female	17.24	4.43	21
	Assumed Male	24.89	9.48	9
Male	Assumed Female	21.14	6.07	22
	Assumed Male	21.50	3.79	6

Note. Modern Sexism scores range from 10-50, with higher scores reflective of more sexist beliefs.

IRAP Performance and Modern Sexism

Spearman's Rho correlations were conducted to examine the relationships between self-reported sexism and overall D_{IRAP} scores for both IRAPs. The non-parametric alternative was used here as scores for the MS were non-normally distributed for this sample. These analyses revealed a weak positive correlation between the MS and gender-science IRAP, $r = .239$, $p = .035$. No other significant correlations were found, though it should be noted that there was a very weak negative relationship between infrahumanisation scores and self-reported sexism.

4.3.3 Summary

Effects on the infrahumanisation IRAP were not as expected. Complex traits were attributed more readily to women rather than men, though overall the effects on the IRAP were weak and significant effects were only found on the women-complex trial-type. Effects on the male-complex trial-type were significant (although not after a Bonferroni correction) and were comparably weaker than the women-complex effects. For the gender-science IRAP, effects were more consistent with existing theory and research in this area but still surprising: significant men-science and men-*not*-arts biases were observed as expected. However, there were also significant women-science biases found and no significant women-arts biases. While somewhat

inconsistent with gender theory in this area, these data do cohere with the results of other studies (i.e., neutral or reversed effects for the trial-type involving women and the stereotypically masculine set of attributes or traits).

Responses on the androcentrism task were both unexpected and inconsistent with the results of the previous study. There was a significant gynocentric response pattern, with nearly three quarters of participants (72.8%) selecting the female response option. No gender differences were found in responses.

Concerning the relationships between measures, effects varied across analyses. There was no relationship between IRAPs and effects on either did not vary depending on participant gender or responses on the androcentrism task. Modern Sexism did weakly correlate with scores on the gender-science IRAP however, and these scores also varied depending on participant gender and androcentric bias. Specifically, scores were higher for participants who selected the male response option, and particularly when those participants were female.

4.4 Discussion

IRAP Performance

Effects on the gender binary IRAP were identical to those observed in the previous chapter. There were significant men-masculine, men-*not*-feminine, and women-feminine biases found, and again no significant women-*not*-masculine bias. As the significance of these effects was discussed in the last chapter and will be reviewed again in the General Discussion, this section will just focus on the two novel IRAPs in Study Seven.

Infrahumanisation IRAP

The first of these IRAPs explored implicit gender infrahumanisation biases, or the gendered associating of women and men with primary and secondary

emotions. Infrahumanisation is a well-documented phenomenon describing the tendency to deny complex or uniquely human characteristics to members of a particular social group (Leyens et al., 2000). Studies indicate this form of "emotional prejudice" is common between members of different racial, cultural and ethnic groups (Cortes, Demoulin, Rodriguez, Rodriguez, & Leyens, 2005; Demoulin et al., 2009; Gaunt, 2009; Gaunt, Leyens & Demoulin, 2002;), as well as between groups of high and low status (Levens et al., 2003). However, in the current study, effects were generally weak and in the unexpected direction, with significant biases found only on the "women-complex" trial-type.

As only a small number of studies have investigated infrahumanisation in the context of gender, it is somewhat difficult to contextualise the current findings. However, the absence of any significant or expected effects on this IRAP could be explained in a few ways. First, it could be that these biases are simply not present in the current sample or population, and that infrahumanisation is not a component of normative gender beliefs. Previous studies have identified a societal tendency to objectify or *de*humanize women (i.e., attribute animalistic or inhumane attributes to them: Hussey et al., 2015; Rudman & Mescher, 2012); however, it is possible that the denial of complex or secondary emotions to women is not a feature of gender prejudice. The previous research in this area only found infrahumanisation biases among males high in hostile sexism (Abrams, 2003; Gaunt, 2013), and indeed the Gaunt paper found a reversed effect in men high in benevolent (i.e., well-intentioned) sexism. Though there was no relationship between sexism, gender and IRAP performance in this study, future research could investigate this further by employing the same sexism measures used in the previous studies.

Another explanation for the effects on this IRAP could be the specific category labels used (i.e., "simple" and "complex"). It is possible that these labels did not adequately represent the individual exemplars, or that the stimuli together did not form a coherent category outside of the IRAP. This could explain the high variance in responses and overall weak effects, as the task may have made less overall sense (relative to other IRAPs in this thesis). As discussed in Chapter Two, research suggests inappropriate or incoherent category labels can confound effects on implicit measures (Fazio & Olsen, 2003; Steffens et al., 2006), and so this may have been a factor here. The weak or unexpected effects could also have been due to the gendered connotations for individual exemplars. For example, "Anger" (used here as a primary/simple emotion) has been shown to be more readily associated with men than women (e.g., Brescoll & Uhlmann, 2001). Similarly, "complexity" – and indeed emotionality more generally – is more coherent with the overall feminine stereotype (Broverman et al., 1990; Eagly & Karau, 2002; Koenig, 2018). As such, it could be that the category labels and individual exemplars used here interacted in unexpected ways with traditional gender stereotypes and influenced the overall results.

Gender-Science IRAP

For the gender-science IRAP, effects on the male trial-types were consistent with the hypothesis and existing research in this area. As in previous explicit and implicit research, participants in this study showed a significant bias towards associating men with science (i.e., responding quicker to men-science-*true* than men-science-*false*) and against associating men with arts/humanities (i.e., responding quicker to men-arts-*false* than men-arts-*true*). This effect is significant because this role congruity between men, agency and science has been widely problematized as

contributing to the gender discrimination experienced by women in STEM (Carli et al., 2016). Research suggests that such stereotypes significantly impede female scientists' career prospects and success, and significantly reduce their likelihood of being hired, promoted or funded (see Reuben, Sapienza & Zingala, 2014 for a comprehensive review). Moreover, studies indicate women in science experience gender discrimination and harassment at above-average rates (Smyth & Nosek, 2018), which has been shown to mediate academic participation and a willingness to pursue or remain in STEM careers (see Lane, Goh & Driver-Linn, 2012). A recent study of women in undergraduate STEM courses, for instance, found that a significant percentage had experienced sexual harassment (61%) and gender bias (78%) while on their course of study, and that this significantly influenced their desire to pursue a scientific career (Leaper & Starr, 2018). Given that longitudinal and cross-cultural studies find no significant or reliable gender differences in scientific/mathematical ability (see Halpern, 2007 for a review of this literature), understanding the ideological barriers to women's progress is particularly important and the current data adds to the literature in an Irish context.

It is important to note, however, that this study also found a significant women-science bias and no significant effect in either direction on the women-arts trial-type. Recent research into gender-science stereotypes suggests the increasing representation of women in science is changing beliefs about women's scientific abilities (Miller et al., 2018). Given that Ireland has implemented a range of initiatives to increase the number of women in STEM in recent years (see SFI, 2018), these data could evidence changing societal attitudes around female gender roles. Another related explanation for the women-science effects could be the makeup of the current sample. Research in this area suggests that gender-science

stereotypes vary depending on the level of education and discipline of the sample employed, with stereotypes weaker among female scientists or in cultural contexts where women are more equally represented in STEM fields (e.g., Miller et al., 2015; Nosek & Smyth, 2016). While information about participants' educational background and field of study was not collected here, the majority of participants were recruited from the Maynooth University undergraduate community which comprises both Arts/Humanities and Science programmes. Regardless of the potential reasons for these effects, the ability to separate out scores at the trial-type level allowed for this analysis of *both* women-science and men-science biases (as opposed to an overall men-science/women-arts bias, as produced by the gender-science IAT: e.g., Zingala et al., 2017). As such, these data provide a novel contribution to this literature and a useful methodology for researchers in this area.

Androcentric Bias

Across the two studies, there was an overall skew towards attributing a female over a male gender to the neutral stimulus. The first study in this chapter found no evidence of either androcentric or gynocentric bias (23 out of 44 selected "female"), while the second study had a significant pro-female response pattern (43 participants out of 59). As with the absence of infrahumanisation biases, this pattern of responses could have been due to a genuine absence of androcentric bias among participants. Because this is the first empirical assessment of androcentric bias in an Irish sample, however, this explanation would require corroboration using larger samples and alternative methods of measurement (e.g., multiple assessments of androcentric tendencies or biases). The results obtained here could also have been due to the abstract way in which androcentrism was assessed. Previous research on androcentric bias examined the influence of particular variables (e.g., sex-typed

language or masculine generics, etc.) on the likelihood of imagining a male over a female person (see Bailey et al., 2018). The current studies forced a male or female choice, and the directness of this question without any additional information may have fostered a knee-jerk female response option (as was potentially observed in the pro-female hiring preferences in Chapters One and Two). Again, this is conjecture and would require a more focused investigation in future research.

Responses may have additionally been influenced by the particular facial stimulus employed in the task. All images including the composite, gender-neutral face were selected from a dataset created by a group of Australian researchers (Rhodes et al., 2000; 2011). Although these images were piloted for gender ratings among similar samples used in the main studies (Australian adults), facial recognition has been shown to vary considerably across cultural contexts (e.g., Dailey et al., 2010; Elfenbein et al., 2007), and as such may have been interpreted differently by Irish participants. Future research may benefit from piloting the stimuli using a comparable sample, or using an alternative paradigm that does not rely on facial stimuli. For example, studies could present participants with a written description of a person and ask them to imagine if that person was female or male.

Relationship between IRAP Scores and Androcentric Bias

Generally, the relationship between androcentric bias and IRAP performance was weak. For the gender-binary IRAP, there was a significant androcentric bias/trial-type interaction effect, with scores on the men-masculine trial-type higher among those who attributed a male gender. While this effect should be interpreted cautiously due to small sample size, the number of statistical comparisons that were made, and the issues discussed above with the androcentrism task, results are somewhat coherent with theories of androcentrism. Bem (1993) and Butler (2002),

for example, both argue that androcentrism is a key feature of binary gender systems which partly maintains the patriarchal gender order. It is therefore interesting and significant that a relationship was found between these measures, although there is no existing theoretical reason the variance should be driven by the men-masculine trial-type. Future research could perhaps explore this in more detail and with a broader range of measures (e.g., other explicit measures of gender beliefs, alternative formats for assessing androcentric bias, etc.), and chase the specific trial-type effects with a larger sample.

There were no significant main or interaction effects for androcentric bias and either IRAP in study five, indicating that neither infrahumanisation or gender-science biases are related to androcentric preference. Due to the weakness of the infrahumanisation biases and the unexpected pattern of responses on this IRAP, it is likely that gender infrahumanisation is generally not a feature of normative gender beliefs (as discussed above). As such, it would not be expected to be related to other types of gender prejudice such as androcentrism. The absence of any relationship between androcentric and gender-science biases is surprising, however, given that gender-science stereotypes are known to correlate with other forms of sexist or discriminatory attitudes (e.g., Miller et al., 2015). Future research could investigate this in more detail, again potentially using a broader range of androcentrism measures and paradigms.

Role of Explicit Sexism

While there was no relationship between explicit sexism and androcentric bias in the first study, significant main and interaction effects were found in study five. MS scores were higher for those who selected the male response option, and this effect interacted with participant gender (with scores especially higher for

female participants who selected male). While the interaction between this bias, gender beliefs, and participant gender is novel in the literature, it can be explained by Ambivalent Sexism Theory (Glick & Fiske, 2000). As mentioned previously, Gaunt (2013) proposes that a tendency to consider men more uniquely human is likely to be related to a person's broader gender beliefs (rather than their specific gender group membership *per se*). That is, as both men and women are exposed to the same gender stereotype content and hegemonic gender ideology, it would not be atypical for women high in sexism to behave in a discriminatory manner. This should again be chased in future studies using a larger range of measures.

Conclusions

Results from this chapter provided a range of novel findings and contributed to the growing literature on "human-as-male" bias, which has received little empirical attention to date. While the two newly developed IRAPs produced unexpected results, each added to our broader understanding of gender beliefs in an Irish context. Specifically, these studies suggested that that gender binary beliefs may play a role in androcentric bias, and also that infrahumanisation may not in fact be a common feature of gender beliefs. These results also provided some methodological insights into the IRAP (e.g., around the significance of category labels and individual stimulus selection), and also attested to the IRAP's general utility in separating out gender biases at the trial-type level. For example, the women-science and men-science biases observed in Study Seven provided new information about the content of implicit gender-science stereotypes, and identified some areas for future investigation.

Chapter 5

Binary Biases and Sexual Harassment

Proclivity

5.1 Introduction

Chapters Three and Four demonstrated the IRAP's utility in measuring gender-as-binary biases. Across five studies, participants completed a range of gender IRAPs as well as tasks assessing their gender beliefs and propensity to prioritize or centralize men or masculinity. The current (and final) experimental chapter in this thesis will examine the role of the binary in another form of gender-based discrimination: a proclivity to engage in sexually coercive, harassing or predatory behaviours. Understanding sexual aggression has been a unifying goal of feminist research and activism for decades. Since the onset of the second wave, feminists have conceptualised sexual aggression as a cornerstone of patriarchy, and one which is fundamentally rooted in a traditional binarist and masculine ideology (Brownmiller, 1975; Dworkin, 1974; Groth & Birnbaum, 1979). It would thus be interesting and theoretically relevant to explore the relationship of binary biases to sexually aggressive behaviours, and specifically a person's proclivity towards sexual harassment.

The two studies in this chapter will examine the relationship between implicit binary biases and sexual harassment proclivity in young, heterosexual Irish men. Sexual harassment proclivity was selected as a useful starting point over other forms of sexually aggressive behaviours given that (a) it would have raised more serious ethical and methodological concerns to ask participants about their personal histories of coercive or harassing behaviours (see Strang & Peterson, 2017), and (b) sexual harassment has been more widely studied in young adult samples than other forms of sexual violence, and thus there was a broader range of validated measures to choose from (see Testa, 2015 for a review). The first study in this chapter (Study Eight) will employ the same gender binary IRAP as previous studies. This study will also

include a new IRAP for gender identity biases measuring the relative extent to which participants associate themselves with traditionally masculine or feminine traits. The identity IRAP is included as a way to examine, first, if the IRAP has utility in measuring identity biases and, second, if such biases are related to other discriminatory or harmful tendencies. Study Nine will also measure sexual harassment proclivity, but will instead examine whether it is related to implicit infrahumanisation biases. This study will use the same infrahumanisation IRAP as the previous chapter.

5.2 Study Eight

Sexual Harassment

Before reviewing the literature on sexual harassment proclivity and its measurement, it is necessary to give a brief overview of the significance and scale of the issue. The term sexual harassment was coined by feminists in the 1970s to make visible the culture of sexual coercion, exploitation, and violence experienced by women within the workforce (Brownmiller, 1999; MacKinnon, 1979; Rowe, 1973). Since then, the term has evolved somewhat and now generally refers to any "unwelcome (as opposed to involuntary) sexual advances, requests for sexual favours, and other verbal or physical conduct of a sexual nature." (UN, 2015). While few national surveys have been conducted (in Ireland or elsewhere), the most recent European statistics estimate that 55% of women have experienced it at least once since the age of 15 (Eurobarometer, 2016). Though most widely studied in organisational settings, harassment has been shown to be widespread across numerous social contexts, including the street, school and university, nightclubs and bars, and the home (for comprehensive reviews, see the Eurobarometer, 2016 report and also Paludi & Paludi, 2003; Sbraga & O'Donohue, 2000).

As with other forms of bullying and discrimination, sexual harassment is associated with an array of negative consequences. These include an increased risk of mental health problems such as depression, anxiety, or post-traumatic stress disorders, as well as frequent feelings of shame, demotivation, helplessness, and a desire to socially withdraw (see Willness, Steel & Lee, 2017 for a meta-analysis). Victims of harassment can also experience many physical health issues (mostly related to prolonged stress), including gastrointestinal problems, disrupted sleep and impaired cognitive function (Campbell, Greeson, Bybee & Raja, 2008; Leidig, 1992). In addition to the effects on the individual, sexually hostile cultures and environments have significant consequences for society at large. These include costs to organisations (e.g., because of prolonged absences and sick leave, or the loss of skilled workers due to victims leaving the organisation), as well as the costs to the state and healthcare providers (Shaw, Hegewisch & Hess, 2018). It is therefore considerably important to better understand harassment at an individual and societal level.

Attempts to explain sexual harassment have followed a similar trajectory as explanations for other forms of gender-based discrimination. Early models tended to normalise such behaviours as a natural and relatively harmless feature of male-female "mating" behaviour (see McPhail, 2016). Though this interpretation is still prevalent in evolutionary models of sexuality (Jonason et al., 2008; 2009), feminist critiques have interrogated this sort of theorizing (i.e., that harassment or assault is just over-aggressive sexual desire; see Berdahl, 2007). Instead, generally speaking they would tend to position harassment within a broader framework of gender-based violence or patriarchal power structures. According to Unger (1979) and Byers (1996), for instance, normative harassment should be conceptualization as the logical

consequence of heteronormative socialization practices, and specifically of conditioning men to be sexually virile, dominant, and assertive and women to be sexually pure, subservient, and passive. Numerous theoretical approaches now exist based on this core premise, which model harassment in terms of gendered power imbalances or abuses (Tangri & Hayes, 1997). These include social or sex role "spillover" accounts, which propose that sexual harassment is simply socio-sexual behaviour playing out in a work or social setting (Gutek, 1985; Gutek, Cohen, & Konrad, 1990; Gutek & Morasch, 1982), as well as situational or organisational power theories, which view sexual harassment as a manifestation of economic and societal male dominance (Evans, 1978; Farley, 1978; MacKinnon, 1979; Nieva & Gutek, 1981; Zalk, 1990).

While any individual instance of sexual harassment is likely the result of various personal, situational, and socio-cultural factors, research generally supports socialisation-based over essentialist explanations (McPhail, 2016; Skaine, 1996). The most persuasive evidence comes from the fact that sexual harassment is a significantly gendered problem, which appears to function in much the same way as other forms of societal gendered power abuses. For example, as with other types of discrimination based around gender-status hierarchies (e.g., workplace bias), men are significantly more likely to perpetrate harassment and women are significantly more likely to experience it (Bastian, Lancaster & Reyst, 1996; Sanders, 2008). Moreover, as would be expected in a power-based dynamic, men are more likely than women to disregard or ignore the severity of the issue. For instance, studies show that men are significantly more likely to view harassment as harmless (e.g., as flirting or "banter"; DeSouza & Solberg, 2004; Russell & Trigg, 2004), question or discount its negative effects (Quinn, 2002), and apportion blame on the victim if a case is brought (De

Judicibus & McCabe, 2001; Kenig & Ryan 1986; Jensen & Gutek 1982). Socialisation or power-based explanations can also better explain why non-binary individuals, gay or bisexual men, and women in positions of power (e.g., female managers or politicians) seem to be particular targets for sexual and gender-based harassment (Berdahl, 2007; Dubois et al., 1998; Foote & Goodman-Delahunty, 1999; McLaughlin, Uggen & Blackstone, 2012), as opposed to just sexually "viable", attractive, and traditionally feminine women. According to Franke (2007), what makes sexual harassment a gendered abuse of power is not that the conduct is sexual, but that it is being used to enforce and perpetuate binary norms. She posits, "the discriminatory wrong of sexual harassment…lies in its power as a regulatory practice that feminizes women and masculinizes men. That renders women sexual objects and men sexual subjects" (p. 691).

In addition to reinforcing binary roles more generally, harassment appears to serve a specific purpose in regulating hegemonic, traditional masculinity. According to Quinn (2002), sexual harassment is just one of the many ways in which masculinity is reproduced or "done" in a social context, and is simply part of men's broader socialised tendency to objectify and sexualise women in the public sphere. Research tends to support the idea that harassment is a form of gender performance, showing for instance that sexual harassment is significantly more likely to occur in male-dominated settings and in front of other men (e.g., by groups of men on the street, or in historically male-dominated work contexts: Berdahl, 2007; Pryor, Giedd & Williams, 1995; Stamarksi & Son Hing, 2015). Moreover, research indicates that men who more strongly endorse gender role distinctions and a traditional masculine gender identification are significantly more likely than other men to sexually harass, and indeed to report an enhanced sense of male identity after doing so (Dall'Ara &

Maass, 1999; Maass et al., 2003). Others have found that men may be particularly likely to harass if they feel their masculinity or status is under threat, such as in contexts where they have a female superior and other instances where they feel subordinate to a woman (Burgess & Borgida, 1999; Fiske & Stevens, 1993).

Measuring Sexual Harassment Proclivity

As with sexual violence more generally, most research into sexual harassment has focused on understanding the scale and nature of victimisation. Comparably few paradigms exist for studying a person's proclivity and/or history of perpetration, due in large part to the obvious difficulty identifying large enough samples of perpetrators (see Strang & Peterson, 2017 and Pina, Gannon & Saunders, 2009 for reviews of this issue). Of the research that has been conducted, most has focused on either men's tolerance of sexual harassment (i.e., the extent to which the evaluate an instance of harassment as normal or acceptable: e.g., Riley, Lott, Cadwell & De Luca, 1992; Russell & Trigg, 2004) or their self-reported tendency to behave in a hostile, predatory, or coercive manner when in a position of power (e.g., Dall'Ara & Maass, 1999; Maass et al., 2003). A small number of questionnaires have been developed for assessing proclivity, namely the Sexual Harassment Proclivities Scale (SHP: Bingham & Burleson, 1996), which measures a person's views about harassment as well as their likelihood of engaging in harassing behaviours, and the Likelihood to Sexually Harass Scale (LSH: Pryor, 1987). The LSH is by far the most widely used and assesses participants' self-reported likelihood of behaving in a harassing manner, if they could be assured they would not face any repercussions (i.e., penalties, job losses, etc.).

Studies of proclivity suggest that harassment is closely related to other problematic gender behaviours and beliefs. Studies using the LSH, for instance, have

found it correlates with more traditional and sexist gender views, authoritarian beliefs, a rejection of feminism, and an increased tolerance of interpersonal violence against women (Barak, Fisher, Belfry, & Lashambe, 1999; Bartling & Eisenman, 1993; Begany & Milburn, 2002; Malamuth & Dean, 1991; Pryor, 1987; Ward, Hudson, Johnston, & Marshall 1997). Similarly, a tolerance of sexual harassment has been shown to predict an endorsement of rape myths, sexually hostile or adversarial attitudes, and a history of sexual aggression (Begany & Milburn, 2002; Pryor, 1987; Riley, Lott, Cadwell & De Luca, 1992). More recently, researchers identified a positive correlation between harassment proclivity and the so-called "Dark Triad" traits (i.e., narcissism, psychopathy, and Machiavellianism: Ziegler-Hill, Besser, Morag & Campbell, 2016), a tendency to deceive or exploit others (Lee, Gizzarone & Ashton, 2003), and belittle women in social settings (Siebler et al., 2008). While very few studies have explored the role of implicit biases in harassment, the one study that did (Rudman & Mescher, 2012) found that implicit dehumanization biases were associated with a significantly increased likelihood of harassing. Together, these data thus support feminist socio-cultural models of rape, and more broadly the argument that harassment participates in and indeed arises from conservative, binarist, and misogynistic ideologies.

Measuring Masculine Identity

Masculine gender identity has been most commonly measured using self-report techniques such as widely used Bem Sex Role Inventory (Bem,1974), the Masculine Behaviours Scale (Snell, 2013), and the Personal Attributes Questionnaire (Spence, Helmreich & Stapp, 1973). Generally, these measures assess the extent to which a person identifies with stereotypically masculine or feminine traits, roles or abilities, and are thus assessments of how well a person adheres to a pre-defined

gender role. Newer measures have tended to combine assessments of masculine identity orientations and masculine ideologies, focusing broadly on the extent to which a person both endorses *and* adheres to traditional masculine values. These include measures such as the Masculine Gender Role Stress Scale (Eisler, 1987) and Gender Role Conflict Scale (O'Neill et al., 1986), as well as the more recently developed "hypermasculinity" inventories (e.g., the Hypermasculine Values Questionnaire: Archer, 2010 or the Conformity to Masculine Norms Inventory: Mahalik, 2005). Implicit measures have not been used to date to examine gender identity in adults, although one study did use the IAT to compare identity biases (i.e., me/not me with male/female) across cis and transgender children (Olsen, Key & Eaton, 2015).

While relatively few studies have directly examined the role of masculinity in harassment proclivity, the few that have suggest these ideologies and beliefs participate to some degree in harassing behaviours. Powell (1986), for example, found that BSRI scores significantly influenced male (but not female) participants' definitions of harassment, with higher masculinity scores associated with labelling fewer behaviours as harmful or problematic. Russell and Trigg (2004) similarly found that both a masculine identification (on the PAQ) and conservative gender ideology (measured using the Ambivalent Sexism Inventory) increased tolerance of harassment, while De Judicibus and McCabe (2004) found a positive association between masculine identification and a likelihood to assign blame to the victim over the perpetrator. Discourse analyses have also examined how harassment may function as way for young men to display their masculinity (Quinn, 2002), and also how men may rationalize harassing behaviours as normative features of male gender roles (Robinson, 2006). It is also worth mentioning the review of masculinity

ideologies and sexual aggression conducted by Murnen and colleagues in 2002. This meta-analysis found a reliable and robust role for masculinity ideologies in men's tolerance of, propensity towards, and histories of engaging in sexually violent behaviours. Though not directly focused on harassment, it is logical to extrapolate here given that many feminists view harassment as the lower end of a broader sexual violence continuum (e.g. Brownmiller, 1975; see also McPhail, 2016 for a recent review).

Current Study

This chapter aims to add to the literature on sexual harassment proclivity and examine whether this tendency is related to automatic gender biases in a sample of young Irish men. Building on the literature reviewed above, this study will specifically focus on the role of implicit binary gender ideologies and masculine identities in harassment proclivity, neither of which have been examined to date. Two IRAPs will be used for this purpose: (1) the gender binary IRAP from previous studies, and (2) a new gender identity IRAP measuring participants' associations between themselves and stereotypically feminine or masculine traits. For the binary IRAP, the same pattern of results observed in previous studies is expected here (i.e., participants are expected to relate women with feminine but not masculine traits, and men with masculine but not feminine traits). As the identity IRAP is the first implicit measure of its kind, no particular pattern of results is expected here, although it is tentatively expected that participants will associate themselves more with masculine than feminine traits. Sexual harassment proclivity will be measured using an existing validated questionnaire, and sexism will again be assessed using the Modern Sexism Scale. Concerning the relationships between measures, it is tentatively hypothesised that more binary-consistent biases and identifications will be associated with an

increased propensity to harass. A relationship is also expected between self-reported sexism and harassment.

5.2.1 Method

Participants

Fifty White Irish male undergraduate students aged between 18 and 23 participated in this study (M_{age}=21.4). All participants self-identified as heterosexual. Sampling, recruitment methods and inclusion criteria were the same as the previous study, with one additional exclusion criterion: no women could participate. The sample was restricted to males for a number of reasons: first, as the aim was to examine the potential role of gender-binarist and identification relations in sexual harassment proclivity and not to examine gender differences in propensity, it seemed prudent to keep gender constant. Second, sexual harassment is typically perpetrated by men against women (e.g., Eurobarometer, 2016) so exploring males' propensity was a more logical starting point. Third, gender identification and construction appears to vary across men and women (see Smiler, 2006) and thus self-gender relations on the IRAP could be difficult to compare. Last, given that the majority of previous research has focused on the role of masculinity in sexual aggression (e.g., Murnen et al., 2002), there is comparably less rationale for exploring the role of feminine gender identification.

Materials

Likelihood to Sexually Harass Scale (LSH)

The LSH (Pryor, 1987, 1998: Appendix 10) assesses an individual's propensity for sexual harassment and coercion. It consists of 10 separate paragraph-length vignettes in which the respondent is asked to imagine they are in a position of power (e.g., a company manager or college professor) over an attractive subordinate

female. Following each scenario, the participant is presented with three questions asking whether they would be likely to show preferential bias for such a woman. Subscale A does not specify a contingency for this preferential bias (e.g., "Would you let her carry out a project for extra credit (e.g. write a paper)?"), subscale B specifies that it is in return for sexual favours (e.g., "Assuming that you are very secure in your job and the university has always tolerated professors who make passes at students, would you offer the student a chance to earn extra credit in return for sexual favours?"), and subscale C specifies that it is in return for going on a date (e.g., "Given the same assumptions as in the question above, would you ask her to join you for dinner to discuss the possible extra credit assignments?"). Items are scored on a 5-point Likert scale ranging from 1 (*Not at all likely*) to 5 (*Very likely*), with higher scores representative of a greater propensity for harassment or exploitation. Scores for each subscale range from 10-50

Gender Binary and Gender Identity IRAPs

The gender binary IRAP was identical to the one employed in previous experiments. The gender identity IRAP employed the same target stimuli as the binary IRAP, but used "I am" and "Other men are" in place of the label stimuli (see Table 5.1). As mentioned in Chapter One, gender identity is a complex, multi-faceted, and contextual phenomenon, so the measure developed here is not intended to be fully reflective of an individual's gender identity. However, with regards to sexual harassment and aggression, research suggests that the extent which an individual identifies as masculine and, importantly*, not-feminine* may be of relevance (see Murnen et al., 2002 for a meta-analysis of the role of masculinity in sexual aggression). As the IRAP allows for a brief assessment of an individual's fluencies with confirming self-masculine and self-feminine statements, it could be considered

a measure of relative masculine–feminine identification. “Other men” was used for the opposing category rather than “Women” or the more general “Others” because research shows that gender identities are constructed relative to others of the same gender (i.e., women define themselves relative to other women, and men to other men: see Wodak, 1997).

Table 5.1

Stimuli used in the gender binary and gender identity IRAPs

Label 1: Men	**Label 2: Women**	**Target 1: Masculine traits**	**Target 2: Feminine traits**	**Rule A**	**Rule B**
Gender Binary IRAP					
Men are	Women are	Witty Charismatic Competitive Decisive	Nurturing Gentle Affectionate Sensitive	Please respond as if men have more stereotypically masculine traits and women more stereotypically feminine traits	Please respond as if women have more stereotypically masculine traits and men more stereotypically feminine traits
Gender Identity IRAP					
I am	Other men are	Witty Charismatic Competitive Decisive	Nurturing Gentle Affectionate Sensitive	Please respond as if I have more stereotypically masculine traits and other men have more stereotypically feminine traits	Please respond as if other men have more stereotypically masculine traits and I have more stereotypically feminine traits

Modern Sexism Scale (MS)

The same 10-item version of the MS scale from previous experiments was used here.

Procedure

The experimental procedures were identical to previous studies. The order of the tasks in this experiment was as follows: Likelihood to Sexually Harass Scale, two IRAPs (presented in a counter-balanced order), and Modern Sexism Scale.

Ethical Issues

Ethical issues and procedures were identical to previous studies, though particular care was taken during the debriefing to stress that no diagnoses or judgments would me made based on individual responses to specific tasks. Participants were informed that we were only interested in group-level analyses and overall comparisons between the measures.

Data processing and analysis

Data were processed in an identical manner to previous studies and participants were excluded using the same criteria. One participant was excluded from the binary (final $n = 49$) and three from the identity IRAP (final $n = 47$) on this basis.

5.2.2 Results

Self-Report Measures

Sexism scores were similar to previous studies (Table 5.2). LSH scores were somewhat varied across the three subscales: for subscale A (preferential treatment with no specified contingency), scores were skewed to the right of the scale with lower variation, suggesting moderate-to-high propensity for specialist treatment. Scores for the B (preference in exchange for sexual favours) and C (preference in exchange for a date) subscales were left-skewed and more varied, suggesting lower relative propensity for more explicitly harassing behaviours. While no normed scores or cut-off values exist for LSH scores, interpretation is guided by the

recommendations in the original Pryor articles (1987, 1988) and more recent research using this measure (e.g. Rudman & Mescher, 2012).

Table 5.2

Descriptive Statistics for the MS and three LSH subscales

	MS	**LSH_A**	**LSH_B**	**LSH_C**
Mean	24.00	29.24	17.38	21.44
SD	6.357	4.326	7.656	8.811

Note. Scores for the MS and all LSH subscales range from 10 to 50, with higher scores reflecting higher sexism and greater propensity for sexual harassment, respectively.

IRAPs

Descriptive statistics for both IRAPs can be found in Table 5.3 (and also graphed in Figures 5.1 and 5.2). Bonferroni-corrected one-sample t-tests were conducted to explore the significance of the effects produced for each IRAP trial-type against zero (Table 5.3). For the gender binary IRAP, significant biases were found on the men-masculine, men-feminine, and women-feminine trial-types, corresponding to men-*are*-masculine, men-are-*not*-feminine, and women-*are*-feminine biases respectively. For the gender identity IRAP, significant effects were found on three of four trial types and near significant effects on the fourth (other men are masculine). However, the response patterns here were somewhat unexpected – significant I-am-masculine *and* I-am-feminine biases were found, in addition to significant other men-*are not*-feminine and near-significant other men-*are*-masculine effects.

Figure 5.1 Mean IRAP trial-type scores for the gender binary IRAP

Gender Binary IRAP

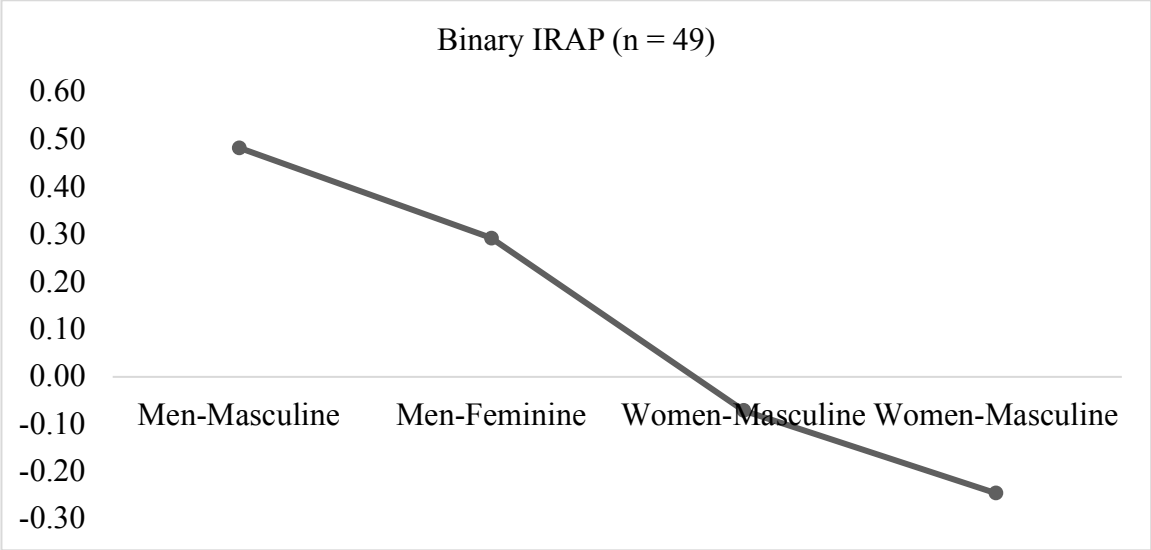


Figure 5.1. Mean scores on the gender binary IRAP. Note: positive scores reflect a "masculine" or "not-feminine" response pattern and negative a "feminine" or "not-masculine" response pattern.

Figure 5.2 Mean IRAP trial-type scores for the gender identity IRAP

Gender Identity IRAP

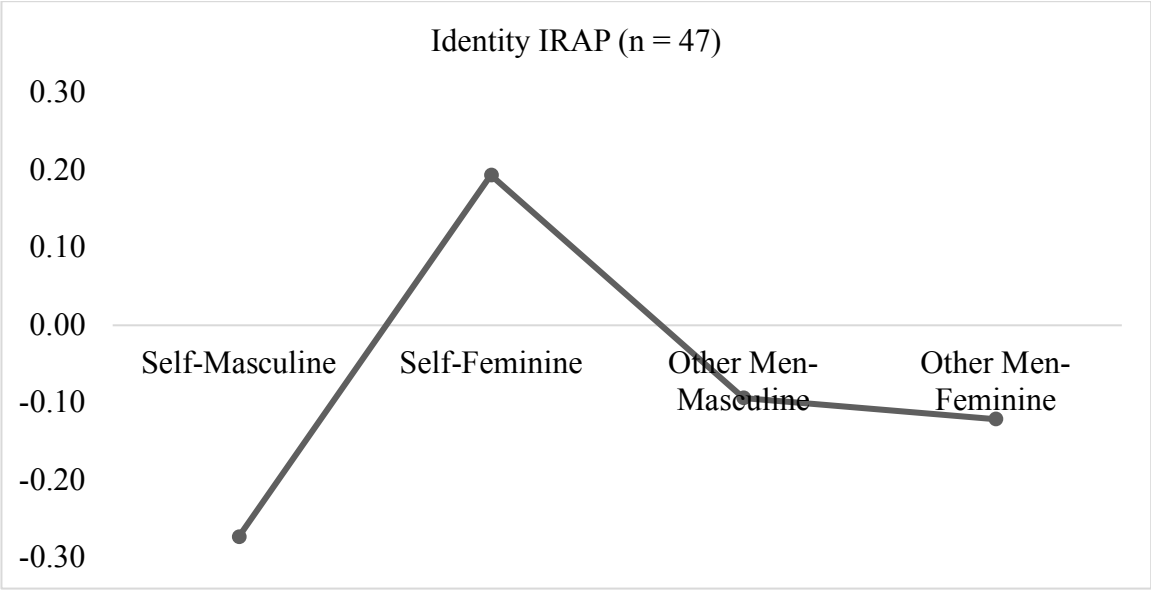


Figure 5.2. Mean scores on the gender identity IRAP. Note: positive scores reflect a "self-feminine" or "others-masculine" response pattern.

Table 5.3

Descriptive statistics and one-sample t-test results for IRAP trial-types

Gender Binary IRAP (*n* = 49)	***M***	***SD***	***t***	**df**	***p***
Men-Masculine	0.48	0.33	10.34	48.00	< .001***
Women-Feminine	0.29	0.31	6.71	48.00	< .001***
Women-Feminine	-0.07	0.35	-1.44	48.00	0.16
Women-Masculine	-0.25	0.32	-5.34	48.00	< .001***
Gender Identity IRAP (*n* = 47)	***M***	***SD***	***t***	**df**	***p***
Self-Masculine	-0.27	0.35	-5.38	46.00	< .001***
Self-Feminine	0.19	0.37	3.61	46.00	< .001***
Other men-Masculine	-0.09	0.40	-1.62	46.00	0.11
Other men-Feminine	-0.12	0.39	-2.14	46.00	0.04

Note. *$p < .05$, **$p < .01$. Acceptable p-values after Bonferroni correction </= .00625.

Measure Comparisons

Correlation analyses were conducted between all measures (Table 5.4). As in previous chapters, overall D_{IRAP} scores were used for these analyses to minimize the number of comparisons and risk of a Type 1 error. Disregarding the correlations between the LSH subscales (as would be expected), positive correlations found were between the two IRAPs ($r = .285$, $p = .046$) and between the MS and LSH_A ($r = .240$, $p = .046$) and B subscales ($r = .350$, $p = .006$). A small negative correlation was also found between the identity IRAP and the LSH C subscale ($r = -.291$, $p = .043$).

Table 5.4

Correlations between IRAPs, LSH subscales and MS

		Binary IRAP	Identity IRAP	LSH_A	LSH_B	LSH_C	MS
Binary IRAP	Pearson's r	—	0.285*	0.179	-0.065	-0.098	0.120
	p-value	—	0.046	0.109	0.672	0.747	0.206
Identity IRAP	Pearson's r		—	0.159	-0.032	-0.291 *	-0.069
	p-value		—	0.143	0.583	0.043	0.676
LSH_A	Pearson's r			—	0.448 ***	0.513 ***	0.240 *
	p-value			—	< .001	< .001	0.046
LSH_B	Pearson's r				—	0.545 ***	0.350 **
	p-value				—	< .001	0.006
LSH_C	Pearson's r					—	0.050
	p-value					—	0.365
MS	Pearson's r						—
	p-value						—

Note . all tests one-tailed. * p < .05, ** p < .01, *** p < .001, one-tailed. Positive scores on the Binary IRAP reflect a "men-masculine/women-feminine" response bias while positive scores on the Identity IRAP reflect a "self-feminine/other men-masculine" response bias. Scores on the Modern Sexism scale range from 10-50, with higher scores representing more sexist beliefs. Scores for each LSH subscale range from 10-50, with higher scores representing a greater propensity to sexually harass.

5.2.3 Summary

Significant biases were found on both IRAPs, though not always in the expected direction. For the binary IRAP, participants ascribed stereotypically masculine traits to men (but not women) and stereotypically feminine traits to women (but not men). This was similar to previous experiments. On the identity IRAP, however, masculine traits were ascribed to both the self and other men but feminine traits were only ascribed to the self. One possible interpretation for this pattern is the inherently relativistic nature of the IRAP block structure, given that participants are responding to themselves as masculine insofar as they are responding to other men as feminine. Alternatively, it could suggest greater flexibility with one's own gender than with the broader gender category as a whole. This will be discussed in more detail later in the chapter.

Concerning the relationships between measures, there was no relationship between the binary IRAP and sexual harassment proclivity (as measured by the LSH). However, there was a small negative correlation between the identity IRAP and the LSH C subscale (preferential treatment in exchange for a date). As a positive score on this IRAP reflects an "other men masculine/self-feminine feminine bias" a negative correlation in fact means harassment propensity was associated with stronger self-masculine effects. A small positive correlation was also found between the two IRAPs, suggesting gender identity may be related to the tendency to binarise gender. Again, given how the scores are interpreted, a positive correlation here means that binary-consistent gender stereotypes (i.e., men-masculine/women-feminine) were associated with a tendency to associate the self with feminine and other men with masculine traits. Lastly, as in previous research on harassment proclivity, a positive relationship was found between self-reported sexism and a

propensity to sexually harass (on two of the three subscales). Again these effects will be discussed later.

5.3 Study Nine

As mentioned previously, there is a paucity of literature concerning the relationship between implicit gender biases and harassment proclivity. While no previous studies could be identified that examined implicit binary or identity biases in this context, two studies have explored the relationship between implicit dehumanization and a propensity for harassment. Rudman and Mescher (2012) tested the theoretical alignment of gender dehumanization and violence against women across two studies using IATs. The first found that men who associated women with primitive constructs (e.g., instinct, nature) had higher scores on the LSH and also reported more negative attitudes towards female rape victims. The second study found that implicit women-animal biases (e.g., between women and words like paw, snout, animal) predicted harassment proclivity, as well as scores on a rape-behaviour analogue measure. Hussey and colleagues replicated the findings of the Rudman and Mescher study in an Irish context, though this study used IRAPs in place of IATs (Hussey et al., 2015).

The current study aims to build on this by exploring whether harassment propensity is associated with the related construct of infrahumanisation. Using the same infrahumanisation IRAP from the previous chapter, this study will explore whether assumptions of complexity in men and simplicity in women are related to sexual harassment proclivity (on the LSH). A second aim of this study is to assess whether the response pattern from Study Seven replicates. In the previous chapter, gender-infrahumanisation biases were found in the opposite direction than expected. Rather than associating men with complexity and women with simplicity,

participants associated both women and men with complex emotions, and effects were actually stronger for women than men. While some reasons for this were discussed in the last chapter (e.g., that the stimulus labels may have influenced effects, etc.) a replication will help inform the conclusions drawn.

5.3.1 Method

Participants

Forty-one White Irish male undergraduate students aged between 18 and 27 participated in this study (M_{age}=20.63). All participation was voluntary and no remuneration was offered. Sampling, recruitment and inclusion criteria were the same as previous studies.

Materials

Self-Report Scales

The versions of the LSH and MS used here were identical to the previous experiment.

Infrahumanisation IRAP

The stimuli for the Infrahumanisation IRAP were identical to Study Seven.

Procedure

The procedure was identical to previous experiments, though participants were again provided with a reference sheet outlining the stimuli belong to the categories "complex" and "simple", as in Study Seven. The order of the tasks was as follows: Likelihood to Sexually Harass Scale, IRAP, Modern Sexism Scale.

Ethical Issues

Ethical issues and procedures were identical to the previous study.

Data processing and analysis

Data were processed in an identical manner to previous studies and participants were excluded using the same criteria. Six participants were excluded from the infrahumanisation IRAP on this basis (final $n = 35$).

5.3.2 Results

Self-Report Measures

Table 5.5 presents the descriptive statistics for the MS and LSH by subscale. Scores on the MS were slightly higher than previous studies and were skewed more towards the higher end of the scale. On the LSH, scores were relatively low for the B and C subscales (preference in exchange for sexual favours or a date), but moderate for the A subscale (no contingency for preferential treatment). Scores for all subscales had similar levels of variance.

Table 5.5

Descriptive statistics for the LSH and MS

	LSH A	**LSH B**	**LSH C**	**MS**
Mean	27.35	13.15	17.43	28.05
SD	6.927	5.921	6.441	2.987

Note. Scores on each scale have a possible range of 10-50, with higher scores reflecting greater sexism/propensity for sexual harassment.

Infrahumanisation IRAP

The response pattern on the Infrahumanisation IRAP was similar to Study Seven: effects were positive for both the men-complex and women-complex trial-types (indicative of both men-*are*-complex and women-*are*-complex biases), negative for the men-simple trial-type (indicative of a men-*are*-simple bias), and very marginally positive for the women-simple trial-type (indicative of a women-*are not*-simple bias; see Table 5.6). Bonferroni-corrected one-sample t-tests were conducted to explore the significance of the effects produced for each IRAP trial-

type against zero. Of the four, only effects on the men-complex trial-type were significant.

Figure 5.3 Mean IRAP trial-type scores for the infrahumanisation IRAP

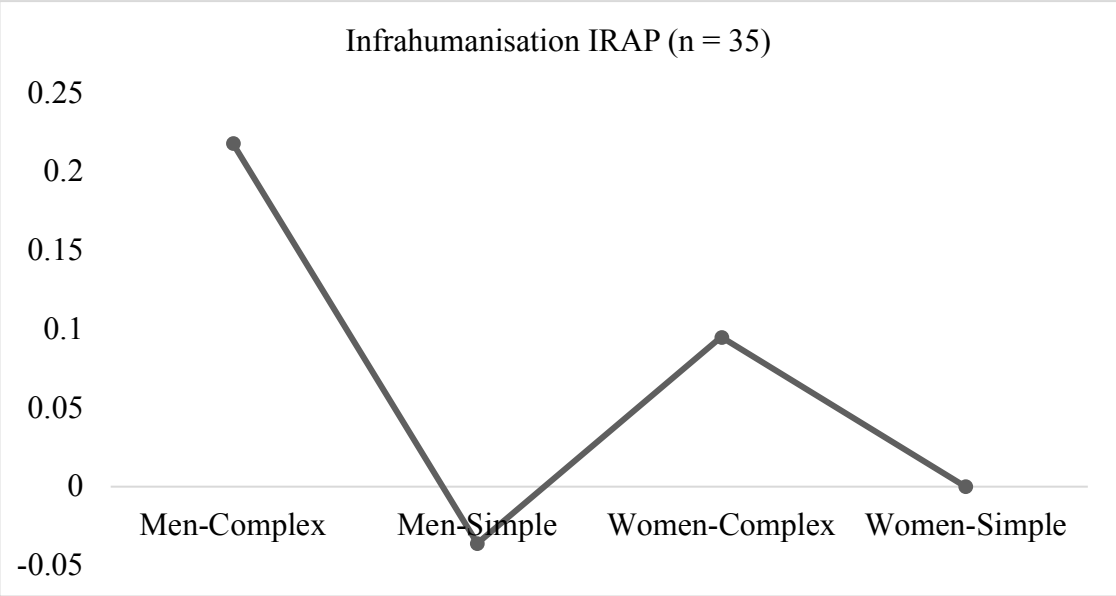


Figure 5.3. Mean scores on the infrahumanisation IRAP. Note: positive scores reflect a "complex" or "not-simple" response pattern.

Table 5.6

Descriptive statistics and one-sample t-tests for IRAP trial-types.

	M	***SD***	***t***	**df**	***p***
Men-complex	0.22	0.34	3.8	34	<.001***
Men-simple	-0.04	0.31	-.68	34	.501
Women-complex	0.10	0.31	1.82	34	.078
Women-simple	0.00	0.31	.009	34	.993

Note. *p < .05, **p < .01. Acceptable p-values after Bonferroni correction </= .0125

Measure Comparisons

Given that this study used only a single IRAP and thus fewer comparisons were being made, correlation analyses were run on the individual trial-type scores rather than overall *D* scores (Table 5.7). Again disregarding the inter-correlations

between effects on IRAP trial types or LSH subscales, significant correlations were only found between the men-complex trial type and the LSH_A subscale (r = -.575, p = <.001) and the men-complex trial and the MS (r = -.367, p = <.03). Interestingly, the direction of this relationship was negative for both of these correlations, and indeed for all involving the two male trial-types. Though these effects should be interpreted tentatively given the relatively small sample size and the number of correlations being run, this suggests that higher harassment proclivity may actually be associated with men-*not*-complex biases. This will be elaborated on in the discussion below.

Table 5.7

Correlations between the IRAP, LSH and MS

		Men-complex	Men-simple	Women-complex	Women-simple	LSH A	LSH B	LSH C	MS
Men-complex	Pearson's r	—	0.024	-0.103	-0.187	-0.575***	-0.124	-0.130	-0.367*
	p-value	—	0.892	0.558	0.282	< .001	0.476	0.457	0.030
Men-simple	Pearson's r		—	-0.507**	-0.040	-0.004	-0.100	0.064	0.100
	p-value		—	0.002	0.821	0.982	0.569	0.716	0.569
Women-complex	Pearson's r			—	0.152	0.081	0.049	0.103	-0.132
	p-value			—	0.384	0.643	0.779	0.558	0.450
Women-simple	Pearson's r				—	0.051	-0.006	-0.136	-0.010
	p-value				—	0.771	0.973	0.434	0.954
LSH_A	Pearson's r					—	0.178	0.102	0.189
	p-value					—	0.273	0.532	0.243
LSH_B	Pearson's r						—	0.649***	0.239
	p-value						—	< .001	0.138
LSH_C	Pearson's r							—	-0.028
	p-value							—	0.865
MS	Pearson's r								—
	p-value								—

* p < .05, ** p < .01, *** p < .001

5.4 Discussion

IRAP Effects

Participants in this chapter completed three IRAPs: one for gender binary biases, one for gender identity biases, and another for gender infrahumanisation biases. Effects on the binary IRAP were identical to previous chapters and so will not be reviewed again here (but see Chapter 7). For the infrahumanisation IRAP, effects were generally similar to the previous study, though the men-complex biases were more pronounced in the current sample. Significant women-complex biases were also found in this sample but were less significant than the effects on the men-complex trial type. The conclusions drawn in the last chapter thus seem appropriate: infrahumanisation does not appear to be a particularly salient or pronounced feature of gender prejudice. At least it does not appear to be a feature at the automatic level, when measured using the current methodology and specific stimuli or category labels (i.e., complex and simple). Future research could investigate these conclusions in more detail in a number of ways; these recommendations will be outlined in the General Discussion.

Gender Identity IRAP

For the gender identity IRAP, significant biases were found on three of the four trial types: self-masculine, self-feminine, and other men-masculine. While the self-feminine effects were not expected, these data do cohere somewhat with existing research on gender self-concept and identity. Previous research using trait rating scales or androgyny questionnaires suggests people may be more flexible with their own gender identity than their gender stereotypes about the typical man or women. That is, while they may rate certain traits as more stereotypically female or male, they tend to define themselves using a mix of both female and male attributes (e.g.,

Bem, 1988; Coleman & Hong, 2007; Oswald & Lindstedt; 2006; Twenge, 1997). Studies also show that identity may vary within individuals depending on the context. A repeated-measures study by Smith, Noll and Bryant (1999), for example, found that scores on the Bem Sex Role Inventory changed significantly depending on where the measure was administered (school, home, work, social interactions with same-sex friends, social interactions with other-gender friends, and situations where they did not know anyone). It is thus possible that the scores here reflect participants' own flexible identities or gender self-concepts. However, as this was the first study of its kind, this explanation would need to be substantiated with more research (potentially using a wider range of identity measures).

Effects could also be due to the specific stimulus categories used in the IRAP (i.e., "I am" versus "other men are"). Previous research using both the IRAP and IAT suggests that bias scores may change depending on the contrast categories employed (see Hussey et al., 2015). To date, the only other assessment of implicit gender identity was carried out with children and used the categories "Me" and "Not me" with male and female words (Smith et al., 1999). These categories were inappropriate for the current research for two reasons: first, the version of the IRAP used in this thesis employs natural language statements and responding to "*not me*-masculine-true/false" would not have made sense in this context. Second, as mentioned, the literature on identity constructions suggests that identity is relational – both to the "opposite" gender (women in this case) and to others within the category (other men: e.g., Bem, 1993). While this study assessed a person's automatic identity biases relative to other men, it's possible that the contrast category influenced results given that participants were required to rate themselves as feminine insofar as other men were masculine and vice versa. As such, it could have

unintentionally functioned as a simultaneous assessment of gender stereotypes (for other men) and gender identity. This could explain the significant self-feminine biases observed here, though future research would need to examine this more directly (e.g., by comparing effects across IRAPs using different contrast categories). Another way to circumvent this issue could be to use an implicit measure that does not require a contrast category, such as the Single Category IAT (SC-IAT: Karpinski & Steinman, 2006). This would allow for an assessment of a person's relative masculine-feminine identification but would not require it to be measured against a specific category.

Relationship between the Binary and Identity IRAPs

A small positive correlation was found between the gender binary and identity IRAPs in Study Eight. While previous research using self-reports has found a relationship between gender stereotypes and self-concept (e.g., Olsen et al., 2015), typically, these have found that more traditional stereotypes are associated with a more pronounced gender identity (i.e., more masculine identity in men, and more feminine identity in women). In the current study, however, stereotype-consistent gender binary biases were associated with a tendency for men to associate themselves with feminine traits and other men with masculine traits. Again, this could have been due to the specific stimulus categories used, or it could be because of the specific identity and stereotype beliefs in the current sample. It is difficult to draw concrete conclusions about this finding in the absence of other information, but future research could assess this by including a broader range of assessments of identity and gender-related beliefs.

Sexual Harassment Proclivity

Mean scores on the three LSH subscales were similar to previous studies of young adult samples in both a North American (e.g., Lee et al., 2003; Rudman & Mescher, 2012) and Irish context (Hussey et al., 2015). Effects were highest for the A subscale (preferential treatment), then the C subscale (preferential treatment in exchange for a date), and lastly the B subscale (preferential treatment in exchange for sexual favours). While the only scores that could be considered mid-to-high across the two studies was the A subscale, all of the subscales included in the LSH have been found to correlate with discriminatory or conservative gender beliefs (see Pryor et al., 1995). That said, it is encouraging and worth noting that the means for the other more severe forms of *quid-pro-quo* harassment (i.e., harassment in exchange for dating or sexual favours) were low.

Relationship between IRAP Scores and Sexual Harassment Proclivity

While no relationship was found between the binary IRAP and harassment proclivity, a small correlation was found between the identity IRAP and the LSH_C subscale. Specifically, these analyses found a self-masculine/other men-feminine bias was associated with an increased tendency to harass. This finding coheres with existing theoretical and empirical research in this area suggesting masculinity plays a role in sexual aggression (e.g., Murnen et al., 2002; Quinn, 2002), and more generally with performative theories of masculinity (e.g., Connell, 2005; Smiler et al., 2015). It also provides support for the IRAP as a measure, and indeed for analysing self-masculinity biases at the level of automatic cognition. Given the number of analyses that were run and the size of the correlation, however, this effect should only be considered a tentative and preliminary finding.

In Study Nine, a significant correlation was found between LSH scores and effects on the men-complex trial type. Specifically, men-*not*-complex biases were

associated here with an increased propensity to engage in coercion or harassment. The two previous studies to date which explored the relationship between harassment and implicit gender biases found a positive relationship between LSH and dehumanisation scores (Hussey et al., 2015; Rudman & Mescher, 2012). It is thus interesting that a relationship was found here, although again the effect was in the opposite direction than expected. However, as discussed in the previous chapter, it is possible that "complex" has more feminine than masculine connotations. As such, the infrahumanisation IRAP may still be tapping into stereotypes, though in a different way. This interpretation would require substantiation in a larger sample and using a broader range of gender stereotype measures and assessments.

Conclusions

The current chapter was an assessment of automatic gender binary, identity, and infrahumanisation biases and their relationship to sexual harassment proclivity. It provided some preliminary evidence to suggest that masculine gender identities play a role in harassment, in addition to counter-intuitive men-*not*-complex assumptions. Somewhat contrary to feminist theorising and literature, gender-as-binary beliefs were unrelated to a propensity for sexual aggression (at least in the current sample and using the current methodology). More generally, studies in this chapter informed the conclusions drawn about the gender binary and infrahumanisation IRAPs, and provided some novel insights into the use of the IRAP for measuring gender identity.

Chapter 6

Pooled Analyses

6.1 Introduction

As mentioned in the General Introduction, IRAP studies typically have a sample size of between 30-50 (Vahey et al., 2015), and one of the aims of this thesis was to generate a comparably larger IRAP dataset by pooling the gender binary IRAP from different studies. While this combined data set notionally allows for several theoretically meaningful analyses to be carried out (e.g., investigations comparing different IRAP scoring metrics and algorithms, explorations of response distributions, block/IRAP order effects, and so on), this chapter will restrict its focus to a small number of questions relevant to the current thesis. The first will be a pooled analysis of the gender binary IRAP dataset. The enhanced statistical power from pooling the data should better inform the conclusions drawn about this thesis, specifically around the significance of the trial type effects or gender differences. As the Modern Sexism Scale was administered in all studies, this analysis will also inform conclusions about the relationship between self-reported sexism and automatic binary biases.

The second analysis will explore the IRAP's measurement invariance across female and male participants. Measurement invariance (also called measurement equivalence) broadly refers to a test's ability to measure the same underlying construct (e.g., gender bias) across different groups or different times (Putnick, 2016). A measure is considered invariant when it is interpreted and responded to in the same way regardless of the time or population being sampled (e.g., men, women, older adults, groups from different cultural backgrounds, etc.). This investigation was considered appropriate as it would provide more insight into gender differences above and beyond a pooled ANOVA. Specifically, this analysis would inform arguments made earlier in this thesis that men and women may navigate the IRAP

differently, potentially due to the different interpretations of the stimulus categories and the "desirable masculine" category in particular.

6.2 Analysis One:

Pooled Analysis of gender binary IRAP data

Data and Inclusion Criteria

Gender binary IRAP data were pooled from Studies Three, Four, Five, Six and Eight, leading to a combined data set of 228 after exclusions (i.e., based on failing one or more pairs of practice blocks). While the binary IRAP from Study Three was slightly different (it used the words "men" and "women" while the others used "men are" and "women are"), this was not considered so significant a difference that it merited exclusion. Analysis One will involve tests for IRAP effects at the trial-type level and gender differences, in addition to a correlation analysis with the MS. As the sample size has increased, these correlations will be run on the individual trial-types as well as the overall D_{IRAP} scores. Given that each experiment included a different domain-specific assessment (i.e., the hiring task, androcentric bias task, or LSH) and this was always administered before the IRAPs, this analysis will also explore any differences based on study.

6.2.1 Results

Trial Type Effects

As evidenced by the data in Table 6.1 and Figure 6.1, the direction and magnitude of trial-type effects did not change notably from the individual experiments: participants associated men with stereotypically masculine traits and women with stereotypically feminine traits. There was a bias against associating men with feminine traits, while effects on the women-masculine trial type were very close to zero. Effects were more pronounced (i.e., were larger in a more binary-consistent

direction) for male than female participants on all trial types except the women-feminine trial type. This suggests an own-group bias of sorts on the role-congruent trial types related to participants' own gender (i.e., the men-masculine trial types for male participants and the women-feminine trial types for female participants).

Table 6.1

Descriptive statistics for binary IRAP trial types by participant gender

Trial Type	Gender	Mean	SD	N
Men-masculine	Female	0.22	0.37	101
	Male	0.44	0.38	128
Men-feminine	Female	0.07	0.33	101
	Male	0.22	0.35	128
Women-masculine	Female	7.228e -4	0.33	101
	Male	-0.05	0.36	128
Women-feminine	Female	-0.28	0.35	101
	Male	-0.18	0.35	128

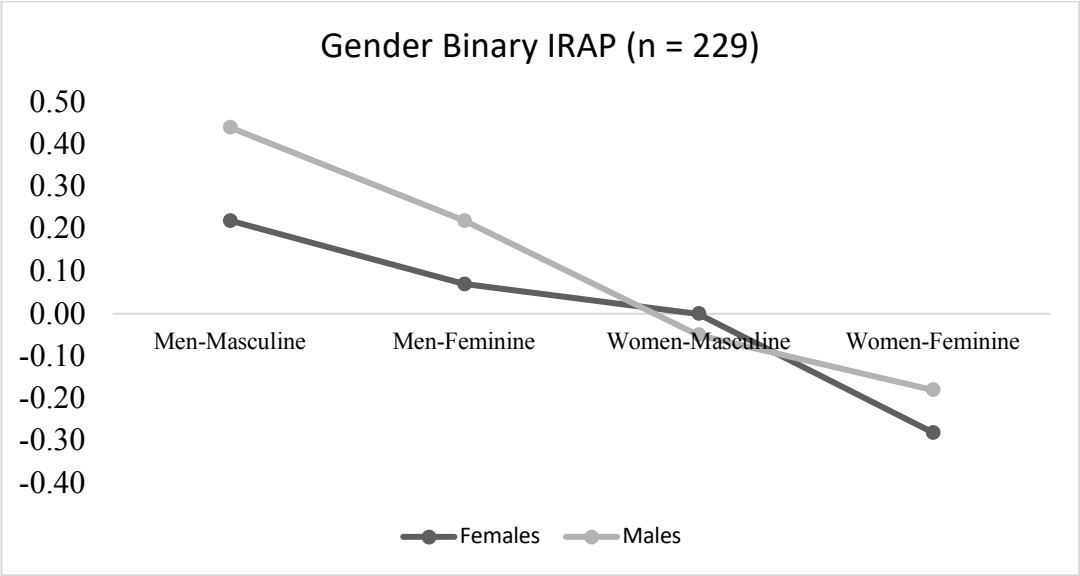


Figure 6.1. Mean D_{IRAP} scores by trial type.

One sample t-tests were run on the pooled data set to examine the significance of these effects at the trial type level (Table 6.2). For consistency and comparison, these analyses were run on the entire sample rather than split by gender.

The effects followed an identical pattern to the individual studies: significant effects were found on the men-masculine, women-feminine and men-feminine trial types. These correspond to men-*are*-masculine, women-*are*-feminine, men-*are not*-feminine biases respectively. No significant effects were found for the women-masculine trial type, indicating there no particular bias in either a role-congruent or incongruent direction.

Table 6.2

One sample t-test results

Trial Type	***t***	**df**	***p***	**Cohen's d**
Men-masculine	13.365	228	< .001	0.883
Men-feminine	6.763	228	< .001	0.447
Women-masculine	-1.189	228	0.236	-0.079
Women-feminine	-9.621	228	< .001	-0.636

Gender Differences

A mixed between-within 2x4 ANOVA was carried out to investigate gender differences in the overall sample. This analysis found significant differences across male and female participants, F (1, 227) = 22.86, p <.001, with a medium effect size (η^2 = .09). Follow-up independent samples t-tests were conducted to identify the specific trial-types driving this effect. After a Bonferroni correction was applied (bringing the acceptable p value cut-off to .0125), significant differences were found on the men trial types only (i.e., men-masculine and men-feminine) with male participants displaying significantly larger men-masculine and men-*not*-feminine biases (see Table 6.3). Effect sizes (calculated using Cohen's d) for both were in the medium range.

Table 6.3

Independent Samples T-Test results comparing trial-type scores by gender

	t	**df**	***p***	**Cohen's d**
Men-masculine	-4.38	227.0	< .001	-0.58
Men-feminine	-3.38	227.0	< .001	-0.45
Women-masculine	1.09	227.0	0.279	0.14
Women-feminine	-2.08	227.0	0.039	-0.28

Study Differences

A separate 4x5 ANOVA was run to investigate any differences based on study; that is, any differences potentially caused by the preceding domain-specific task (i.e., Likelihood to Sexually Harass scale, androcentric bias task, or hiring preference task). This analysis found a significant difference based on study, F (4, 224) = 3.049, p = .018, with a small-to-medium effect size (η^2 = .052). A significant study-trial type interaction effect was also found here, F (12, 672) = 1.196, p = .022, though the effect size was small (η^2 = .052). As evidenced by Table 6.4, means were highest on all trial types for Studies Six (using the androcentric bias task) and Eight (using the Likelihood to Sexually Harass scale). Means for the other studies did not vary reliably across studies (see Table 6.4). While these results will be discussed in more detail in the next chapter, they suggest that IRAP effects were at least partially influenced by the preceding domain-specific tasks. However, is important to note that effects likely interact with participant gender, given that study Eight used male participants only. Analysing by gender and conducting follow-up tests would not be appropriate here as it would (a) require removing Study Eight from the analysis, and (b) conducting multiple 2x5 follow up ANOVAs per trial type. This would substantially reduce the statistical power of the analysis and increase the risk of a Type 1 error. As such, it is sufficient at this stage to descriptively note that there are

differences based on study type, and that these should be reviewed later and followed up in future studies.

Table 6.4

Descriptive statistics for IRAP trial types by study number

Trial type	Study	Mean	SD	N
Men-masculine	Eight	0.48	0.33	49
	Five	0.40	0.33	37
	Four	0.30	0.39	60
	Six	0.31	0.44	39
	Three	0.24	0.42	44
Men-feminine	Eight	0.29	0.31	49
	Five	0.12	0.32	37
	Four	0.14	0.39	60
	Six	0.18	0.34	39
	Three	0.04	0.33	44
Women-masculine	Eight	-0.07	0.35	49
	Five	-0.01	0.32	37
	Four	0.03	0.31	60
	Six	-0.09	0.35	39
	Three	-0.01	0.40	44
Women-feminine	Eight	-0.25	0.32	49
	Five	-0.11	0.29	37
	Four	-0.22	0.36	60
	Six	-0.31	0.36	39
	Three	-0.21	0.40	44

Relationship between IRAP scores and Modern Sexism

Spearman's rho correlations were conducted to investigate the relationship between IRAP performance and self-reported sexism. The non-parametric alternative was used here because MS scores were not normally distributed. These analyses found no significant relationship between any of the four trial types and MS scores (see Table 6.5).

Table 6.5

Correlations between the MS and four binary IRAP trial types

		Men-masculine	Men-feminine	Women-masculine	Women-feminine	Modern Sexism Scale
Men-masculine	Spearman's rho	—				
	p-value	—				
Men-feminine	Spearman's rho	0.332***	—			
	p-value	< .001	—			
Women-masculine	Spearman's rho	-0.132*	-0.215**	—		
	p-value	0.046	0.001	—		
Women-feminine	Spearman's rho	-0.181**	-0.171**	0.276***	—	
	p-value	0.006	0.009	< .001	—	
Modern Sexism Scale	Spearman's rho	0.019	0.084	-0.067	0.004	—
	p-value	0.777	0.205	0.311	0.958	—

* p < .05, ** p < .01, *** p < .001

6.3 Analysis Two:

Assessing the IRAP's Measurement Invariance Between Female and Male Participants

Analysis Two aims to apply measurement invariance to the pooled IRAP data set to further inform the interpretation of IRAP effects and, specifically, the observed gender differences. As mentioned, measurement invariance refers to a measure's ability to assess the same underlying construct in the same way across different populations or contexts. Measurement invariance has most frequently been applied to questionnaire or other self-report data, usually to examine whether validated scales are appropriate for a specific population or investigate whether a scale may be used in a repeated-measures or longitudinal design (see Flake, Pek, & Hehman, 2017). While measurement invariance has not been examined in the context of implicit measures to date, such an approach may provide useful and important insights into both the IRAP and its interpretation. For example, the IRAP could fail to meet measurement invariance if certain stimulus categories had different meanings or salience for men versus women (e.g., due to differences in socialisation). This could mean men and women approach the completion of the task in an importantly different way, differentially pay attention to some trial types more than others, or some other unforeseen ways.

It is important to note that passing or failing measurement invariance would not undermine the observation that men and women differ in their IRAP performances (as seen in previous studies and in the pooled analysis); rather it helps guide the interpretation of these differences. For example, it would elucidate whether these effects are due to differences in a latent variable that is being assessed in the same way between the groups (e.g., differential automatic binary biases between

groups), or differences in the way men and women experience, interpret, or produce behaviour within the IRAP (e.g., differential meanings or salience of words, or how differences in how attention directed within the task).

Component Tests

Testing for measurement invariance is done using a Confirmatory Factor Analysis (CFA) framework. CFA is an extension of regression modelling, where one assesses whether multiple observed indicator variables (often individual self-report items, or IRAP trial types in this case) can be said to collectively measure an indirectly observed or latent variable (e.g., Implicit Gender Bias). Measurement invariance involves fitting increasingly constrained models to the same data and checking goodness of fit, where each level of constraint represents the test of a specific assumption. It typically involves three component tests: configural invariance, metric invariance, and scalar invariance (Putnick & Bornstein, 2016; van de Schoot, Lugtig, & Hox, 2012; Vandenberg & Lance, 2000). These tests are hierarchical: if one test is failed the next test does not need to be run, as measurement invariance has not been met. The nature of these three tests will be described first, followed by a description of the specific statistics used to assess model within these tests.

Configural invariance involves fitting the same measurement model to each group that is being tested for measurement invariance. In this case, this means fitting separate CFA models to the men and women subsamples. The base model assumes a single latent variable (e.g., "Implicit Gender Bias"), and that this latent variable is formed by four observed variables, the IRAP trial types (i.e., men-masculine, men-feminine, women-masculine, and women-feminine). This model was selected because it formalizes that which the analyses in the previous chapters implicitly rely

on (e.g., by examining differences between the IRAP trial types between genders, and by correlating the overall D_{IRAP} score with other variables). Configural invariance therefore tests the assumption that the same base measurement model can be said to fit in both samples. Failure to find configural invariance can be interpreted loosely as the possibility that men and women interpret the IRAP differently as they complete it (e.g., attribute differential meanings or salience to words, follow its instructions differently, or pay attention to different features of the task).

Testing for metric invariance involves fitting a more constrained model to the whole dataset. Specifically, it tests the additional assumption that each indicator (in this case trial type) is an equally good measure of the latent variable, and therefore that no trial types (or items were this a self-report measure) are redundant in driving inter-individual or intergroup differences. Technically speaking, this is done by forcing the factor loadings for all indicators to be equivalent. Failure to find metric invariance can be interpreted loosely as the possibility that the IRAP trial types are a better measure of Implicit Gender Bias in one gender than the other. That is, any differences in IRAP scores may be due to differences in quality of measurement rather than differences in the underlying thing being measured.

Finally, testing for scalar invariance involves fitting an even more constrained model to the whole dataset. Specifically, it tests the additional assumption that the indicators link the observed and latent variables in the same way. For example, for all indicators, whether a D_{IRAP} score of 0.30 on two different trial-types can be said to represent a comparable degree of bias, and therefore allow for a common interpretation. This is done by forcing both the factor loadings and also the intercepts to be equivalent. Failure to find scalar invariance can be interpreted loosely as the possibility that the IRAP trial types are interpreted on a different scale

between the groups. That is, the same IRAP D_{IRAP} scores in men versus women may refer to different levels of underlying Implicit Gender Bias. This is another form of differences in quality of measurement rather than differences in the underlying thing being measured.

Should all tests be passed, measurement invariance is said to be met. More technically, this means that correlations between the measure and with external variables can be interpreted in a comparable manner between the invariant subgroups.

Assessment of Model Fit

Unlike null hypothesis significance testing, which relies on a narrow range of statistics (e.g., *p* values) and cut-off values (e.g., alpha = 0.05) to make conclusions, the fit of a CFA model to the data can be assessed using a wide variety of both different statistics and cut-off values. This has led to great degree of heterogeneity in the methods used to assess measurement invariance. Contemporary recommendations are that multiple fit metrics should be reported for completeness, and a subset of these should be employed for decision-making (see Putnick & Bornstein, 2016; Vandenberg & Lance, 2000 for reviews and recommendations). The following indices were therefore calculated and reported: measures of absolute fit: Chi squared tests and Root Mean Square of the Residual (RMSR); measure of relative fit: the Tucker Lewis Fit Index (TLI); and noncentrality indices: Comparative Fit Index (CLI) and the Root Mean Square Error of Approximation (RMSEA and its 95% CIs). For the sake of decision making, metrics and cut-off criteria were selected on the basis of the recommendations of commonly cited simulation studies (Chen, 2007; see also Hu & Bentler, 1999). Specifically, configural fit was assessed using RMSEA ≤ 0.06 and CFI $\geq .95$. Metric and scalar

invariance were assessed by assessing the change-in-fit between models (i.e., between configural and metric, metric and scalar) meeting both ΔCFI > -.01 and ΔRMSEA < .015. This use of a two-metric decision making strategy is consistent with Hu and Bentler's (1999) recommendations on minimizing combined false positive and false negative rates, and is now the current modal reporting practice according to a recent review (Putnick & Bornstein, 2016).

6.3.1 Results

As measurement invariance cannot be conducted in JASP (which was used for all previous analyses), it was conducted using R and specifically the packages lavaan and semTools (Jorgensen et al., 2018; Rosseel, 2012). Configural, metric, and scalar tests were fit to the data using the parameter constraints described above. Results of each test can be found in Table 6.6, and the CFA configural model plots for the men and women subsamples can be found in Figure 6.2. Results demonstrated that the configural model found good fit in female participants but bad fit in male participants. As such, the IRAP in the current sample was found not to meet measurement invariance. While there is technically no need to run the tests for metric and scalar invariance, they are reported for completeness.

Figure 6.2. Plot of the CFA configural models fit within male and female participants.

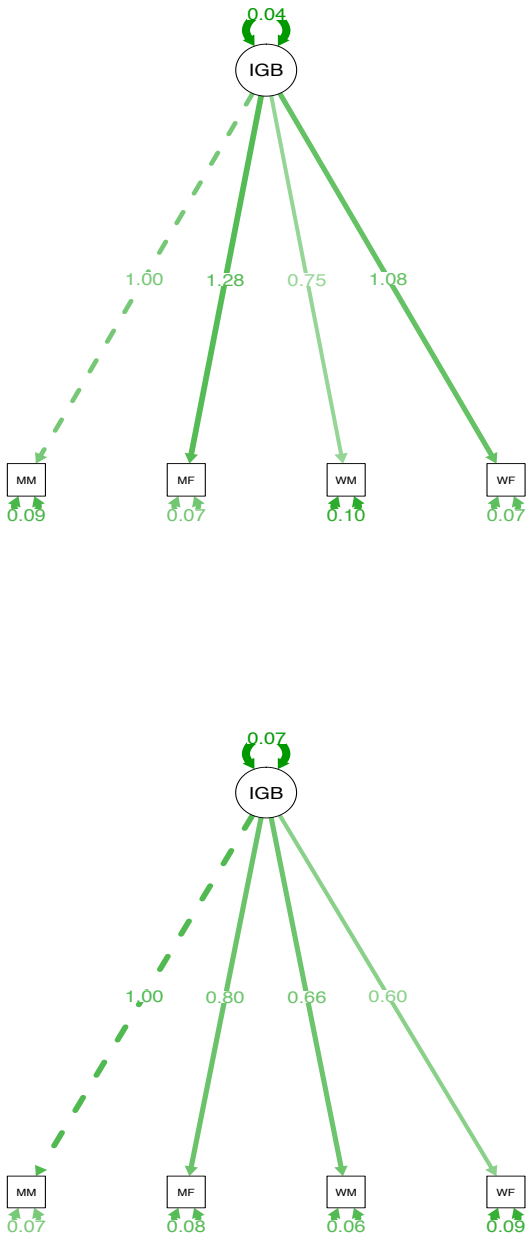


Figure 6.2. Plot of the CFA configural models fit within male (upper panel) and female participants (lower panel). IGB refers to the latent variable (Implicit Gender Bias) and squares refer to the IRAP trial types. Lines between them refer to the relative factor loadings. Curved lines refer to the associated error terms. Differences in factor loads can be seen between the panels, which contribute to measurement invariance between the samples

Table 6.6

Results of the measurement invariance tests.

										95% CI						
Test	Subset	Parameters	χ^2	df	*p*	SRMR	TLI	CFI	RMSEA	lower	upper	ΔSRMR	ΔTLI	ΔCFI	ΔRMSEA	Decision
Configural	Men	8	5.73	2	.057	0.05	0.78	0.93	0.13	0.00	0.27	-	-	-	-	Fail
	Women	8	0.16	2	.921	0.01	1.16	1.00	0.00	0.00	0.08	-	-	-	-	Pass
	All	24	5.90	4	.207	0.03	0.93	0.98	0.07	0.00	0.19	-	-	-	-	-
Metric	All	21	8.64	7	.280	0.05	0.97	0.98	0.05	0.00	0.14	0.02	0.03	0.00	-0.02	Pass
Scalar	All	18	27.72	10	.002	0.09	0.75	0.79	0.14	0.08	0.20	0.04	-0.22	-0.19	0.09	Fail

Notes: Good configural fit refers to meeting both of Root Mean Square Error of Approximation (RMSEA) ≤ 0.06 and Comparative Fit Index (CFI) ≥ 0.95; good metric and scalar fit refers to meeting ΔRMSEA < .015 and ΔCFI > -0.01. The configural model using all participants is included only to calculate change in fit metrics for metric and scalar invariance: no decision making is done on the basis of this model.
SRMR = Standardised Root Mean Square residual. TLI = Tucker Lewis Index.

6.4 Results Summaries

Pooled Analyses

Analysis One substantially informed the findings and interpretations of the current thesis. First, it demonstrated that automatic binary biases were indeed large and significant for the overall sample. In so doing, this analysis attested to the IRAP's utility in measuring the binary at the level of automatic cognition. Second, it shed light on the inconsistent gender differences observed throughout this thesis and also the specific trial types driving these effects. Third, the pooled analysis showed that the IRAP may be susceptible to seemingly extraneous influences within specific studies (i.e., the tasks presented prior to completing the IRAP). Lastly, it demonstrated that automatic binary biases were unrelated to self-reported sexist beliefs, or at least those captured on the Modern Sexism Scale. Each of these findings will be discussed in the next chapter.

Measurement Invariance

The absence of measurement invariance suggests that differences between men and women on the observed variable (the D_{IRAP} scores) should not be interpreted as reflecting differences in an overarching latent variable (Implicit Gender Bias). Instead, differences in IRAP scores might be attributable to one or more of several things, for example differences in the meaning of the word stimuli employed within the task, differential salience of/attention to the stimuli within the task, differences in strategies to complete the task, or differences in how they interpret or follow the instructions. Some evidence for differential meaning or salience can be found in the differential factor loadings between men and women (see Figure 6.2). Specifically, the strongest factor load for men was the men-feminine trial type. This indicates that inter-individual differences in responding on this particular trial type were driving the IRAP effect for men. In contrast, the

strongest factor load for women was the men-masculine trial type. This indicates that inter-individual differences in responding on this particular trial type were driving the IRAP effect for women. This could potentially be the result of the fact that women rated these traits as significantly less masculine than men did, though this will be reviewed in the next chapter. The absence of measurement invariance also has important implications for the broader use of the IRAP, and indeed other popular implicit measures. This will again be discussed in the next chapter, but in brief it suggests that the effects captured by the IRAP may not be generalisable across populations and may be more susceptible than previously thought to the effects of stimulus categories or exemplars.

Chapter 7

General Discussion

7.1 Summary of the Findings from this Thesis

The central goal of this thesis was to examine and measure the social construction of gender-as-binary. It aimed to add to the growing literature into binarist beliefs and attitudes and, specifically, explore whether such biases would be readily demonstrated on an implicit measure. Using the IRAP methodology, this work examined the extent to which young Irish adults differentially associated women and men with distinct sets of traits, attributes, and abilities. Studies in this thesis primarily looked at the binarization of gender according to traditionally feminine and masculine attributes, but a small number of other gender-related concepts were also assessed (specifically competency, uniquely human emotions, scientific ability, and identity). A second aim of this thesis was to assess the role of the binary in inequality. To this end, each study included a measure of self-reported sexism (the Modern Sexism Scale) as well as one other measure relating to either gendered hiring preferences, androcentric bias, or sexual harassment proclivity. A final aim of this work was to gather a larger-than-average IRAP dataset for a pooled analysis. This was done to both inform the conclusions drawn throughout this thesis (i.e., about the effects and the IRAP's utility in this context), and also to allow for a more detailed investigation of gender differences. Accordingly, across five studies, over 229 participants completed the gender binary IRAP. This is the largest IRAP data set in the literature to date. Before discussing the significance of the results in light of the above aims, the findings from each chapter will each be summarised in turn.

Chapter Two

The aim of Chapter Two was to identify gender stimuli for use in different IRAPs and measures throughout this thesis. Generally speaking, the field of implicit testing has struggled with stimulus selection, despite the large body of research

showing how stimuli may influence and/or confound results (e.g., Bluemke & Friese, 2004; Gast & Rothermund, 2010; Mitchell et al., 2003). No standardised selection guidelines or criteria exist in the literature, and the vast majority of studies have used ad hoc methods when choosing stimulus categories, labels, and exemplars (see Steffens et al., 2008). Therefore, to ensure the stimulus categories in this thesis were adequately clear, familiar, and representative of the overarching constructs (i.e., masculinity and femininity), the first two focused entirely on selecting exemplars for the core masculinity and femininity stimulus categories. Across two separate studies (one pilot and one larger replication) a sample of participants rated 60 personality traits in terms of their gender and their desirability ratings in women and men. Once this dataset of trait ratings was generated, four distinct stimulus categories were selected by a team of researchers: "Desirable Feminine"; "Desirable Masculine"; "Undesirable Feminine"; and "Undesirable Masculine". Analyses showed these were (a) sufficiently different to one another in both valence and gender, and (b) significantly gendered (i.e., the masculine categories were more masculine than feminine, and vice versa). While male and female participants did not differ in their ratings for most of the categories, they did differ significantly in the "Desirable Masculine" category, with women evaluating these traits as more gender-neutral than men. This was noted as an interesting finding, but also important caveat for future IRAP analyses (discussed in more detail later).

A second and more general aim of this chapter was to examine explicit gender trait ratings in a contemporary Irish sample. Across the two studies, ratings were consistent with existing literature into gender trait stereotypes: masculinity was associated with independence, assertiveness, and aggression (i.e., traits associated with agency) and femininity was associated with nurture, care, and helplessness (i.e., traits related to communion: Broverman, Vogel, Broverman, Clarkson, &

Rosenkrantz, 1994; Eagly & Karau, 2002; Newport, 2001; Gaucher, Friesen, & Kay, 2011; Madera, Hebl, & Martin, 2009; Moscatelli, Ellemers, Menegatti, & Rubini, 2016; Rudman, Moss-Racusin, Phelan, & Nauts, 2012; Spence & Buckner, 2000; Williams & Best, 1990). As has also been found in previous research, results showed a gender-valence interaction (or a "women-are-wonderful" effect) with the feminine traits considerably more positively valenced than the masculine (Broverman et al., 1994; Eagly & Karau, 2002; Newport, 2001; Rudman et al., 2012; Spence & Buckner, 2000). One unexpected and novel finding, however, was the asymmetry in the number of highly masculine and highly feminine traits in the dataset. While several traits were rated as highly feminine, only a handful had a mean rating towards the low, masculine end of the scale. Though not hypothesised, this finding coheres with feminist arguments around the androcentric structure of the binary, and specifically the tendency to equate masculine and human stereotypes and/or or evaluate femininity as the more gender-specific "other" under patriarchy (Bem, 1993; Hegarty et al., 2018; Hyde, 1984). Put simply, these data provided evidence that femininity may be more gendered than masculinity, at least based on the traits included here.

Chapter Two lastly sought to examine the implicit societal overlap between masculinity and competency stereotypes by assessing participants' hypothetical hiring preferences between men, women, stereotypically masculine, and stereotypically feminine individuals. To this end, a brief hiring preference measure was developed and piloted that employed the same masculine and feminine stimulus categories outlined above. Results were broadly consistent with existing research into occupational gender bias (e.g., Carli et al., 1995; Ealy et al., 2000; Eagly & Wood, 2016; Heilman et al., 1989, 1995; Phelan & Rudman, 2010): While the majority of participants did not express a preference for hiring a man over a woman,

nearly 90% of the sample elected to hire the witty, decisive, charismatic, and assertive individual. This person was not explicitly referred to as agentic or stereotypically masculine; however, as mentioned, analyses showed these traits were rated as significantly more masculine than feminine. As such, these data provide clear evidence for a link between men, masculinity, agency, and occupational success in an Irish context, which had not been examined to date.

Chapter Three

Chapter Three expanded on these findings by examining, first, the automaticity of binarist men-masculine and women-feminine associations, and second, the relationship between these biases and occupational preferences. While previous research had used the IAT to measure overall men-masculine/women-feminine biases (or men-agentic/women-communal biases: e.g., Rudman & Kilianski, 2000), this chapter represented the first to use the IRAP in this domain, and thus the first to separate automatic gender biases at the trial type level. Using the stimuli obtained in the previous chapter, Studies Three, Four and Five investigated the strength of both role-congruent and role-incongruent biases (i.e., the extent to which *both* women and men are automatically associated with *both* masculine and feminine traits). Response patterns were the same across all three studies, with significant effects found on the men-masculine, women-feminine, and men-feminine trial types. These corresponded to men-*are*-masculine, women-*are*-feminine, and men-*are not*-feminine biases respectively. No significant biases were found on the women-masculine trial type in either direction. These data thus cohere with existing research showing that women and men tend to be differentially associated with feminine-communal and masculine-agentic traits (e.g., Carnes et al., 2015; Rudman & Kilianski, 200; Rudman & Glick, 2001). However, they also provided novel insights into the relevance and strength of role-incongruent biases to the construction

of gender. Specifically, while these data suggest that women are defined more by their role-congruent attributes (i.e., their femininity), men seem almost equally defined by their masculinity and indeed their "not-femininity". These effects were replicated several times throughout this thesis, and will be discussed in more detail later.

In addition to the core gender binary IRAP, Chapter Three examined the impact of a small number of theoretically-relevant procedural modifications on IRAP effects. Building on research suggesting trait valence influences outcomes on implicit measures (e.g., Bluemke & Friese, 2004; Rudman, 2000), Study Three compared effects across two IRAPs: one containing the desirable or positively valenced traits and the other the undesirable or negatively valenced traits. No significant differences were found, suggesting binary biases (at least those measured by the IRAP) are not affected by the valence of the stimuli employed. The next study (Study Four) investigated the theoretical claim that gender stereotypes are simultaneously descriptive and prescriptive by manipulating the relational terms used to connect the stimulus pairs on-screen. The first employed descriptive terms (i.e., men and women *are* masculine/feminine) and the second prescriptive terms (i.e., men and women *should be* masculine/feminine). Again, no significant differences were found, which could either support the claim above or indicate that the IRAP is less sensitive to relational qualifiers than has been previously claimed (Barnes-Holmes et al., 2010).

The final aim of Chapter Three was to investigate the role of the binary in gender inequality, and specifically in an occupational context. Several contemporary models of inequality suggest that women's discrimination arises not because of explicitly negative views about women or their competence, but rather because of (a) the implicit overlap between male, masculine, agentic, and competent traits; and (b)

the denial of these same traits to women (Eagly & Karau, 2002; Koenig et al., 2011). To test this claim, studies in this chapter compared IRAP results and performance to responses on different hiring preference measures. Using the same single-item hiring task from the previous chapter, Study Three found that the traits associated with men on the IRAP were evaluated as more hireable by the vast majority of the sample. As such, they demonstrated that the hireable (i.e., masculine) traits were readily ascribed to men and not women. Studies Four expanded on this by employing a multi-item hiring task and examining the relationship of binary biases to overall gender preferences. Across both, binary biases were positively correlated with a tendency to hire the agentic individual. Study Five additionally assessed the automaticity of gender-competency biases by investigating the extent to which feminine and masculine traits were associated with competent and incompetent traits on an IRAP. Similar to other studies on gender-competency biases, significant masculine-competent and masculine-*not*-incompetent biases were found. Interestingly, however, there were no feminine-incompetent effects. Overall, these data thus provide clear evidence that (i) men but not women are associated with traditional masculine and agentic attributes; (ii) agentic attributes are evaluated as considerably more hireable and desirable in an occupational context; and (iii) the extent to which a person binarises gender (i.e., differentially associates women and men with communal and agentic attributes) increases their preferences for agentic individuals in an occupational setting. While no correlations were found in any study between self-reported sexism and IRAP performance, these data shed light on the broader and subtler relationship between the binary and male dominance and/or supremacy in certain contexts.

Chapter Four

Chapter Four expanded on the previous studies by assessing the relationship of the binary to another theoretically-significant construct: androcentric bias. Feminist theory has long proposed that society has a tendency to equate maleness and humanity, and indeed the results from Chapter Two supported this (i.e., by suggesting that male traits are not as strongly "gendered" as female traits). Studies in this chapter built on this claim and previous data in a few different ways. First, they examined participants' tendencies to assume maleness when presented with an ambiguous stimulus (i.e., they tested the *default male hypothesis*: Smith & Zaraté, 1992). Using a task that was developed based on previous androcentric bias measures (e.g. Bailey et al., 2018), Studies Six and Seven assessed whether participants would attribute a male gender to a gender-neutral facial stimulus. Contrary to existing research, however, no evidence of androcentric bias was found in either study, and in fact a slight gynocentric response pattern emerged. While analyses suggested there may be a relationship between androcentric and binary biases (and specifically men-masculine biases), effect sizes were small and the sample size for this study was relatively low. There was, however, a relationship between self-reported sexism and androcentric bias, but again the effect size was small.

In addition to the explicit measure of androcentrism, studies in Chapter Four also examined "human-as-male" bias at the automatic level. Using two separate IRAPs, Study Five examined the extent to which men and women were differentially associated with two sets of "uniquely human" attributes: (1) complex emotions and (2) scientific ability. The denial of complex emotions to one social group over another is called infrahumanisation, and has been most widely studied in the context of race and ethnicity (Cortes et al., 2005; Demoulin et al., 2009; Gaunt, 2009; Gaunt et al., 2002). Though a small number of studies using self-report methods have

suggested it may also occur in the context of gender (Bearman et al., 2009; Gaunt, 2013), the infrahumanisation IRAP used here found no evidence that men were more readily associated with emotional complexity. Rather, both women-complex and men-complex biases were found and indeed the women-complex biases were stronger. For the gender-science IRAP, results did show a bias towards associating men but with science and not arts (in keeping with existing research: see Miller et al., 2018). However, this study also found a significant bias towards associating women with science, which was unexpected. Overall, therefore, studies in Chapter Four provided weak evidence that androcentrism is a normative feature of gender beliefs, at least in the current population.

Chapter Five

The final IRAP chapter in this thesis examined in the binary in the context of sexually coercive and aggressive behaviours. Along with workplace or economic gender equality, sexual violence is one of the most widely discussed issues in contemporary feminism (see McPhail, 2016). Since the 1970s, feminists have conceptualised rape and other forms of sexual aggression as a key manifestation of masculine dominance, and one which both reflects and reifies our binary patriarchal gender system (Brownmiller, 1975; MacKinnon, 1979). Chapter Five sought to investigate this empirically by assessing the relationship between gender-related biases and a proclivity to engage in sexual harassment in samples of young heterosexual Irish men. The first study in this chapter (Study Eight) built on research suggesting a link between gender normative beliefs and masculine gender identities (see Murnen et al., 2002) by including two IRAPs: one for gender binary biases and another for gender identity biases (i.e., self-masculine or self-feminine associations). Contrary to what was expected, results found no relationship between scores on the harassment propensity measure and the gender binary IRAP. However, there was a

relationship between gender identity and harassment, with the "self-masculine" trial type associated with a higher propensity. This coheres with existing research and feminist theorising (e.g. De Judicibus & McCabe, 2001; Quinn, 2002), though there are a number of methodological and statistical qualifiers to this finding (i.e., small effect size, relatively small sample, etc.).

The second study in this Chapter (Study Nine) investigated the relationship between infrahumanisation biases and harassment proclivity. This study was a partial replication of two previous studies showing that implicit *de*humanisation of women predicts a propensity to engage in harassment (Hussey et al., 2015; Rudman & Mescher, 2012), though it examined gendered associations of complexity rather than explicit humanity. Using the same infrahumanisation IRAP from Study Seven, this study found a significant correlation between propensity scores and effects on the men-complex trial type. However, effects were in the opposite direction than expected, with men-*not*-complex biases associated with an increased propensity to engage in coercion or harassment. Similar to the previous chapter, therefore, while these studies provided interesting insights into the use of the IRAP, these studies do not suggest a role of implicit biases in harassment propensity.

Chapter Six

One of the aims of this research was to generate a larger-than-average IRAP dataset for a pooled analysis. This was in the service of enhancing the statistical power of the analyses (i.e., around trial type, gender difference, and measure comparison tests), and thus more generally increasing the confidence in the conclusions drawn. The primary goal of this work was to examine whether automatic binary biases would be readily demonstrated on an implicit measure, and pooling the data allowed for a more fine-grained analysis of the effects produced in individual studies. Results of the pooled gender binary IRAP dataset (n = 229) showed strong

and significant effects on the three trial types mentioned previously (men-masculine, men-feminine, and women-feminine) and again an absence of any effect for the women-masculine trial type in either direction. Interestingly, results also found a significant main effect for Study, suggesting that the IRAP was influenced at least to some degree by the study's context and/or the additional preceding measures (i.e., the other IRAPs or domain-specific outcome measure). No relationship was found between self-reported sexism and effects on any trial type.

Pooling the IRAP data also allowed for a deeper and more thorough analysis of gender differences in performance. Though not included in the summaries above (for brevity), gender differences on the binary IRAP varied considerably throughout this thesis. As mentioned in the individual chapters, some found significant finding main effects for gender, others found no main effects but some gender interaction effects, and others finding no main or interaction effects. Two separate analyses were conducted to inform these inconsistencies: the first was a simple pooled ANOVA, which did find large and significant gender differences in the data overall. Interestingly, these differences seemed to have been driven by effects on the two trial types involving men (i.e., the men-masculine and men-feminine trial types). The second analysis was a test for measurement invariance across male and female participants. Somewhat surprisingly, the IRAP did not meet the criteria for measurement invariance in this analysis, meaning that it cannot be said to measure the same latent construct (i.e., implicit binary bias) in women and men. While this could be due to a number of reasons, the specific factor loadings and the results of the explicit trait ratings in Chapter Two suggest this is due to different interpretations of the stereotypically masculine traits. This will be discussed in more detail later in this chapter.

7.2 Significance of Findings

Gender Binary Measurement

The most significant contribution of this work was putting forward a new measurement tool for gender binary biases. To date, most research into the binary has relied on abstract or indirect analytic techniques, such as historical re-interpretation, cultural critique, or linguistic and discourse analysis (see Gillis & Jacobs, 2016 and King, 2016 for reviews of common methods in gender studies). Though some questionnaires or self-report measures have been developed to assess psychologically-similar constructs (e.g., gender essentialism, heteronormativity, etc.), very few measures exist for quantifying the binary at the individual level. The current thesis filled this gap by proposing, first, a novel and technical way to operationalise the binary (i.e., as the differential association of women and men with traditional masculine and feminine traits), and second, conducting an in-depth assessment of these automatic associations on a quantitative measure. Results from five binary IRAP studies found strong, robust, and reliable biases towards associating women (but not men) with stereotypically feminine attributes and men (but not women) with stereotypically masculine attributes. In so doing, this work adds to the growing and rich literature into the gender binary, introduces a new experimental paradigm for quantitative researchers, and adds empirical weight to the feminist arguments that the binary is indeed a highly automatic, foundational, and axiomatic feature of gender beliefs (e.g., Ridgeway, 2011).

In addition to providing a new way to operationalise and measure the binary, this work also provides novel and interesting insights into the structure of gender itself. Most mainstream models of gender acknowledge the relational nature of gender categories (e.g., Butler, 1990; West & Zimmerman, 1980), and the broad way in which men and women are framed as "complementary opposites" (Jost & Kay, 2005; Koenig, 2018). Though research to date has generally struggled with

demonstrating this empirically, the IRAP's procedural properties make it ideal for assessing this process. Specifically, because the IRAP can be scored at the trial-type level, this work could assess the extent to which *both* men and women are associated with *both* masculine and feminine traits. That is, it can examine the significance of both role-congruent and incongruent associations to the overall structure of gender roles. Results provide strong evidence that women and men are indeed defined by both their role congruent and incongruent attributes (that is, men are defined by their masculinity and "not-femininity", and women by their femininity and "not-masculinity"). As such, these data support the theoretical argument that male and female traits are not merely distinct, but also *mutually exclusive*. The ability to separate biases in this way provides therefore distinguishes the current work from existing implicit and explicit studies of gender roles stereotypes (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1994; Eagly & Karau, 2002; Newport, 2001; Gaucher, Friesen, & Kay, 2011; Madera, Hebl, & Martin, 2009; Moscatelli, Ellemers, Menegatti, & Rubini, 2016; Rudman, Moss-Racusin, Phelan, & Nauts, 2012; Spence & Buckner, 2000; Williams & Best, 1990) and allows for stronger theoretical conclusions about the relational nature of the binary .

Trial type analyses also shed light on the potentially asymmetrical ways in which we may "gender" men relative to women. Across nearly all studies and IRAPs, participants demonstrated strong resistance towards associating men with traditionally feminine attributes (i.e., strong men-*not*-feminine biases). However, the same was generally not true for women, with results on the women-masculine trial tending to show no significant bias in either direction. While this was not explicitly hypothesised at the outset, this response pattern does cohere with existing theories of gender identity and gender role development. Several researchers have argued that masculinity is potentially a more rigid social construct than femininity with more

well-defined boundaries (Bem, 1993; Leaper & Friedman, 2007; Thorne, 1993). Observational, discourse-analytic, and survey studies have supported this, showing for instance that gender-nonconforming behaviour tend to be more severely punished in men relative to women, particularly in early life (Adams & Coltrane, 2004; Bem, 1993; Kimmel & Messner, 2009; Leaper, 2002), and also that men struggle to a greater degree with gender role stress than women (Levant et al., 2003; Smiler, 2006). As discussed previously, the typical explanations for this effect are, first, that many of our patriarchal or male-dominated social spheres (such as politics or business) place more value on masculine traits, meaning that gender-role deviations are more problematic in men than women (see Coltrane & Adams, 2008). The second explanation is that women are encouraged to aspire to masculinity (as the higher-status category in the gender order), and thus that "progress" tends to manifest as women adopting more agentic characteristics (e.g., Smiler, 2006; Wetherell, 1998). These results provide tentative support for these explanations, though it would be interesting to elaborate on findings with additional measures of gender role beliefs, potentially at multiple levels of analysis (e.g., discursive, attitudinal, etc.).

It is also interesting at this point to draw a comparison between implicit and explicit assessments of gender trait evaluations. Chapter Two gathered explicit trait ratings of various different traits (some stereotypically masculine, some neutral, and some stereotypically feminine) and found that, overall, there were very few "extreme" ratings in the set (i.e., very few traits with a mean rating close to the low or high end of the scale). This was particularly pronounced in the case of the masculine trait ratings. On the implicit measures, however, participants demonstrated strong and reliable biases towards associating women and men with feminine and masculine traits. While the comparisons were between and not within

groups (and thus do not allow for concrete conclusions), the differences in the two datasets potentially attest to the importance of the relational context on trait "gendering". Specifically, because the IRAP requires participants to gender traits in a relational way (i.e., form men-masculine and women-feminine associations at the same time) but the explicit scale requires participants to gender traits in the abstract, these data provide further support that gender categories are relationally defined (e.g., Bem, 1993; Butler, 1990; Harding, 1990).

Additional IRAPs

While the effects on the core gender binary IRAP were both strong and reliable, it is important to note that response patterns on the other IRAPs varied considerably. Throughout this thesis, effects were either weak (e.g., the gender-competency IRAP) and/or broadly inconsistent with existing research or theory (e.g., the infrahumanisation, gender-science, and or identity IRAPs). As discussed in the individual chapters, it is likely these inconsistencies were due in part to different methodological issues with the individual measures. On the infrahumanisation IRAP, for instance, outcomes may have been unintentionally confounded by conflicting gendered connotations associated with the individual exemplars, or indeed the overarching category labels (e.g., "angry men" or "complicated women"). Similarly, the contrast category chosen for the identity IRAP (other men) may have influenced outcomes, while the competency IRAP may have been too complicated for participants to interpret in a quick or intuitive way. Another possible explanation is that the stimulus categories used in the gender-competency and infrahumanisation IRAPs were too weak (i.e. they did not form a strong enough category to use as stimulus labels). In both of these studies, reference sheets were given to participants to clarify the label stimuli (masculine/feminine and complex/simple traits) but this may not have been sufficient to coalesce the categories. Indeed, it may have

influenced results in an unknown way. Though these IRAPs were all included in the service of exploratory, secondary or supplementary research questions (and thus do not undermine the significance of the overall findings too strongly), these data further attest to the importance of careful stimulus and category selection (Nosek, 2007; Steffens et al., 2008). In addition, they demonstrate the benefits of piloting, replicating and validating newly developed implicit measures (see also Levin, 2007).

Outside of the methodological issues, the absence of theoretically-consistent biases on these IRAPs may also be due to the nature of the beliefs or attitudinal constructs themselves. As mentioned in the General Introduction, implicit measures are generally only seen as appropriate for biases that are highly automatic as opposed to those which are more complicated and/or elaborated (Fazio & Olsen, 2003; Greenwald et al.,2002). While the gender binary seems to be a particularly well-entrenched and normative feature of gender cognition, it is possible that the others are too subtle for an implicit assessment and/or less readily demonstrated under conditions of automaticity. Identity, for instance, may require a more detailed assessment paradigm that can properly account for the many contextual, personal, and ideological factors that make up a person's gender self-concept (Van Anders, 2015). Similarly, while the overlap between masculine traits and competency may be evidenced on self-report measures such as the hiring task employed in this thesis, gender role-competency biases may be too complex or contextual for an implicit measure. This explanation is of course conjecture and would require a more focused investigation, and perhaps a more direct comparison of implicit and explicit measures of the same construct. Overall, however, results support the theory of automaticity in the context of social cognition (Bargh, 1994; Gawronski, 2013), and more generally the argument that different types of attitudes may require different types of measures (De Houwer, 2001; Hughes et al., 2011).

Gender Differences in Binary Biases

The next finding that has significant implications for the literature concerns the gender differences. While differences in male and female performances varied across individual studies, the pooled analysis in Chapter Six found large and significant differences on the binary IRAP, with men showing more binary-consistent effects on nearly all trial-types (women had higher scores on the women-feminine trial-type). A face value interpretation of this finding is that women and men simply differ in terms of the underlying construct being measured (i.e., binary biases), and thus that men have stronger automatic biases in the context of gender. This interpretation would be consistent with the broader body of literature around explicit and implicit gender beliefs, and the reliable finding that men hold more sexist, traditional, and gender-normative views than women (see Russell & Trigg, 2004). As the IRAP failed measurement invariance, however, these differences should not be attributed to genuine variation in the latent variable. Rather, they appear to result from differences in how women and men navigate, respond to, or interpret the task. While measurement invariance analyses cannot provide definitive causes for why different groups may complete a measure in the same way (Flake et al., 2017), the likely explanation here is differing interpretations of the stimuli used in the measure.

To elaborate, based on the results from the trait rating scale in Chapter Two and the confirmatory factor analysis in Chapter Six, the logical explanation would be that women and men differentially evaluate or "gender" the masculine-agentic traits. As mentioned previously, research conducted over the past few decades suggests women's stereotypes have been gradually moving in a more androgynous and/or agentic direction (Brewster & Padavic, 2000; Cotter et al., 2011; Eagly & Karau, 2002; Miller et al., 2018). In the IRAP, therefore, if these traits were not gendered in

the same way by female and male participants, then the entire structure and format of the task would have been different for these two groups. For instance, if the feminine traits were in fact the more gendered attributes for women, then they may have formed a more salient or coherent category. As such, they may have unintentionally paid greater attention to the feminine trial types over the others, thus influencing reaction times and potentially overall effects. A related possibility is that women had slower relative reaction times to the men-masculine pairings during the binary-consistent IRAP blocks (i.e., the blocks where they associated men with masculine and women with feminine), given that these traits were not seen as strongly male. These are just two possible explanations, however, and would require a more focused analysis in order to be substantiated.

It is important to note, of course, that the lack of measurement invariance may not specific to the gender binary IRAP. That is, it may be a more general issue with the IRAP or indeed implicit measures more generally. Research shows that men and women often in terms of their performances on cognitive and behavioural tasks (see Richardson, 2013 for a comprehensive review), and that they may occasionally use different strategies when navigating through psychological paradigms (e.g., generate different mental rules or heuristics, engage in more or less proactive behaviours, etc.: see Scheuringer, Wittig & Pletzer, 2017). While no other measurement invariance analyses of implicit measures could be found in the literature for a direct comparison, studies of questionnaire data suggest variance across gender groups is common (e.g., Nien & Duda, 2008; Levant et al., 2013). Ambivalent Sexism Theory (Glick & Fiske, 2001) do recommend treating gender as a mediator and not a moderator of gender beliefs, and indeed results here support this explanation, and indeed perhaps changing how gender is analysed in psychological experiments. This is of course a much broader issue than could be meaningfully

discussed here, but generally the issue of gender variance in measure development requires addressing (Flake et al., 2017; Hatlevik et al., 2017).

The IRAP Methodology

Though analysing the IRAP's methodological features and properties was not a goal of this work, results do inform our understanding of the measure and its use in social psychological research. The most significant contribution was elucidating the role of stimuli, and thus demonstrating the importance of careful stimulus selection procedures. Without the explicit trait ratings collected in Chapter Two, for instance, it would have been very difficult to interpret the results of the measurement invariance analysis or indeed the gender differences more broadly. It also would have been difficult to draw strong conclusions about the measure itself and its ability to detect role-congruent or role-incongruent gender biases (that is, because the attributes could not have been said to meaningfully represent masculine and feminine attributes for the populations sampled in later studies). This research demonstrates the utility and benefits of both conducting pilot research and following clear stimulus selection procedures, and thereby adds support to the arguments made elsewhere around the importance of stimuli in implicit measures (Steffens et al., 2008).

Results from this thesis also informed our understanding of the IRAP's sensitivity (or lack of sensitivity) to certain procedural modifications. To look first at the impact of study context, the pooled analyses found that binary IRAP effects differed significantly across the five experiments. This of course could have been due to natural variation in the samples, and indeed the unequal gender distribution across studies may have influenced outcomes (i.e. Study Eight was a male-only sample). Generally speaking, however, the samples were comparable in terms of their demographics and self-reported gender beliefs and it is therefore likely that

performances were influenced at least in part by the other measures. The domain-specific measure was always presented before the IRAPs in this thesis, and these different measures varied considerably in both content and length. Specifically, the brief hiring preference measure was quite opaque in its reference to gender, while the longer sexual harassment measure was very clearly about gender-related abuse. Though the research into the impact of additional tasks on implicit measures is scarce, some studies do suggest they are susceptible to procedural influences or primes (e.g., Gawronski, Geschke & Banse, 2003; Payne, Burkley & Stokes, 2008). In addition to the domain-specific task, most studies in this thesis also included a second IRAP, which may have also influenced results. Though the order of IRAPs were always counterbalanced across participants (in keeping with convention: see Hussey et al., 2015), this variety would have meant differences in both IRAP experience and fatigue across studies. As both are known to influence implicit effects, the variation in the number of IRAPs included across studies may have been a factor here (see Nosek et al., 2005). Regardless of the source of the effect, future research should examine this more directly (e.g., by using a single IRAP in a homogenous sample and manipulating the content, length or style or a preceding measure).

Though the IRAP was unexpectedly sensitive to changes in the study's context, it was surprisingly insensitive to other procedural changes made in early experiments. Study Three found no differences in IRAP performance depending on the valence of the traits used, and analyses also found no differences in the extent to which theses biases were associated with scores on other measures (i.e., the sexism or hiring preference measure). Similarly, Study Four found no difference depending on the relational terms used (i.e., *are* versus *should be*), though these IRAPs did differ slightly in their relationship to hiring preferences. As mentioned in the

individual chapters, the lack of difference across IRAPs could be evidence of the simultaneously positive, negative, prescriptive, and descriptive nature of contemporary gender roles (Eagly & Karau, 2002; Prentice & Carranaza, 2002; Koenig, 2018). Alternatively, it could be that participants are simply responding to the most salient information for each of the categories (i.e., "masculine" or "feminine") and ignoring other more sensitive procedural information. This would contradict claims made elsewhere in the literature that the IRAP can detect subtle stimulus relations (e.g. Barnes-Holmes et al., 2010), though again this is conjecture and would require a follow up in a future study.

This work's final contribution to our understanding of the IRAP comes from the measurement invariance analysis. While the significance of the measurement variance to our understanding of gender differences was discussed above, the analysis also raises more general questions around the use of the IRAP in social psychological research. To put it simply, if a gender IRAP is variant across gender groups, then it is highly likely IRAPs in other contexts would be variant as well (i.e., race IRAPs for different racial groups, age IRAPs for different age groups, and so on). This would have significant consequences for the interpretation of the effect (i.e., as indicative of the latent construct or not: see Flake, 2017), but would also significantly limit the IRAP's use in correlational, group comparison, or longitudinal designs. According to Putnick & Bornstein (2016), measurement invariance should be considered akin to other forms of validity and reliability testing, and should be factored into the measure development process. When a measure fails tests for invariance, these researchers suggest continually modifying and re-testing different items, stimuli, or procedural properties until the measure meets criteria. In our case, this may mean returning to the individual stimuli used and working on finding more representative exemplars for the individual categories, though future research could

also go back and examine the impact of other manipulations (e.g., response rules or instructions, etc.).

The Role of the Binary in Inequality

The final significant finding of this research concerns the role of the binary in inequality. Feminist theorising has long implicated binary ideologies in women's oppression (Bem, 1993; Butler, 1990; Harding, 1990; Kimmel, 2009), but experimental work in this area is scarce. The current research thus conducted several investigations into the relationship between binary biases and other forms of self-reported discrimination, bias, and prejudice. Results varied considerably across studies and domains, with evidence found in some contexts and not in others. In the context of workplace inequality, results from Chapters Two and Three provided clear evidence that the binary underpins male dominance in occupational settings. Across both implicit and explicit assessments, there was a clear and direct link between men, masculine, and hireable traits, and a recurrent resistance towards associating these same hireable traits with women. Direct evidence for this relationship came from Studies Four and Five, which both found moderate positive correlations between binary biases on the IRAP and overall masculine hiring preferences. Overall, these findings thus support previous research identifying a link between masculine traits and competency or leadership (Eagly & Carli, 2007; Koenig et al., 2011; Koenig, 2018) and more broadly the argument that Western societies continue to implicitly prioritize masculine values and traits (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Johnson, 2005; Serano, 2009).

In the other contexts, however, the relationships were either weak, underpowered, or in the unexpected direction. With regards to infrahumanisation, effects were inconsistent with previous research and there were no meaningful relationships between such biases and performances on the IRAP. Similarly, while

there were some links between binary biases and sexual harassment proclivity, generally speaking the effects were weak and should be interpreted cautiously given the number of comparisons. The pooled analysis also found no relationship between self-reported sexism and scores any binary IRAP trial-type. Though these effects should all be considered tentative given the number of statistical comparisons being made in each study and the various methodological issues with individual IRAPs (discussed above), it is nonetheless important to note that studies in this thesis provided little evidence that the binary underpins inequality. That is, at least at the level of belief-belief or belief-behaviour relationships, conclusions would be very different at a sociological or cultural level of analysis.

There are, however, some important methodological caveats to both the significant and insignificant results that require addressing. First, the analyses throughout this thesis were conducted on the entire sample (i.e., female and male participants), but the IRAP failed measurement invariance. When a measure is variant, combined analyses are not considered appropriate (Flake, 2017) and this calls the above findings and measure comparison conclusions generally into question. Second, it should be noted that there is a general debate within the attitude literatures regarding the predictive utility of implicit measures, and precisely when (or indeed if) implicit and explicit assessments should be expected to correlate (Nosek, 2007; Payne et al., 2008). Attitudes are multifaceted constructs that are likely subject to a range of personal, situational, and cultural influences (Deaux & Snyder, 2012). The current research assessed only one component aspect of gender bias at one specific level of analysis, but in reality social attitudes and behaviours operate at multiple levels of complexity, awareness, and intentionality (De Houwer, 2001; Hughes et al., 2011). It could thus be that binary biases (measured in this way) are more of a feature or reflection of binarist ideologies, and not necessarily a

"stage" in the discrimination process. While not a limitation of the current work as such, this debate is a useful reminder to be precise around the language and terminology used (i.e., not to conflate an automatic bias with a deliberated and elaborated set of beliefs, or indeed with harmful real-world behaviours), and also tentative regarding the nature of the relationship that is expected between measures, even in theory.

7.3 Issues for Consideration

Cultural and Contextual Factors

While each of the main findings have now been discussed, it is important to acknowledge some more general limitations and issues for consideration or follow up. Firstly, though this was mentioned briefly in the General Introduction, it is important to acknowledge that this thesis worked entirely within a Western framework. The roles used here reflect the specific spatial and temporal location of the modern West and should thus not be considered reflective of gender categories on the whole (see Fausto-Sterling, 2000; and Oakley, 2016 for a more detailed review of role differences across cultures and time periods). Also, while it was necessary for theoretical and design purposes, this thesis also treated masculinity and femininity as homogenous categories when of course they are varied, complex social constructs. Radical and third-wave feminists increasingly acknowledge the intersectional nature of social categories, and the ways in which different social dimensions interact in the context of inequality (Collins, 2015; Crenshaw, 1989; Davis, 2008). For example, research into the lived experiences of Black women shows that their oppression is more than the simple combined experiences of being both Black and a woman. Rather, these individuals can experience entirely unique forms of prejudice that neither Black men or White women experience (e.g., hyper-sexualisation: see Benard, 2016 for a review). This is of course just one example of

intersectional prejudice, and there are several dimensions that are known to influence and interact with gender evaluations (e.g., class, physical ability, and so on: see hooks, 2014). While it would be difficult to bring this level of complexity to bear in an implicit assessment, future research should be both conscious of these intersections and potentially focused on examining their influence on bias.

It is also very important to note that this research worked entirely within a binary system of knowledge. Participants across all studies identified as either female or male, and no studies investigated evaluations or gendered associations about non-binary individuals. As reviewed at length in the introduction, non-binary individuals experience considerable social prejudice and discrimination (Hegarty et al., 2018; Murjan & Bouman, 2015; Van Anders et al., 2017) and it is important that research continues to understand and interrogate these biases. Moreover, it is essential that researchers themselves take active steps to reach and include marginalised groups in their studies, especially when their beliefs and experiences would be of relevance to the construct under investigation (as it would have been here). As noted by Baeur, Braimoh, Scheim and Dharma (2017), it has become increasingly normal to exclude small groups from research or analyses for statistical reasons, meaning that over time these groups are essentially erased from experimental research. Though it would indeed have been statistically difficult to account for smaller non-binary groups in analyses here, one study could have directly targeted these individuals and potentially included a third group for analysis. The absence of non-binary groups and constructs should thus be considered a limitation of this work.

Lastly, it should be acknowledged that the past five years have seen a considerable change in Ireland (and indeed the broader West's) cultural climate around gender issues. In the period since this data were collected, there have been a

number of significant social advances around sexual and gender equality, including online movements (e.g., the #MeToo, #IBelieveHer, #TimesUp: see Jackson, 2018) and significant legislative change (including the Marriage Equality and Reproductive Rights referenda). However, there has also been a significant increase in reactionary movements (e.g., the Men's Rights movements, or the more recent "involuntary celibate" online communities: Gotell & Dutton, 2016; Jordan, 2016; Schmitz & Kazyak, 2016), and a notable (although anecdotal) surge in public displays of gender-motivated or anti-women violence (Boyd, 2018; Williams, 2018). This is a highly active and labile period for gender rights and discussions, particularly on college campuses (Haines, Deaux & Lofaro, 2016), and it is therefore likely that attitudes have changed among the samples employed. While future research should (and is beginning to) explore the impact of such movements on gender beliefs (e.g., Kunst, Bailey, Prendergast & Gundersen, 2018; Swank & Fahs, 2017), this period of change should be considered an important caveat when interpreting the generalisability of the current results.

Methodological Factors

Though the current work examined and/or controlled for the influence of some methodological properties on IRAP results (i.e., stimulus gender ratings, stimulus valence, relational terms, etc.), the impact of other features was not directly assessed. These include, but are not limited to, the influence of: block or IRAP order; response rules (both their presence and their phrasing); response key placement; stimulus type (i.e., pictorial, symbolic, auditory, or verbal); block number; trial number, and so on. This was of course because of both space and time constraints, but it should be noted that all of the above factors are modifiable and have been shown to occasionally influence outcomes on implicit measures (Fazio & Olsen, 2013; De Houwer, Teige-Mocigemba, Spruyt & Moors, 2009; Garwonski & De

Houwer, 2014; Greenwald, Poehlman, Uhlmann & Banaji, 2009). As such, future research should consider examining these variables in more detail, and in the specific context of a gender IRAP.

Another methodological factor to consider is the particular scoring technique employed here. In this work, scores were produced using the D_{IRAP} scoring algorithms on the metric that is widely used within the IRAP and broader implicit measures literature (i.e., mean differences in reaction time: Nosek, 2007). Though this was primarily for consistency and ease of comparison with existing work, it should be noted that concerns have been raised in the literature around these conventions. Researchers have questioned, for instance, whether response latencies are capturing all of the necessary information in a measure (e.g., relative to other metrics like accuracy or fluency: see Gavin, Roche & Ruiz, 2008). Others have more generally questioned whether the mean is an appropriate measure of central tendency for implicit measures, given that reaction times are not normally distributed (Whelan, 2010). Future research could investigate these claims by systematically comparing different metrics and algorithms on some relevant dimension of interest (e.g., predictive validity).

Outside of the IRAP, there are some methodological issues to consider with the self-report tasks. Both the androcentrism and hiring preference tasks were developed for the purpose of this work, and there may have therefore been any number of issues related to their stimuli, format, and scoring. For example, in the androcentrism task, running a focus group in advance may have caught that the gender-neutral face seemed more feminine than gender-neutral. Similarly, while the scoring technique employed in the expanded hiring task (i.e. summing and averaging preferences) was logical and allowed for simple analyses, psychometric tests of its validity etc. could have led to alternative and more appropriate ways of scoring this

measure. Though there was not sufficient time to explore these issues in detail in this thesis (though see the individual chapter discussions for a review), such investigations would have strengthened the conclusions drawn about these biases.

Participant Factors

Lastly, as in any research, it is important to examine generalisability and the ways in which participants may have influenced results. Participants in this study were predominantly recruited from the Maynooth University student populations and as such were significantly more likely to be White, well-educated, and middle class. This should of course not be considered representative of the broader Irish population, even within that age range, and future research should explore the potential impact of demographics on binary biases. This work did also not carry out comprehensive assessments of participants' gender-related beliefs, identities, and attitudes, and as such participants may have varied in unforeseen ways across different studies. Though it would not have been feasible to assess beliefs in this way, doing so would have considerably strengthened the conclusions drawn about binary biases and indeed the role of the binary in gender attitudes more broadly. Future studies should consider exploring this in more detail.

It is also worth noting that all of the researchers involved in data collection for this thesis were women. The influence of researcher gender on experimental outcomes is well documented (see Chapman, Benedict & Schioth, 2018, for a recent meta-analysis of these effects). While certain steps were taken to minimize the potential impact of researcher gender (e.g. emphasising anonymity, having the participants complete the experiment on their own in a cubicle), the fact that the research was evidently about gender and inequality issues may have led to social desirability biases or other unintended influences on the data.

7.4 Concluding Remarks

In conclusion, the current work was a unique, quantitative, and direct assessment of gender binary biases in an Irish context. Across Nine studies and two pooled analyses, this work shed light on the social construction of the binary at the individual level, and attested to the centrality and automaticity of the binary within gender cognition. In so doing, it introduced new measurement tool for those interested in gender ideology, and indeed added empirical weight to some central and historic feminist arguments. This work also explored the binary in the context of three key manifestations of patriarchy – male supremacy, male centrality, and male aggression – and provided some evidence that the binary may indeed form the foundation of women's oppression, at least in some settings. Findings from this thesis also informed our understanding of the IRAP methodology, and indeed gender attitude measurement itself.

References

Aboud, F. E. (1984). Social and cognitive bases of ethnic identity constancy. *The Journal of Genetic Psychology, 145*(2), 217-229.

Acker, J. (2009). From glass ceiling to inequality regimes. *Sociologie du Travail, 51,* 199-217.

Adams, M., & Coltrane, S. (2004). Boys and men in families. *Men and Masculinities*, *21,* 230-248.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652.

Anderson, K. J., & Leaper, C. (1998). Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how. *Sex Roles, 39*(3-4), 225-252.

Anderson, P. S. (2003). Autonomy, vulnerability and gender. *Feminist Theory, 4*(2), 149-164.

Archer, E. M. (2010). The enemy within: stereotype threat and the disarming of American servicewomen. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 70*(9-A), 3628.

Aries, E. J., & Johnson, F. L. (1983). Close friendship in adulthood: Conversational content between same-sex friends. *Sex Roles, 9*(12), 1183-1196.

Arkkelin, D., & Simmons, R. (1985). The "good manager": Sex-typed, androgynous, or likable?. *Sex Roles, 12*(11), 1187-1198.

Atwater, L. E., Carey, J. A., & Waldman, D. A. (2001). Gender and discipline in the workplace: Wait until your father gets home. *Journal of Management*, *27*(5), 537-561.

Ayman, R., Korabik, K. & Morris, S. (2009). Is transformational leadership always evaluated as effective? Male subordinates' devaluations of female transformational leaders. *Journal of Applied Social Psychology, 39*(4), 852-879.

Babcock, L., & Laschever, S. (2009). *Women don't ask: Negotiation and the gender divide.* Princeton, NJ: Princeton University Press.

Bacchi, C. L. (1999). *Women, policy and politics: The construction of policy problems*. Thousand Oaks, CA: Sage.

Baer, J. (2011). Feminist theory and the law. *The Oxford handbook of political science*. Oxford: Oxford University Press.

Bailey, A. H., & LaFrance, M. (2016). Anonymously male: Social media representations are implicitly male and resistant to change. *Cyberpsychology, 10,* 4, 8-12.

Bailey, A. H., & LaFrance, M. (2017). Who counts as human? Antecedents to androcentric behavior. *Sex Roles*, *76*(11-12), 682-693.

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2018). Is man the measure of all things? A social cognitive account of androcentrism. *Personality and Social Psychology Review*, *23*(4), 307-331.

Bandura, A. (1977). *Social Learning Theory*. New York, NY: General Learning Press.

Barak, A., Fisher, W. A., Belfry, S., & Lashambe, D. R. (1999). Sex, guys, and cyberspace: Effects of internet pornography and individual differences on men's attitudes toward women. *Journal of Psychology & Human Sexuality, 11*(1), 63-91.

Barbercheck, M. (2001). *Mixed messages: Men and women in advertisements in science*. *Sex Roles, 65,* 117-131

Barbero-Rubio, A., López-López, J. C., Luciano, C., & Eisenbeck, N. (2016). Perspective-taking measured by implicit relational assessment procedure (IRAP). *The Psychological Record, 66*(2), 243-252.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*(3), 527-542.

Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure: Exploring the impact of private versus public contexts and the

response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60*(1), 57-79.

Barnett, H. (1998). *Introduction to feminist jurisprudence.* London: Cavendish Publishing Limited.

Baron, A. S., Schmader, T., Cvencek, D., & Meltzoff, A. N. (2014). The gendered self-concept: How implicit gender stereotypes and attitudes shape self-definition. *Gender and Development, 24,* 109-132.

Bartling, C. A., & Eisenman, R. (1993). Sexual harassment proclivities in men and women. *Bulletin of the Psychonomic Society, 31*(3), 189-192.

Bauer GR, Braimoh J, Scheim AI, & Dharma C (2017). Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PLoS ONE,* 12(5), e234456.

Bearman, S., Korobov, N., & Thorne, A. (2009). The fabric of internalized sexism. *Journal of Integrated Social Sciences, 1*(1), 10-47.

Begany, J. J., & Milburn, M. A. (2002). Psychological predictors of sexual harassment: Authoritarianism, hostile sexism, and rape myths. *Psychology of Men & Masculinity*, *3*(2), 119.

Beggs, J. M., & Doolittle, D. C. (1993). Perceptions now and then of occupational sex typing: A replication of Shinar's 1975 study. *Journal of Applied Social Psychology*, *23*(17), 1435-1453.

Bem, S. (1993). *The lenses of gender: Transforming the debate on sexual inequality*. New Haven, CT: Yale University Press.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 155-162.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*(4), 354.

Benard, A. A. (2016). Colonizing Black female bodies within patriarchal capitalism: Feminist and human rights perspectives. *Sexualization, Media, & Society*, *2*(4), 237-254.

Bennett, J. M. (1989). Feminism and history. *Gender & History*, *1*(3), 251-272

Benwell, B. (2017). Language and Masculinity. In Susan Ehrlich, Miriam Meyerhoff & Janet Holmes (eds.). *The Handbook of Language, Gender, and Sexuality.* Oxford: Oxford University Press.

Berdahl, J. L. (2007). The sexual harassment of uppity women. *Journal of Applied Psychology, 92*(2), 425.

Berger, J., Rosenholtz, S. & Zelditch, M. (1980). Status organizing processes. *Annual Review of Sociology, 6,* 479-508

Besnier, N., Phillips, S. U., Ehrlich, S., Meyerhoff, M., & Holmes, J. (2014). Ethnographic methods for language and gender research. *The Handbook of Language, Gender and Sexuality*, 123-40.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389-391.

Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues, 57*(4), 707-724.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science, 16*(3), 162-166.

Bigler, R., Hayes, A. R., & Hamilton, V. (2013). The role of schools in the early socialization of gender differences. *Encyclopaedia on Early Childhood Development,* 1-4.

Bing, J. M., & Bergvall, V. L. (1996). The question of questions: Beyond binary thinking. *Rethinking Language and Gender Research: Theory and Practice, 1,* 30.

Bingham, S. G., & Burleson, B. R. (1996). The development of a sexual harassment proclivity scale: Construct validation and relationship to communication competence. *Communication Quarterly, 44*(3), 308-325.

Bird, S. R. (1996). Welcome to the men's club: Homosociality and the maintenance of hegemonic masculinity. *Gender & Society, 10*(2), 120-132.

Bivens, R. (2017). The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. *New Media & Society*, *19*(6), 880-898.

Blaise, M. (2005). *Playing it straight!: Uncovering gender discourses in the early childhood classroom.* New York, NY: Routledge.

Blashill, A. J., & Powlishta, K. K. (2009). Gay stereotypes: The use of sexual orientation as a cue for gender-related attributes. *Sex Roles*, *61*(11-12), 783-793.

Blashill, A. J., & Powlishta, K. K. (2009). The impact of sexual orientation and gender role on evaluations of men. *Psychology of Men & Masculinity*, *10*(2), 160.

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, *42*(2), 163-176.

Boccato, G., Cortes, B. P., Demoulin, S., & Leyens, J. P. (2007). The automaticity of infra-humanization. *European Journal of Social Psychology*, *37*(5), 987-999.

Bodine, A. (1975). Androcentrism in prescriptive grammar: Singular 'They', Sex-indefinite 'He', and 'He or She'. *Language in Society, 4*(2), 129-146

Bolzendahl, C. I., & Myers, D. J. (2004). Feminist attitudes and support for gender equality: Opinion change in women and men, 1974–1998. *Social Forces*, *83*(2), 759-789.

Bowles, H. R., & Babcock, L. (2009). Are outside offers and answer to the compensation negotiation dilemma for women? *Academy of Management Proceedings, 1*, 1-6

Braun, F., Sczesny, S. & Stalhberg, D. (2005). Cognitive effects of masculine generics in German: An overview of empirical findings. *Communications, 30*(1), 1-21

Brescoll, V. L. (2011). Who takes the floor and why: Gender, power, and volubility in organizations. *Administrative Science Quarterly*, *56*(4), 622-641.

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science*, *19*(3), 268-275.

Brescoll, V. L., Uhlmann, E. L., & Newman, G. E. (2013). The effects of system-justifying motives on endorsement of essentialist explanations for gender differences. *Journal of Personality and Social Psychology*, *105*(6), 891.

Brescoll, V., & LaFrance, M. (2004). The correlates and consequences of newspaper reports of research on sex differences. *Psychological Science, 15*(8), 515-520.

Brewster, K. L., & Padavic, I. (2000). Change in gender-ideology, 1977–1996: The contributions of intracohort change and population turnover. *Journal of Marriage and Family*, *62*(2), 477-487.

Brochu, P. M., & Morrison, M. A. (2007). Implicit and explicit prejudice toward overweight and average-weight men and women: Testing their correspondence and relation to behavioral intentions. *The Journal of Social Psychology*, *147*(6), 681-706.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1994). Sex-role stereotypes: A current appraisal. *Journal of Social Issues 28*(2), *59-78.*

Brown, L. S. (1990). Taking account of gender in the clinical assessment interview. *Professional Psychology: Research and Practice, 21*(1), 12-17.

Brownmiller, S. (1975). *Against our will: Men, women and rape.* New York, NY: Pelican Books.

Brownmiller, S. (1999). *In our time: Memoir of a revolution*. New York, NY: Dial Press.

Bruckmüller, S., Hegarty, P., & Abele, A.E. (2012). Framing gender differences: Linguistic normativity affects perceptions of gendered power and gender stereotypes. *European Journal of Social Psychology, 42,* 210-218.

Bunreacht na hÉireann, the Constitution of Ireland, adopted by referendum 1 July 1937, entered into force 29 December 1937.

Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, *5*(3), 665.

Burgess-Jackson, K. (1996). *Rape: A philosophical investigation*. Brookfield, VT: Dartmouth.

Burton, L. J., Barr, C. A., Fink, J. S., & Bruening, J. E. (2009). Think athletic director, think masculine?: Examination of the gender typing of managerial subroles within athletic administration positions. *Sex Roles, 61*, 416-426.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676-688.

Butler, D., & Geis, F. L. (1990). Nonverbal affect responses to male and female leaders: Implications for leadership evaluations. *Journal of Personality and Social Psychology, 58(*1), 48-59.

Butler, J. (1988). Performative acts and gender constitution: An essay in phenomenology and feminist theory. *Theatre Journal*, *40*(4), 519-586.

Butler, J. (1990). Gender trouble, feminist theory, and psychoanalytic discourse. In L. Nicholson, *Feminism/Postmodernism* (pp. 324-340). New York, NY: Routledge.

Butler, J. (2002). *Gender trouble: feminism and the subversion of identity.* New York, NY: Routledge.

Byers, E. S. (1996). How well does the traditional sexual script explain sexual coercion? Review of a program of research. *Journal of Psychology & Human Sexuality, 8*(1-2), 7-25.

Cahill, B., & Adams, E. (1997). An exploratory study of early childhood teachers' attitudes toward gender roles. *Sex Roles*, *36*(7-8), 517-529.

Cameron, D. (1997). Theoretical debates in feminist linguistics: Questions of sex and gender. *Gender and Discourse*, *1*, 21-36.

Campbell, B. Schellenberg, E. G. and Senn, C. Y. (1997). Evaluating measures of contemporary sexism. *Psychology of Women Quarterly, 21,* 89-102.

Campbell, R., Greeson, M. R., Bybee, D., & Raja, S. (2008). The co-occurrence of childhood sexual abuse, adult sexual assault, intimate partner violence, and sexual harassment: a mediational model of posttraumatic stress disorder and physical health outcomes. *Journal of Consulting and Clinical Psychology, 76*(2), 194.

Cann, A., Siegfried, W. D., & Pearce, L. (1981). Forced attention to specific applicant qualifications: Impact on physical attractiveness and sex of applicant biases. *Personnel Psychology, 34*(1), 65-75.

Carilli, T., & Campbell, J. (Eds.). (2005). *Women and the media: Diverse perspectives*. Lanham, MD: University Press of America.

Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes about gender and science: Women ≠ scientists. *Psychology of Women Quarterly, 4*(2), 244-260.

Carli, L. L., LaFleur, S. J., & Loeber, C. C. (1995). Nonverbal behavior, gender, and influence. *Journal ofPpersonality and Social Psychology*, *68*(6), 1030.

Carnes, M., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., ... & Palta, M. (2015). Effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. *Academic Medicine: Journal of the Association of American Medical Colleges*, *90*(2), 221.

Ceci, J. S. & Williams, W. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8), 3157-3162

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin, 25*(4), 413-423.

Central Statistics office. (2016). Women and Men in Ireland Report. Retrieved from https://www.cso.ie/en/releasesandpublications/ep/p-wamii/womenandmeninireland2016/

Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior, 31*(5) 334-347.

Cherry, F., & Deaux, K. (1978). Fear of success versus fear of gender-inappropriate behavior. *Sex Roles, 4*(1), 97-101.

Cheshire, J., & Trudgill, P. (Eds.). (1998). *The sociolinguistics reader: Gender and discourse*. London: Arnold Press.

Coates, G. (1997). Organisation man: Wmen and organisational culture. *Sociological Research Online, 2*, 3-19.

Coates, J. (2015). *Women, men and language: A sociolinguistic account of gender differences in language.* New York, NY: Routledge.

Cohen-Kettenis PT, van Goozen SHM (1997), Sex reassignment of adolescent transsexuals: a follow-up study. *J Am Acad Child Adolesc Psychiatry 36*, 263–271

Collins, P. H. (1990). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. New York, NY: Routledge.

Coltrane, Scott, and Michele Adams. (2008). *Gender and families.* New York, NY: Rowman & Littlefield.

Connell, R. W. (1982). Class, patriarchy, and Sartre's theory of practice. *Theory and Society*, *11*(3), 305-320.

Connell, R. W. (1985). Theorising gender. *Sociology*, *19*(2), 260-272.

Connell, R. W. (2005). Change among the gatekeepers: Men, masculinities, and gender equality in the global arena. *Signs: Journal of Women in Culture and Society*, *30*(3), 1801-1825.

Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society, 19*(6), 829-859.

Connell, R. W., Ashenden, D., & Kessler, S. (1982). *Making the difference: Schools, families and social division.* Sydney, Australia: Allen & Unwin.

Conway, M., Pizzamiglio, M. T., & Mount, L. (1996). Status, communality, and agency: Implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology, 71,*1, 25-38.

Cortes, B. P., Demoulin, S., Rodriguez, R. T., Rodriguez, A. P., & Leyens, J. P. (2005). Infrahumanization or familiarity? Attribution of uniquely human emotions to the self, the ingroup, and the outgroup. *Personality and Social Psychology Bulletin*, *31*(2), 243-253.

Cotter, D., Hermsen, J. M., & Vanneman, R. (2011). The end of the gender revolution? Gender role attitudes from 1977 to 2008. *American Journal of Sociology*, *117*(1), 259-89.

Crenshaw, K. (1989). Demarginalizing the intersection of race and Sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum, 8*, 12-42.

Cronin, A. M. (2015). Gendering friendship: couple culture, heteronormativity and the production of gender. *Sociology, 49*(6), 1167-1182.

Cuddy, A. J., Rock, M. S., & Norton, M. I. (2007). Aid in the aftermath of Hurricane Katrina: Inferences of secondary emotions and intergroup helping. *Group Processes & Intergroup Relations*, *10*(1), 107-118.

Cullen, C., & Barnes-Holmes, D. (2008). Implicit pride and prejudice: A heterosexual phenomenon. *Modern Prejudice, 1*, 95-223.

Cutler, A., & Scott, D. R. (1990). Speaker sex and perceived apportionment of talk. *Applied Psycholinguistics, 11*(3), 253-272.

Dailey M.N., Joyce C., Lyons M.J. (2010). Evidence and a computational explanation of cultural differences in facial expression recognition. *Emotion, 10,* 874–893.

Dall'Ara, E., & Maass, A. (1999). Studying sexual harassment in the laboratory: Are egalitarian women at higher risk?. *Sex Roles*, 41(9), 681-704.

Daly, M. (1985). *Beyond God the father: Toward a philosophy of women's liberation*. Boston, MA: Beacon Press.

Daly, M., & Wilson, M. (1983). *Sex, evolution, and behaviour.* Boston, MA: Willard Grant Press.

Dar-Nimrod, I., & Heine, S. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin*, *137*(5), 800-818.

Davidson, M. J., & Burke, R. J. (Eds.). (2000). *Women in management: Current research issues* (Vol. 2). York, CA: Sage.

De Beauvoir, S., (1979). *The second sex*. New York, NY: Vintage books.

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*(6), 443-451.

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass, 8*(7), 342-353.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In: *Implicit measures of attitudes: Procedures and controversies* (pp 179-194). New York, NY: Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*(3), 347.

De Judicibus, M., & McCabe, M. P. (2001). Blaming the target of sexual harassment: Impact of gender role, sexist attitudes, and work role. *Sex Roles, 44*(7-8), 401-417.

De Visser, R. O., Smith, J. A., & McDonnell, E. J. (2009). 'That's not masculine' Masculine Capital and Health-related Behaviour. *Journal of health psychology*, *14*(7), 1047-1058.

Dean, K. E., & Malamuth, N. M. (1997). Characteristics of men who aggress sexually and of men who imagine aggressing: Risk and moderating variables. *Journal of Personality and Social Psychology*, *72*(2), 449.

Deaux, K., & Snyder, M. (Eds.). (2012). *The Oxford handbook of personality and social psychology*. Oxford: Oxford University Press.

DeLamater, J., & Hyde, J. (1998). Essentialism vs. social constructionism in the study of human sexuality. *Journal Of Sex Research*, *35*(1), 10-18.

DeLoache, J. S., Cassidy, D. J., & Carpenter, C. J. (1987). The three bears are all boys: Mothers' gender labeling of neutral picture book characters. *Sex Roles*, *17*(3-4), 163-178.

Demoulin, S., Cortes, B. P., Viki, T. G., Rodriguez, A. P., Rodriguez, R. T., Paladino, M. P., & Leyens, J. (2009). The role of in-group identification in infra-humanization. *International Journal of Psychology, 44*(1), 4–11.

Demoulin, S., Torres, R. R., Perez, A. R., Vaes, J., Paladino, M. P., Gaunt, R., ... & Leyens, J. P. (2004). Emotional prejudice can lead to infra-humanisation. *European Review of Social Psychology*, *15*(1), 259-296.

Deogracias, J. J., Johnson, L. L., Meyer-Bahlburg, H. F., Kessler, S. J., Schober, J. M., & Zucker, K. J. (2007). The gender identity/gender dysphoria questionnaire for adolescents and adults. *Journal of Sex Research, 44*(4), 370-379.

DeSouza, E. (2004). Women's and men's reactions to man-to-man sexual harassment: Does the sexual orientation of the victim matter? *Sex Roles, 50*(9-10), 623-639.

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*(10), 1171-1188.

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*(10), 1171-1188.

Dodge, K. A., Gilroy, F. D., & Fenzel, L. M. (1995). Requisite management characteristics revisited: Two decades later. *Journal of Social Behavior and Personality, 10*(4), 253.

Donaldson, M. (1993). What is hegemonic masculinity? *Theory and society, 22*(5), 643-657.

Drake, C. E., Kellum, K. K., Wilson, K. G., Luoma, J. B., Weinstein, J. H., & Adams, C. H. (2010). Examining the implicit relational assessment procedure: Four preliminary studies. *The Psychological Record, 60*(1), 81-100.

Dreger, Alice. 1999. *Intersex in the age of ethics.* Hagerstown, MD: University Publishing Group.

Dubois, C. L., Knapp, D. E., Faley, R. H., & Kustis, G. A. (1998). An empirical examination of same-and other-gender sexual harassment in the workplace. *Sex Roles, 39*(9-10), 731-749.

Duehr, E. E., & Bono, J. E. (2006). Men, women, and managers: are stereotypes finally changing? *Personnel Psychology*, *59*(4), 815-846.

Dworkin, A., & MacKinnon, C. (1988). *Pornography and civil rights: a new day for women's equality*. Minneapolis, MS: Organizing Against Pornography.

Dworkin, A., (1997). *Life and death: Unapologetic writings on the continuing war against women.* New York, NY: Free Press.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. New York, NY: Psychology Press.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573.

Eagly, A. H., & Kite, M. E. (1987). Are stereotypes of nationalities applied to both women and men? *Journal of Personality and Social Psychology, 53,* 451–462.

Eagly, A. H., & Mitchell, A. A. (2004). Implications for the sociopolitical attitudes of women and men. In: M. A. Paludi (Ed.). *Praeger guide to the psychology of gender* (pp 183-206). Westport, CT: Praeger.

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, *15*(4), 543-558.

Eagly, A. H., & Wood, W. (2016). Social role theory of sex differences. *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*, 1-3.

Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Boston, MA: Harvard Business Press.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: a meta-analysis. *Psychological bulletin*, *117*(1), 125-152.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, *111*(1), 3-18.

Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are Women Evaluated More Favorably Than Men?: An Analysis of Attitudes, Beliefs, and Emotions. *Psychology of Women Quarterly*, *15*(2), 203–216.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The Developmental Social Psychology of Gender*, *12*, 174.

Ebert, I. D., Steffens, M. C., & Kroth, A. (2014). Warm, but maybe not so competent? - Contemporary implicit stereotypes of women and men in Germany. *Sex Roles*, *70*(9-10), 359-375.

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis) match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General*, *137*(2), 262 -675.

Eisler, R. M., & Skidmore, J. R. (1987). Masculine gender role stress: Scale development and component factors in the appraisal of stressful situations. *Behavior modification*, *11*(2), 123-136.

Eisler, R. M., Skidmore, J. R., & Ward, C. H. (1988). Masculine gender-role stress: predictor of anger, anxiety, and health-risk behaviors. *Journal of Personality Assessment, 52(*1), 133-141.

Elfenbein, H. A. (2007). 7 Emotion in organizations: a review and theoretical integration. *The Academy of Management Annals*, *1*(1), 315-386.

Elsaid, A., & Elsaid, E. (2012). Sex stereotyping managerial positions: A cross-cultural comparison between Egypt and the USA. *Gender in Management: An International Journal*, *27*(2), 81-99.

Epstein, M., & Ward, L. M. (2011). Exploring parent-adolescent communication about gender: Results from adolescent and emerging adult samples. *Sex Roles*, *65*(1-2), 108-118.

Eurobarometer (2016). Gender-based violence. *Special Eurobarometer Report, 449*.

European Institute for Gender Equality. (2017). Gender Equality Index 2017 Measuring gender equality in the European Union 2005-2015. Retrieved from https://eige.europa.eu/rdc/eige-publications/gender-equality-index-2017-measuring-gender-equality-european-union-2005-2015-report

Evans, L. J. (1978). Sexual harassment: Women's hidden occupational hazard. In J. R. Chapman & M. Gates (Eds.), *The victimization of women* (p. 203–223). Beverly Hills, CA: Sage.

Fairclough, N. (1992). *Discourse and social change (Vol. 10)*. Cambridge: Polity press.

Farley, L. (1978). *Sexual shakedown: The sexual harassment of women on the job.* New York, NY: McGraw-Hill.

Farrell, L., & McHugh, L. (2017). Examining gender-STEM bias among STEM and non-STEM students using the Implicit Relational Assessment Procedure (IRAP). *Journal of Contextual Behavioral Science*, *6*(1), 80-90.

Fausto-Sterling, A. (1992). Building two-way streets: The case of feminism and science. *The National Women's Studies Association Journal*, *4*(3), 336-349.

Fausto-Sterling, A. (2000). *Sexing the body: Gender politics and the construction of sexuality*. New York, NY: Basic Books.

Fazio, R. H., & Olsen, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297– 327.

Feldstein, R. (2002). The personal is (still) political: Marriage, citizenship, and women's and gender history. *Reviews in American History*, *30*(1), 106-113

Ferree, M. M., Lorber, J., & Hess, B. B. (Eds.). (1999). *Revisioning gender*. London: Rowman Altamira.

Fine, C. (2010). *Delusions of gender: How our minds, society, and neurosexism create difference*. New York, NY: WW Norton & Company.

Fine-Davis, M. (2014). *Gender roles in Ireland: Three decades of attitude change*. London,: Routledge.

Fineman, M. A. (Ed.). (2010). *Transcending the boundaries of law: Generations of feminism and legal theory*. New York, NY: Routledge.

Fineman, M. A. (Ed.). (2013). *Vulnerability: Reflections on a new ethical foundation for law and politics*. Farnham, CA: Ashgate Publishing.

Firestone, S. (1970). *The dialectic of sex: The case for feminist revolution.* New York, NY: William Morrow and Company.

Fischer, C. S., & Oliker, S. J. (1983). A research note on friendship, gender, and the life cycle. *Social Forces, 62,* 124.

Fishman, P. (1983). Interaction: The work women do. In B. Thorne, C. Kramarae, & N. Henley (Eds.), *Language, Gender and Society* (pp. 89-101). Cambridge, MA: Newbury House.

Fisk, N. M & Atun R. (2009). Systematic analysis of research underfunding in maternal and perinatal health. *British Journal of Obstetrics & Gynaecology, 116,* 347-56.

Fiske, S. T., & Stevens, L. E. (1993). *What's so special about sex? Gender stereotyping and discrimination.* New York, NY: Sage Publications.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science, 8*(4), 370-378.

Foote, W. E., & Goodman-Delahunty, J. (1999). Same-sex harassment: implications of the *Oncale* decision for forensic evaluation of plaintiffs. *Behavioral Sciences & the Law, 17*(1), 123-139.

Foroni, F., & Bel-Bahar, T. (2010). Picture-IAT versus Word-IAT: level of stimulus representation influences on the IAT. *European Journal of Social Psychology*, *40*(2), 321-337.

Foucault, M. (1986). *The care of the self: The history of sexuality,* Vol. 3. New York: Pantheon.

Frank, S.N. (2017) Feminist Historiography As If: Performativity and Representation in Feminist Histories of Rhetoric, *Rhetoric Review,* 36(3), 187-199

Franke, K. M. (1997). What's wrong with sexual harassment? *Stanford Law Review, 49,* 691–772.

Freud, S. (1905). *Three essays on the theory of sexuality. Standard edition of the complete psychological works of Sigmund Freud.* London: Hogarth Press.

Friedan, B. (1963). *The feminine mystique*. New York, NY: W.W. Norton & Company.

Fuegen, K., Biernat, M., Haines, E., & Deaux, K. (2004). Mothers and fathers in the workplace: How gender and parental status influence judgments of job-related competence. *Journal of Social Issues, 60*(4), 737-754.

Gallagher, M. (2013). Media and the representation of gender. In *The Routledge companion to media & gender* (pp. 41-49). New York, NY: Routledge.

Gardner, C. B. (1980). Passing by: Street remarks, address rights, and the urban female. *Sociological Inquiry, 50,* 21-43.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall.

Gast, A., & Rothermund, K. (2010). When old and frail is not the same: Dissociating category and stimulus effects in four implicit attitude measurement methods. *The Quarterly Journal of Experimental Psychology*, *63*(3), 479-498.

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine generics. *Sex Roles*, *23*(11-12), 629-643.

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, *101*(1), 109.

Gaunt, R. (2006). Biological essentialism, gender ideologies, and role attitudes: What determines parents' involvement in child care. *Sex Roles, 55*(7-8), 523-533.

Gaunt, R. (2009). Superordinate categorization as a moderator of mutual infrahumanization. *Group Processes & Intergroup Relations*, *12*(6), 731-746.

Gaunt, R. (2013). Ambivalent sexism and perceptions of men and women who violate gendered family roles. *Community, Work & Family, 16*(4), 401-416.

Gaunt, R., Leyens, J. P., & Demoulin, S. (2002). Intergroup relations and the attribution of emotions: Control over memory for secondary emotions associated with the ingroup and outgroup. *Journal of Experimental Social Psychology*, *38*(5), 508-514.

Gavin, A., Roche, B., & Ruiz, M. R. (2008). Competing contingencies over derived relational responding: A behavioral model of the implicit association test. *The Psychological Record, 58*(3), 427-441.

Gawronski, B., & Bodenhausen, G. V. (2015). *Theory and explanation in social psychology*. New York, NY: Guilford Press.

Gawronski, B., & De Houwer, J. (2014). *Implicit measures in social and personality psychology.* In: H. T. Reis & C. M Judd (Eds.). *Handbook of research methods in social and personality psychology* (pp 283-310). New York, NY: Cambridge University Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences, 8*(9), 404-409.

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). The developmental course of gender differentiation. *Monographs of the Society for Research in Children Development, 69*(1), 116-127.

Germain, C. B. (1983*). In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Gillham, B. (2000). *Case study research methods*. London: Bloomsbury Publishing.

Gillia, M.J. & Jacobs, A.T. (2016*). Introduction to Women's and Gender Studies: An Interdisciplinary Approach*. Oxford: Oxford University Press.

Gilligan, C. (1982). New maps of development: New visions of maturity. *American Journal of Orthopsychiatry*, *52*(2), 199.

Gino, F., Wilmuth, C. A., & Brooks, A. W. (2015). Compared to men, women view professional advancement as equally attainable, but less desirable. *Proceedings of the National Academy of Sciences, 112*(40), 12354-12359.

Glenn, E. N. (1999). The social construction and institutionalization of gender and race. *Revisioning Gender*, 3-43.

Glick, P. & Fiske, S. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *The American psychologist.* 56, 109-18.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., ... & Annetje, B. (2000). Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology, 79*(5), 763.

Glick, P., Wilk, K., & Perreault, M. (1995). Images of occupations: Components of gender and status in occupational stereotypes. *Sex Roles, 32*(9), 565-582.

Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology*, *55*(2), 178.

Golijani-Moghaddam, N., Hart, A., & Dawson, D. L. (2013). The implicit relational assessment procedure: Emerging reliability and validity data. *Journal of Contextual Behavioral Science, 2*(3-4), 105-119.

Goodwin, M.H. & Kyratzis, A. (2011). Peer language socialisation. In Alessandro Duranti, Elinor Ochs, and Bambi B. Schieffelin, (Eds.) *Handbook of Language Socialization*. (pp. 365-390). Malden, MA: Wiley-Blackwell.

Gottell, L. & Dutton, E. (2016). Sexual Violence in the 'Manosphere': Antifeminist Men's Rights Discourses on Rape. *Crime Justice Journal, 5*(2), 65-80.

Grant, J. M., Mottet, L. A., Tanis, J., Herman, J. L., Harrison, J. & Keisling, M. (2010). *National Transgender Discrimination Survey Report on Health and Health Care.* National Center for Transgender Equality and National Gay and Lesbian Task Force; Washington, DC.

Greenhaus, J. H., & Parasuraman, S. (1999). Research on work, family, and gender: Current status and future directions. *Sex Roles,* 32(9), 565-582.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*(1), 3.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17.

Grossman, A. L., & Tucker, J. S. (1997). Gender differences and sexism in the knowledge and use of slang. *Sex Roles, 37*(1), 101-110.

Groth, A. N., & Birnbaum, H. J. (1979). *Men who rape: The psychology of the offender*. New York, NY: Springer.

Grundy, I. (2000). (Re)discovering women's texts. In Vivien Jones (ed.), *Women and Literature in Britain 1700-1800*. Cambridge: Cambridge University Press.

Gutek, B. A. (1985). *Sex and the workplace.* San Francisco, CA: Jossey-Bass Inc.

Gutek, B. A., & Morasch, B. (1982). Sex-ratios, sex-role spillover, and sexual harassment of women at work. *Journal of Social Issues, 38*(4), 55-74.

Gutek, B. A., Cohen, A. G., & Konrad, A. M. (1990). Predicting social-sexual behavior at work: A contact hypothesis. *Academy of Management Journal, 33*(3), 560-577.

Habarth, J. M. (2015). Development of the heteronormative attitudes and beliefs scale. *Psychology & Sexuality*, *6*(2), 166-188.

Habarth, J. M., Makhoulian, S. C., Nelson, J. C., Todd, C. D., & Trafalis, S. (2019). Beyond Simple Differences: Moderators of Gender Differences in Heteronormativity. *Journal of Homosexuality*, 1-28.

Haider, S. (2016). The shooting in Orlando, terrorism or toxic masculinity (or both?). *Men and Masculinities*, *19*(5), 555-565.

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing … or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly, 40*(3), 353-363.

Halim, M. L., & Ruble, D. (2010). *Gender identity and stereotyping in early and middle childhood.* In: In J. C. Chrisler & D. R. McCreary (Eds.), *Handbook of gender research in psychology* (pp. 495-525). New York, NY: Springer.

Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology*, *77*(2), 350.

Hamilton, M. C. (1991). Masculine bias in the attribution of personhood: People = male, male = people. *Psychology of Women Quarterly, 15*, 393–402.

Harding, J. M. (1998). *Sex acts: Practices of femininity and masculinity*. New York, NY: Sage.

Harding, S. (1986). *The science question in feminism.* Ithaca, N.Y: Cornell University Press.

Harding, S., & Hintikka, M. B. (Eds.). (2003). *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology, and philosophy of science* (Vol. 161). Netherlands: Springer Science & Business Media.

Harper, C. (2007). *Intersex*. Oxford: Berg press.

Harris, A. P. (1990). Race and essentialism in feminist legal theory. *Stanford Law Review*, 581-616.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*(1), 113-127.

Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology, 41*(1), 87-100.

Hegarty, P. & Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965-2004. *Review of General Psychology*, 10(4), 378-402.

Hegarty, P., Ansara, Y. G., & Barker, M. J. (2018). Nonbinary gender identities. *Gender, Sex, and Sexualities: Psychological Perspectives*, 53-76.

Heider, N., Spruyt, A., & De Houwer, J. (2018). Body dissatisfaction revisited: On the importance of implicit beliefs about actual and ideal body image. *Psychologica Belgica*, *57*(4), 158.

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior, 5,* 269-298.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113-135.

Heilman, M. E., & Okimoto, T. G. (2008). Motherhood: a potential source of bias in employment decisions. *Journal of Applied Psychology, 93*(1), 189.

Heilman, M. E., Block, C. J., & Martell, R. F. (1995). Sex stereotypes: do they influence perceptions of managers? *Journal of Social Behavior and Personality*, *10*(4), 237.

Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology*, *74*(6), 935.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, *89*(3), 416.

Herek, G. M. (2004). Beyond "homophobia": Thinking about sexual prejudice and stigma in the twenty-first century. *Sexuality Research & Social Policy, 1*(2), 6-24.

Herring, S. C. (1999). The rhetorical dynamics of gender harassment on-line. *The Information Society, 15*(3), 151-167.

Herring, S. C., & Stoerger, S. (2014). Gender and (a)nonymity in computer-mediated communication. *The Handbook of Language, Gender, and Sexuality, 2,* 567-586.

Hey, V. (1996). *The company she keeps: An ethnography of girls' friendships*. Buckingham: Open University Press.

Himmelstein, M. S & Sanchez, D. T. (2016). Masculinity impediments: Internalized masculinity contributes to healthcare avoidance in men and women. *Journal of Health Psychology, 21*(7), 1283-92.

Hird, M. J. (2000). Gender's nature: Intersexuality, transsexualism and the 'sex'/'gender' binary. *Feminist Theory*, *1*(3), 347-364.

Hogue, M., & Lord, R. G. (2007). A multilevel, complexity theory approach to understanding gender bias in leadership. *The Leadership Quarterly*, *18*(4), 370-390.

Hollinger, K. (2012) Feminist film studies and race. In K. Hollinger, Feminist film studies (pp. 190-227). London: Routledge.

Holmes, C. S., Dunlap, W. P., Chen, R. S., & Cornwell, J. M. (1992). Gender differences in the learning status of diabetic children. *Journal of Consulting and Clinical Psychology, 60*(5), 698.

Holmes, J., & Meyerhoff, M. (Eds.). (2008). *The Handbook of Language and Gender* (Vol. 25). Hoboken, NY: John Wiley & Sons.

Holmes, Janet (2005). Power and discourse at work: Is gender relevant? In Lazar, Michelle (ed.), *Feminist Critical Discourse Analysis,* 31–60.

Hosoda, M., & Stone, D. L. (2000). Current gender stereotypes and their evaluative content. *Perceptual and Motor Skills, 90*(3), 1283-1294.

Houwer, J. D. (2003). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* pp. 219–244. Hoboken, NJ: Lawrence Erlbaum

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal, 6*(1), 1-55.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record, 61*(3), 465-496.

Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science*, *4*(3), 157-162.

Hyde, J. S. (1984). Children's understanding of sexist language. *Developmental Psychology, 20,* 697– 706.

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

Isaac, C., Lee, B., & Carnes, M. (2009). Interventions that affect gender bias in hiring: A systematic review. *Academic medicine: journal of the Association of American Medical Colleges*, *84*(10), 1440.

Jackson, D., Engstrom, E., & Emmers-Sommer, T. (2007). Think leader, think male and female: Sex vs. seating arrangement as leadership cues. *Sex Roles, 57*(9-10), 713-723.

Jackson, S. (2006). Interchanges: Gender, sexuality and heterosexuality: The complexity (and limits) of heteronormativity. *Feminist Theory, 7*(1), 105-121.

Jackson, S. (2018). Young feminists, feminism, and digital media. *Feminism & Psychology, 28,* 32-49.

Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: a meta-analysis. *Psychological Bulletin*, *126*(5), 703.

Jewkes, R., Flood, M., & Lang, J. (2015). From work with men and boys to changes of social norms and reduction of inequities in gender relations: a conceptual shift in prevention of violence against women and girls. *The Lancet, 385*(9977), 1580-1589.

Johnson, S. K., Murphy, S. E., Zewdie, S., & Reichard, R. J. (2008). The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational Behavior and Human Decision Processes*, *106*(1), 39-60.

Jonason, P. K. (2008). A mediation hypothesis to account for the sex difference in reported number of sexual partners: An intrasexual competition approach. *International Journal of Sexual Health*, *19*(4), 41-49.

Jones, S. R., & McEwen, M. K. (2000). A conceptual model of multiple dimensions of identity. *Journal of College Student Development, 41*(4), 405-414.

Jordan, A. (2016). Conceptualizing Backlash: (UK) Men's Rights Groups, Anti-Feminism, and Postfeminism. *Canadian Journal of Women and the Law, 28*(1), 18-44.

Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, *88*(3), 498.

Jost, J. T., Kivetz, Y., Rubini, M., Guermandi, G., & Mosso, C. (2005). System-justifying functions of complementary regional and ethnic stereotypes: Cross-national evidence. *Social Justice Research*, *18*(3), 305-333.

Juarascio, A. S., Forman, E. M., Timko, C. A., Herbert, J. D., Butryn, M., & Lowe, M. (2011). Implicit internalization of the thin ideal as a predictor of increases in weight, body dissatisfaction, and disordered eating. *Eating Behaviors, 12*(3), 207-213.

Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Psychology, 7*, 956.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*(1), 16.

Kavanagh, D., Barnes-Holmes, Y., Barnes-Holmes, D., McEnteggart, C., & Finn, M. (2018). Exploring differential trial-type effects and the impact of a read-aloud procedure on deictic relational responding on the IRAP. *The Psychological Record, 68*(2), 163-176.

Kawakami, K., Dovidio, J. F., van Kamp, S. (2007). The impact of counterstereotypic training and related correction processes on the application of stereotypes. *Group Processes Intergroup Relations, 10,* 139 –156.

Keller, J. (2005). In genes we trust: the biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology, 88*(4), 686.

Kessels, U., Rau, M., & Hannover, B. (2006). What goes well with physics? Measuring and altering the image of science. *British Journal of Educational Psychology*, *76*(4), 761-780.

Kessler, S. J., & McKenna, W. (1985). *Gender: An ethnomethodological approach*. Chicago, IL: University of Chicago Press.

Kessler, S., Ashenden , D.J., Connell , R.W. & Dowsett, G.W. (1992) Gender relations in secondary schooling. *Sociology of Education,* 58, 34-48

Kian, E. M., & Hardin, M. (2009). Framing of sport coverage based on the sex of sport writers: Female journalists counter traditional gendering of media coverage. *International Journal of Sport Communication, 2*, 185–204.

Kiesling, S. F. (1997). Power and the Language of Men. In: S. Johnson and U. Meinhof (Eds), *Language and Masculinity*, 65–85. Oxford: Blackwell.

Kimmel, M. S. (1997). Masculinity as homophobia: Fear, shame and silence in the construction of gender identity. In M. M. Gergen & S. N. Davis (Eds.), *Toward a new psychology of gender* (p. 223–242). London: Routledge.

Kimmel, M. S. (2005). *The history of men: Essays on the history of American and British masculinities*. Albany: State University of New York Press.

Kimmel, M., & Messner, M. (1995). *Men's lives.* Needham Heights, MA: Allyn & Bacon.

King, R. (2016). Gender differences in motivation, engagement and achievement are related to students' perceptions of peer—but not of parent or teacher—attitudes toward school. *Learning and Individual Differences, 52,* 60-71.

Kinzler, K. D., Shutts, K., & Correll, J. (2010). Priorities in social categories. *European Journal of Social Psychology, 40*(4), 581-592.

Kissling, E. A. (1991). Street harassment: The language of sexual terrorism. *Discourse & Society, 2*(4), 451-460.

Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion theory. *Psychology of Women Quarterly*, *11*(1), 83-96.

Kitzinger, C. (2005). Heteronormativity in action: Reproducing the heterosexual nuclear family in after-hours medical calls. *Social Problems, 52*(4), 477-498.

Koenig A.M., Eagly A.H., Mitchell A.A., Ristikari T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychology Bulletin, 137*(4), 616-42.

Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, *9*, 1086.

Kramarae, C. (1992). The condition of patriarchy. In C. Kramarae and D. Spender (eds.), *The knowledge explosion: Generations of feminist scholarship,* (pp. 125-149) New York, NY: Routledge.

Kunst, J. R., Bailey, A., Prendergast, C., & Gundersen, A. (2018). Sexism, rape myths and feminist identification explain gender differences in attitudes toward the #metoo social media campaign in two countries. *Media Psychology*. Advance online publication.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., & Banaji, M. R. (2018). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist. American Psychologist, 74*(5), 569–586.

Lakoff, R. (1975). Language and woman's place. *Language in Society, 2*(1). 45-80.

Lambdin, J. R., Greer, K. M., Jibotian, K. S., Wood, K. R., & Hamilton, M. C. (2003). The animal male hypothesis: Children's and adult's beliefs about the sex of non sex-specific stuffed animals. *Sex Roles, 48,* 471–482.

Lane, K. A., Goh, J. X., & Driver-Linn, E. (2012). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles*, *66*(3), 220-234.

Langford, T., & MacKinnon, N. J. (2000). The affective bases for the gendering of traits: Comparing the United States and Canada. *Social Psychology Quarterly, 63*(1), 34.

Laslett, B., & Brenner, J. (1989). Gender and social reproduction: Historical perspectives. *Annual Review of Sociology*, *15*(1), 381-404

Latu, I. M., Stewart, T. L., Myers, A. C., Lisco, C. G., Estes, S. B., & Donahue, D. K. (2011). What we "say" and what we "think" about female managers: Explicit versus implicit associations of women with success. *Psychology of Women Quarterly, 35*(2), 252-266.

Leaper, C., & Friedman, C. K. (2007). The socialization of gender. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 561-587). New York, NY: Guilford Press.

Leaper, C., Breed, L., Hoffman, L., & Perlman, C. A. (2002). Variations in the gender-stereotyped content of children's television cartoons across genres. *Journal of Applied Social Psychology, 32*(8), 1653-1662.

Lee, K., Gizzarone, M., & Ashton, M. C. (2003). Personality and the likelihood to sexually harass. *Sex Roles*, *49*(1-2), 59-69.

Leidig, M. W. (1992). The continuum of violence against women: Psychological and physical consequences. *Journal of American College Health, 40*(4), 149-155.

Levant, R. F. (2011). Research in the psychology of men and masculinity using the gender role strain paradigm as a framework. *American Psychologist, 66*(8), 765-776.

Levant, R. F., & Pollack, W. S. (Eds.). (1995). *A new psychology of men.* New York, NY: Basic Books.

Levinson, R. M. (1982). Sex discrimination and employment practices: An experiment with unconventional job inquiries. In R. Kahn-Hut, A. N. Daniels, & R. Colvard (Eds.), *Women and work: Problems and perspectives* (pp. 54-65). New York, NY: Oxford University Press.

Levy, S.R., & Killen, M. (2008). *Intergroup attitudes and relations in childhood through adulthood*. Oxford: Oxford University Press.

Leyens, J. Ph., Cortes, B. P., Demoulin, S., Dovidio, J. F., Fiske, S. T., Gaunt, R., et al. (2003). Emotional prejudice, essentialism, and nationalism. *European Journal of Social Psychology, 33,* 704-717.

Leyens, J. Ph., Paladino, P. M., Rodriguez, R. T., Vaes, J., Demoulin, S., Rodriguez, A. P., et al. (2000). The emotional side of prejudice: The role of secondary emotions. *Personality and Social Psychology Review, 4,* 186-197.

Leyens, J. Ph., Rodriguez, A. P., Rodriguez, R. T., Gaunt, R., Paladino, P. M., Vaes, J., et al. (2001). Psychological essentialism and the attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology,* 31, 395-411.

Litosseliti, L., & Sunderland, J. (2002). Gender identity and discourse analysis. London: John Benjamins Publishing.

Livia, A., & Hall, K. (1997). *Queerly phrased: Language, gender, and sexuality*. Oxford: Oxford University Press.

Locke, B. D., & Mahalik, J. R. (2005). Examining masculinity norms, problem drinking, and athletic involvement as predictors of sexual aggression in college men. *Journal of Counseling Psychology, 52*(3), 279.

Lorber, J. (1994). *Paradoxes of gender*. Westport, CT: Yale University Press.

Lorber, J. (1997). *Gender and the social construction of illness.* Thousand Oaks, CA: Sage.

Lorde, A. (1986). *Our dead behind us: Poems.* New York, NY: WW Norton.

Losert, A. (2008). Coping with workplace heteronormativity among lesbian employees: A German study. *Journal of Lesbian Studies, 12*(1), 47-58.

Lucal, B. (1999). What it means to be gendered me: Life on the boundaries of a dichotomous gender system. *Gender & Society, 13*(6), 781-797

Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology, 91*(4), 777.

Maass, A., Cadinu, M., Guarnieri, G., & Grasselli, A. (2003). Sexual harassment under social identity threat: The computer harassment paradigm. *Journal of Personality and Social Psychology, 85*(5), 853.

Maccoby, Eleanor E. (1998). *The two sexes: Growing up apart, coming together*. Boston, MA: Harvard University Press.

Mackay, F. (2015). *Radical Feminism: Feminist activism in movement.* London: Palgrave MacMillan Press.

MacKinnon, C.A. (1983). Feminism, Marxism, method, and the state: Toward feminist jurisprudence. *Signs, 8*(4), 635-658.

MacKinnon, C. A. (2005). *Women's lives, men's law.* Cambridge, MA: Harvard University Press.

MacKinnon, C. A., & MacKinnon, C. A. (1979). *Sexual harassment of working women: A case of sex discrimination.* Hartford, CT: Yale University Press.

Maddox, B. (2002). *Rosalind Franklin: The dark lady of DNA*. London: Penguin Publishers.

Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: agentic and communal differences. *Journal of Applied Psychology*, *94*(6), 1591.

Mahalik, J. R., Talmadge, W. T., Locke, B. D., & Scott, R. P. (2005). Using the Conformity to Masculine Norms Inventory to work with men in a clinical setting. *Journal of Clinical Psychology, 61*(6), 661-674.

Malamuth, N. M., & Dean, K. E. (1989). Attraction to sexual aggression. *The Journal of Sex Research*, *26*(1), 26-49.

Marsh, H. W., & Kleitman, S. (2003). School athletic participation: Mostly gain with little pain. *Journal of Sport and Exercise Psychology, 25*(2), 205-228.

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin, 128*(6), 903-928.

Massey, S. G. (2009). Polymorphous prejudice: Liberating the measurement of heterosexuals' attitudes toward lesbians and gay men. *Journal of Homosexuality, 56*(2), 147-172.

Masters, N. T., Casey, E., Wells, E. A., & Morrison, D. M. (2013). Sexual scripts among young heterosexually active men and women: Continuity and change. *Journal of Sex Research, 50*(5), 409-420.

Mavin, S. (2006). Venus envy: problematizing solidarity behaviour and queen bees. *Women in Management Review, 21*(4) 264-76.

Mavin, S., Williams, J., Grandy, G. (2014). Negative intra-gender relations between women: Friendship, competition and female misogyny. In Kumra, Savita, Simpson, Ruth & Burke, R.J. (Ed.s). *The Oxford Handbook of Gender in Organizations.* Oxford: Oxford University Press.

McCreary, D. R. (1994). The male role and avoiding femininity. *Sex Roles, 31*(9-10), 517-531.

McLaughlin, H., Uggen, C., & Blackstone, A. (2012). Sexual harassment, workplace authority, and the paradox of power. *American Sociological Review, 77*(4), 625-647.

McPhail, B. A. (2016). Feminist framework plus: Knitting feminist theories of rape etiology into a comprehensive model. *Trauma, Violence, & Abuse*, *17*(3), 314-329.

Mead, M. (1935). *Sex and Temperament in Three Primitive Societies.* Oxford: William Morrow.

Menegatti, M., & Rubini, M. (2017). Gender bias and sexism in language. In J. P. Forgas, J. Laszlo, & O. Vincze (Eds.), *Social cognition and communication* (pp. 331–346). New York: Psychology Press.

Meyer, M., & Gelman, S. A. (2016). Gender essentialism in children and parents: Implications for the development of gender stereotyping and gender-typed preferences. *Sex Roles, 75*(9-10), 409-421.

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, *107*(3), 631.

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science stereotypes: a meta-analysis of 5 decades of US draw-a-scientist studies. *Child Development*, *89*(6), 1943-1955.

Miller, D. T., Taylor, B., & Buck, M. L. (1991). Gender gaps: Who needs to be explained?. *Journal of Personality and Social Psychology*, *61*(1), 5.

Miller, P.H. & Scholnick, E.K. (2015). Feminist theory and contemporary developmental psychology: The case of children's executive function. *Feminism & Psychology, 25*(3), 266-284.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*(3), 455-465.

Morton, T. A., Postmes, T., Haslam, S. A., & Hornsey, M. J. (2009). Theorizing gender in the face of social change: Is there anything essential about essentialism? *Journal of Personality and Social Psychology, 96*(3), 653.

Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men & Masculinity*, *11*(2), 140.

Moyer, R. (1997). Covering gender on memory's front page: Men's prominence and women's prospects. *Sex Roles, 37*, 595–618.

Muehlenhard, C. L., & McCoy, M. L. (1991). Double standard/double bind: The sexual double standard and women's communication about sex. *Psychology of Women Quarterly, 15*(3), 447-461.

Muehlenhard, C. L., Goggins, M. F., Jones, J. M., & Satterfield, A. T. (1991). Sexual violence and coercion in close relationships. *Sexuality in Close Relationships, 12,* 155-175.

Mueller, C. M. (1986). Nurturance and mastery: Competing qualifications for women's access to high public office? In: G. Moore and G. D. Spitze (Eds.), *Women and politics: Activism, attitudes, and office-holding*. Greenwich, CT: JAI Press.

Mulvey, L. (1975). Visual pleasure and narrative cinema. *Screen, 8,* 6-18.

Murjan, S., & Bouman, W. P. (2015). Transgender—living in a gender different from that assigned at birth. In In C. Richards & M. Barker (Eds.). *The Palgrave Handbook of the Psychology of Sexuality and Gender* (pp. 198-215). London: Palgrave Macmillan.

Murnen, S. K., Wright, C., & Kaluzny, G. (2002). If "boys will be boys," then girls will be victims? A meta-analytic review of the research that relates masculine ideology to sexual aggression. *Sex Roles, 46*(11-12), 359-375.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., ... & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, *51*(2), 77.

Neumark, D., Bank, R. J., & Van Nort, K. D. (1996). Sex discrimination in restaurant hiring: An audit study. *The Quarterly Journal of Economics*, *111*(3), 915-941.

Newport, F. (2001, February, 21). Americans see women as emotional and affectionate, men as more aggressive: Gender specific stereotypes persist in recent Gallup poll. Retrieved February 23, 2019, from http://institution.gallup.com.turing.library.northwestern.edu/content/ default.aspx?ci-1978.

Ng, S. H. (1990). Androcentric coding of man and his in memory by language users. *Journal of Experimental Social Psychology, 26,* 455–464.

Ng, S. H. (2007). Language-based *discrimination*: Blatant and subtle forms. *Journal of Language and Social Psychology*, *26*, 106–122.

Nicholson, E., & Barnes-Holmes, D. (2012). Developing an implicit measure of disgust propensity and disgust sensitivity: Examining the role of implicit disgust propensity and sensitivity in obsessive-compulsive tendencies. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(3), 922-930.

Nicholson, E., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record, 62*(2), 263-277.

Nicholson, E., Dempsey, K., & Barnes-Holmes, D. (2014). The role of responsibility and threat appraisals in contamination fear and obsessive-compulsive tendencies at the implicit level. *Journal of Contextual Behavioral Science*, *3*(1), 31-37.

Nicholson, E., McCourt, A., & Barnes-Holmes, D. (2013). The Implicit Relational Assessment Procedure (IRAP) as a measure of obsessive beliefs in relation to disgust. *Journal of Contextual Behavioral Science, 2*(1-2), 23-30.

Nien, C. L., & Duda, J. L. (2008). Antecedents and consequences of approach and avoidance achievement goals: A test of gender invariance. *Psychology of Sport and Exercise*, *9*(3), 352-372.

Nieva, V. F., & Gutek, B. A. (1981). *Women and work: A psychological perspective.* New York, NY: Praeger Publishers.

Nolan, J., Murphy, C., & Barnes-Holmes, D. (2013). Implicit relational assessment procedure and body-weight bias: Influence of gender of participants and targets. *The Psychological Record, 63*(3), 467-488.

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science, 16*(2), 65-69.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social cognition, 19*(6), 625-666.

O'Connor, P. (1992) *Friendships between women.* New York, NY: Guildford Press.

Oakley, A. (2016). *Sex, gender and society*. London: Routledge.

Olson, K. R., Key, A. C., & Eaton, N. R. (2015). Gender cognition in transgender children. *Psychological Science, 26*(4), 467-474.

O'Neil, J. M., Helms, B. J., Gable, R. K., David, L., & Wrightsman, L. S. (1986). Gender-Role Conflict Scale: College men's fear of femininity. *Sex Roles*, *14*(5-6), 335-350.

Oswald, D. L., & Lindstedt, K. (2006). The content and function of gender self-stereotypes: An exploratory investigation. *Sex Roles, 54*(7-8), 447-458.

Paladino, M. P., Vaes, J., Castano, E., Demoulin, S., & Leyens, J. P. (2004). Emotional infra-humanization in intergroup relations: The role of national identification in the attribution of primary and secondary emotions to Italians and Germans. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*. *22*(4-5), 519–536.

Paluck, E. L., Ball, L., Poynton, C., & Sieloff, S. (2010). *Social norms marketing aimed at gender based violence: A literature review and critical assessment.* New York, NY: International Rescue Committee.

Paludi, M. A., Paludi, M., & Paludi, C. A. (Eds.). (2003). *Academic and workplace sexual harassment: A handbook of cultural, social science, management, and legal perspectives*. New York, NY: Praeger.

Parker, R. B., Larkin, T., & Cockburn, J. (2018). Gender bias in medical images affects students' implicit but not explicit gender attitudes. *AERA Open*, *4*(3), 1-7.

Parsons, T. (1955). Family structure and the socialization of the child. *Family, Socialization and Interaction Process*, 35-131.

Payne, K., & Lundberg, K. (2014). The affect misattribution procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass, 8*(12), 672-686.

Pederson, S. (1990). Gender, welfare, and citizenship in Britain during the Great War. *The American Historical Review, 95,* 229-245.

Pereira, C., Vala, J., & Leyens, J. P. (2009). From infra-humanization to discrimination: The mediation of symbolic threat needs egalitarian norms. *Journal of Experimental Social Psychology*, *45*(2), 336-344.

Phelan, J. E., & Rudman, L. A. (2010). Prejudice toward female leaders: Backlash effects and women's impression management dilemma. *Social and Personality Psychology Compass*, *4*(10), 807-820.

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly*, *32*(4), 406-413.

Pilcher, J. and Whelehan, I. (2004). *Fifty key concepts in Gender Studies.* London: Sage.

Pina, A., Gannon, T. A., & Saunders, B. (2009). An overview of the literature on sexual harassment: Perpetrator, theory, and treatment issues. *Aggression and Violent Behavior, 14*(2), 126-138.

Pleck, J. H. (1981). *The myth of masculinity.* Cambridge, MA: MIT Press.

Potter, E. F., & Rosser, S. V. (1992). Factors in life science textbooks that may deter girls' interest in science. *Journal of Research in Science Teaching*, *29*(7), 669-686.

Potter, J., & Wetherell, M. (1994). Analyzing discourse. *Analyzing Qualitative Data,* 47-66.

Powell, G. N., Butterfield, D. A., & Parent, J. D. (2002). Gender and managerial stereotypes: have the times changed? *Journal of Management, 28*(2), 177-193.

Pratto, F., Korchmaros, D. J. & Hegarty, P. (2007). When race and gender go without saying. *Social Cognition, 25*(2), 221-247.

Prentice, D. A., & Miller, D. T. (2006). Essentializing differences between women and men. *Psychological Science*, *17*(2), 129-135.

Pryor, J. B. (1987). Sexual harassment proclivities in men. *Sex Roles*, *17*(5-6), 269-290.

Pryor, J. B., Giedd, J. L., & Williams, K. B. (1995). A social psychological model for predicting sexual harassment. *Journal of Social Issues, 51*(1), 69-84.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90.

Quinn, B. A. (2002). Sexual harassment and masculinity: The power and meaning of "girl watching". *Gender & Society, 16*(3), 386-402.

Ragan, S. L., Bystrom, D. Kaid, L.L & Beck, C. (1996). *The lynching of language: Gender, politics, and power in the Hill-Thomas hearings.* Urbana: University of Illinois Press.

Reilly, M. E., Lott, B., Caldwell, D., & DeLuca, L. (1992). Tolerance for sexual harassment related to self-reported sexual victimization. *Gender & Society, 6*(1), 122-138.

Remue, J., De Houwer, J., Barnes-Holmes, D., Vanderhasselt, M. A., & De Raedt, R. (2013). Self-esteem revisited: Performance on the implicit relational assessment procedure as a measure

of self-versus ideal self-related cognitions in dysphoria. *Cognition & Emotion*, *27*(8), 1441-1449.

Remue, J., Hughes, S., De Houwer, J., & De Raedt, R. (2014). To be or want to be: Disentangling the role of actual versus ideal self in implicit self-esteem. *Journal of Personality and Social Psychology, 98,* 645-658.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, *111*(12), 4403-4408.

Rhodes, G., Hickford, C., & Jeffery, L. (2000). Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive?. *British Journal of Psychology*, *91*(1), 125-140.

Rhodes, G., Jaquet, E., Jeffery, L., Evangelista, E., Keane, J., & Calder, A. J. (2011). Sex-specific norms code face identity. *Journal of Vision*, *11*(1), 1-1.

Rich, A. (1980). Compulsory heterosexuality and lesbian existence. *Signs: Journal of Women in Culture and Society*, *5*(4), 631-660.

Richards, C., & Barker, M. J. (2015). *The Palgrave handbook of the psychology of sexuality and gender.* London: Palgrave MacMillan.

Richards, C., Bouman, W. P., & Barker, M. J. (Eds.). (2017). *Genderqueer and non-binary genders*. London: Palgrave Macmillan

Richardson, K., & Norgate, S. H. (2015). Does IQ really predict job performance?. *Applied Developmental Science*, *19*(3), 153-169.

Richardson, S. S. (2013). *Sex itself: the search for male and female in the human genome*. Chicago, IL: University of Chicago Press.

Ridgeway, C. L. (1997). Interaction and the conservation of gender inequality: Considering employment. *American Sociological Review*, *12,* 218-235.

Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. Oxford: Oxford University Press.

Ridgeway, C. L., & Correll, S. J. (2004). Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & Society*, *18*(4), 510-531.

Ridgeway, C. L., & Smith-Lovin, L. (1999). The gender system and interaction. *Annual Review of Sociology*, *25*(1), 191-216.

Risman, B. J. (1998). *Gender vertigo: American families in transition*. New York, NY: Yale University Press.

Risman, B. J. (2004). Gender as a social structure: Theory wrestling with activism. *Gender & Society*, *18*(4), 429-450.

Robinson, L. F., & Reis, H. T. (1989). The effects of interruption, gender, and status on interpersonal perceptions. *Journal of Nonverbal Behavior, 13*(3), 141-153.

Roefs, A., Huijding, J., Smulders, F. T., MacLeod, C. M., de Jong, P. J., Wiers, R. W., & Jansen, A. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin, 137*(1), 149.

Roen K. (2015) Intersex/DSD. In: Richards C., Barker M.J. (Eds) *The Palgrave handbook of the psychology of sexuality and gender* (pp. 239-278). London: Palgrave Macmillan.

Rohmann, A., Niedenthal, P. M., Brauer, M., Castano, E., & Leyens, J. P. (2009). The attribution of primary and secondary emotions to the in-group and to the out-group: The case of equal status countries. *The Journal of Social Psychology*, *149*(6), 709-730.

Rosette, A. S., & Tost, L. P. (2010). Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. *Journal of Applied Psychology*, *95*(2), 221 - 235.

Ross, K.., & Byerly, C. M. (Eds.). (2008). *Women and media: International perspectives*. New York, NY: John Wiley & Sons.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the implicit association test: dissociating salience from associations. *Journal of Experimental Psychology: General, 133*(2), 139.

Rubin, D. (1993). *Gender influences: Reading student texts.* Carbondale, IL: SIU Press.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, *74*(3), 629.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*(4), 743-762.

Rudman, L. A., & Mescher, K. (2012). Of animals and objects: Men's implicit dehumanization of women and likelihood of sexual aggression. *Personality and Social Psychology Bulletin, 38*(6), 734-746.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81*(5), 856-868.

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin*, *27*(9), 1164-1178.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165-179.

Sanders, T. (2016). Inevitably violent? Dynamics of space, governance, and stigma in understanding violence against sex workers. In *Problematizing Prostitution: Critical Research and Scholarship* (pp. 93-114). Birmingham: Emerald Group Publishing.

Sapiro, V. (1983). *The political integration of women: Roles, socialization, and politics.* Urbana: University of Illinois Press.

Sbraga, T. P., & O'donohue, W. (2000). Sexual harassment. *Annual Review of Sex Research*, *11*(1), 258-285.

Scanlon, G., McEnteggart, C., Barnes-Holmes, Y., & Barnes-Holmes, D. (2014). Using the implicit relational assessment procedure (IRAP) to assess implicit gender bias and self-esteem in typically-developing children and children with ADHD and with dyslexia. *Behavioral Development Bulletin*, *19*(2), 48 -59.

Schein, V. E. (1975). Relationships between sex role stereotypes and requisite management characteristics among female managers. *Journal of Applied Psychology*, *60*(3), 340.

Scheuringer, A., Wittig, R., & Pletzer, B. (2017). Sex differences in verbal fluency: the role of strategies and instructions. *Cognitive Processing*, *18*(4), 407-417.

Schmitz, R. M., & Kazyak, E. A. (2016). Masculinities in cyberspace: An analysis of portrayals of manhood in men's rights activistwebsites. *Social Sciences, 5*(2), 22-38.

Schmitz, S., & Höppner, G. (2014). Neurofeminism and feminist neurosciences: a critical review of contemporary brain research. *Frontiers in Human Neuroscience*, *8*, 546.

Schneider, C. M & Bos, L. A. (2014). Measuring Stereotypes of Female Politicians. *Political Psychology, 3*(2), 245-266.

Schneider, C., Cerwenka, S., Nieder, T. O., Briken, P., Cohen-Kettenis, P. T., De Cuypere, G., ... & Richter-Appelt, H. (2016). Measuring gender dysphoria: a multicenter examination and comparison of the Utrecht gender dysphoria scale and the gender identity/gender dysphoria questionnaire for adolescents and adults. *Archives of Sexual Behavior, 45*(3), 551-558.

Seal, M. (2016) Critical Realism's Potential Contribution to Critical Pedagogy and Youth and Community Work: Human Nature, Agency and Praxis Revisited*, Journal of Critical Realism, 15*(3), 263-276.

Sedgewick, E. (1991). *Epistemology of the closet.* Berkeley, CA: University of California Press.

Serano, J. M. (2010). The case against autogynephilia. *International Journal of Transgenderism*, *12*(3), 176-187.

Shaw, E., Hegewisch, A., & Hess, C. Sexual Harassment and Assault at Work: Understanding the Costs. Retrieved February 20, 2020, from https://iwpr.org/publications/sexual-harassment-work-cost/.

Sharpe, P. (Ed.). (2002). *Women, gender and labour migration: Historical and cultural perspectives*. London: Routledge.

Shepard, D., Gurewich, D., Lwin, A., Reed, G., & Silverman, M. (2015). Suicide and suicidal attempts in the United States: costs and policy implications. *Suicide and Life-Threatening Behavior, 46*(3), 352-362

Shinar, E. H. (1975). Sexual stereotypes of occupations. *Journal of Vocational Behavior*, *7*(1), 99-111.

Sibley, C. G., Wilson, M. S., & Duckitt, J. (2007). Antecedents of men's hostile and benevolent sexism: The dual roles of social dominance orientation and right-wing authoritarianism. *Personality and Social Psychology Bulletin, 33*(2), 160-172.

Siebler F, Sabelus S, Bohner G. (2008). A refined computer harassment paradigm: Validation, and test of hypotheses about target characteristics. *Psychology of Women Quarterly, 32*, 22–35.

Signorielli, N., & Kahlenberg, S. (2001). Television's world of work in the nineties. *Journal of Broadcasting & Electronic Media*, *45*(1), 4-22.

Silveira, J. (1980). Generic masculine words and thinking. *Women's Studies International Quarterly, 3,* 165-178.

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, *26*(11), 1329-1342.

Skewes, L., Fine, C., & Haslam, N. (2018). Beyond Mars and Venus: The role of gender essentialism in support for gender inequality and backlash. *PloS one*, *13*(7), e0200921.

Smart, C. (2002). *Feminism and the power of law*. London: Routledge.

Smiler, A. P. (2004). Thirty years after the discovery of gender: Psychological concepts and measures of masculinity. *Sex Roles*, *50*(1-2), 15-26.

Smiler, A. P. (2006). Conforming to masculine norms: Evidence for validity among adult men and women. *Sex Roles, 54*(11-12), 767-775.

Smiler, A. P. (2006). Living the image: A quantitative approach to delineating masculinities. *Sex Roles, 55*(9-10), 621-632.

Smiler, A. P., & Gelman, S. A. (2008). Determinants of gender essentialism in college students. *Sex Roles*, *58*(11-12), 864-874.

Smith, C. J., Noll, J. A., & Bryant, J. B. (1999). The effect of social context on gender self-concept. *Sex Roles, 40*(5-6), 499-512.

Smith, D. E. (1987). *The everyday world as problematic: A feminist sociology*. Toronto: University of Toronto Press.

Smith, K, & Moen, P. (1988). Passage through midlife: Women's changing family roles and economic well-being. *The Sociological Quarterly, 29,* 503-524.

Smith, L. (2008). Heritage, gender and identity. *The Ashgate Research Companion to Heritage and Identity*, 159-178.

Smyth, F. L., & Nosek, B. A. (2015). On the gender–science stereotypes held by scientists: Explicit accord with gender-ratios, implicit accord with scientific identity. *Frontiers in Psychology*, *6*, 415.

Snell Jr, W. E. (2013). Development and validation of the Masculine Behavior Scale (MBS): A measure of behaviours stereotypically attributed to males vs. females. *Sex Roles, 21,* 749-767.

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes. What do they signify? *Psychology of Women Quarterly,* 24 (1), 44–62.

Spence, J. T., & Helmreich, R. L. (1972). The Attitudes Toward Women Scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. *Psychology of Women Quarterly, 21*(1), 17-34.

Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* Austin, TX: University of Texas Press.

Spender, D. (1985). *For the record The making and meaning of feminist knowledge.* London: Women's Press.

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*(3), 404-420.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. *Social Communication,* 163-187.

Stamarksi, C. S., & Hing, L. S. S. (2015). Gender inequalities in the workplace: The effects of organizational structures, processes, practices, and decision makers' sexism. *Frontiers in Psychology, 6,* 140-168.

Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles*, *64*(5-6), 324-335.

Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für experimentelle Psychologie, 48*(2), 123-134.

Steffens, M. C., & Schulze König, S. (2006). Predicting spontaneous big five behavior with implicit association tests. *European Journal of Psychological Assessment*, *22*(1), 13-20.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology*, *41*(1), 37.

Strang, E., & Peterson, Z. D. (2017). Unintentional misreporting on self-report measures of sexually aggressive behavior: An interview study. *The Journal of Sex Research*, *54*(8), 971-983.

Stroessner, S. J. (1996). Social categorization by race or sex: Effects of perceived non-normalcy on response times. *Social Cognition,1*(4), 247-276.

Sutherland, C. M. (2002). Feminist historiography: Research methods in rhetoric. *Rhetoric Society Quarterly*, *32*(1), 109-122.

Swank, E., & Fahs, B. (2017). College students, sexualities identities, and participation in political marches. *Sexuality Research and Social Policy, 14*(2), 122-132.

Swim J. K., Aikin K. J., Hall W. S. & Hunter B. A. (1995). Sexism and racism: Old-fashioned and modern prejudice. *Journal of Personality and Social Psychology, 68*, 199-214.

Switzer, T. (1990). The impact of generic word choices: An empirical investigation of age- and sex-related differences. *Sex Roles, 22*(172), 69–81.

Tangri, S. S., & Hayes, S. M. (1997). Theories of sexual harassment. In W. O'Donohue (Ed.), *Sexual harassment: Theory, research, and treatment* (pp 112–128). Boston, MA: Allyn & Bacon.

Tannen, D. (1994). *Gender and discourse.* Oxford: Oxford University Press.

Tepper, B. J., Brown, S. J., & Hunt, M. D. (1993). Strength of subordinates' upward influence tactics and gender congruency effects. *Journal of Applied Social Psychology*, *23*(22), 1903-1919.

Terman, L. M. (2013). *Sex and personality studies in masculinity and femininity*. London: Read Books Ltd.

Testa, R. J., Habarth, J., Peta, J., Balsam, K., & Bockting, W. (2015). Development of the gender minority stress and resilience measure. *Psychology of Sexual Orientation and Gender Diversity*, *2*(1), 65.

Thane, P. (1992). The history of the gender division of labour in Britain: Reflections on "'herstory' in accounting: The first eighty years". *Accounting, Organizations and Society*, *17*(3-4), 299-312.

Thompson, T. L., & Zerbinos, E. (1995). Gender roles in animated cartoons: Has the picture changed in 20 years? *Sex Roles, 32*(9-10), 651-673.

Thornham, S. (2007). *Women, feminism and media*. Edinburgh: Edinburgh University Press.

Tibboel, H., De Houwer, J., Dirix, N., & Spruyt, A. (2017). Beyond associations: Do implicit beliefs play a role in smoking addiction? *Journal of Psychopharmacology*, *31*(1), 43-53.

Timko, C. A., England, E. L., Herbert, J. D., & Forman, E. M. (2010). The Implicit Relational Assessment Procedure as a measure of self-esteem. *The Psychological Record, 60*(4), 679-698.

Timmins, L., Barnes-Holmes, D., & Cullen, C. (2016). Measuring implicit sexual response biases to nude male and female pictures in androphilic and gynephilic men. *Archives of Sexual Behavior, 45*(4), 829-841.

Tinsley, C. H., Howell, T. M., & Amanatullah, E. T. (2015). Who should bring home the bacon? How deterministic views of gender constrain spousal wage preferences. *Organizational Behavior and Human Decision Processes, 126*, 37-48.

Turner, S. (2000). Vindicating masculinity: the fate of promoting gender equality. *Forced Migration Review, 9*, 2-18.

Twenge, J. M. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex Roles*, *36*(5-6), 305-325.

Twenge, J. M., & Campbell, S. M. (2008). Generational differences in psychological traits and their impact on the workplace. *Journal of Managerial Psychology, 23*(8), 862-877.

Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes, 104*(2), 207-223.

Unger, R. K. (1979). Toward a redefinition of sex and gender. *American Psychologist, 34*(11), 1085-1102.

United Nations. (2018). UN Annual Report on Women. Accessed on 24th February 2019 at http://www2.unwomen.org/-/media/annual%20report/attachments/sections/library/un-women-annual-report-2017-2018-en.pdf

Utaker, A. (1974). On the binary opposition. *Linguistics, 12*(134), 73-93.

Vaes, J., Paladino, M. P., & Leyens, J. Ph. (2004). Perspective taking in an intergroup context and the use of uniquely human emotions: Drawing an E on your forehead. *International Review of Social Psychology, 17*(3), 5-26.

Vahey, N. A., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record, 59*(3), 371-387.

Vahey, N. A., Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 59-65.

Van Anders, S. M. (2015). Beyond sexual orientation: Integrating gender/sex and diverse sexualities via sexual configurations theory. *Archives of Sexual Behavior*, *44*(5), 1177-1213.

Van Anders, S. M., Schudson, Z. C., Abed, E. C., Beischel, W. J., Dibble, E. R., Gunther, O. D., ... & Silver, E. R. (2017). Biological sex, gender, and public policy. *Policy Insights from the Behavioral and Brain Sciences, 4*(2), 194-201.

Van Berkel, L., Molina, L. E., & Mukherjee, S. (2017). Gender asymmetry in the construction of American national identity. *Psychology of Women Quarterly*, *41*(3), 352-367.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.

Van Gelder, L. & Sharpe, K. (2009). Women and girls as upper palaeolithic cave 'artists': Deciphering the sexes of finger fluters in Rouffignac Cave. *Oxford Journal of Archaeology, 28*(4). 323-333.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70.

Vasey, P. L., & Bartlett, N. H. (2007). What can the Samoan "fa'afafine" teach us about the Western concept of gender identity disorder in childhood? *Perspectives in Biology and Medicine, 50*(4), 481-490.

Verniers, C., Vala, J. (2018). Correction: Justifying gender discrimination in the workplace: The mediating role of motherhood myths. *PLoS One,* 13(7), e0201150.

Vetter, L. (2010). *Gender and women's leadership: A reference handbook.* Thousand Oaks, CA: Sage.

Viki, G. T., & Abrams, D. (2003). Infra-humanization: Ambivalent sexism and the attribution of primary and secondary emotions to women. *Journal of Experimental Social Psychology*, *39*(5), 492-499.

Viki, T. G., & Calitri, F. (2009). Infrahuman outgroup or suprahuman ingroup: The role of nationalism and patriotism in the infrahumanization of outgroups. *European Journal of Social Psychology, 38,* 1054–1061.

Waechter, R., & Ma, V. (2015). Sexual violence in America: Public funding and social priority. *American Journal of Public Health*, *105*(12), 2430-2437.

Walby, S. (2009). The cost of domestic violence: up-date 2009. Lancaster: Lancaster University Press.

Ward, J., & Schneider, B. (2009). The reaches of heteronormativity: An introduction. *Gender & Society, 23*(4), 433-439.

Ward, T., Hudson, S. M., Johnston, L., & Marshall, W. L. (1997). Cognitive distortions in sex offenders: An integrative review. *Clinical Psychology Review, 17*(5), 479-507.

Waylen, G., Celis, K., Kantola, J., & Weldon, L. (2013). *The Oxford handbook of gender and politics*. Oxford: Oxford University Press.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society*, *1*(2), 125-151.

Wetherall, M. (1986) Linguistic repertoires and literary criticism: new directions for a social psychology of gender. In: S. Wilkinson (Ed.), *Feminist Social Psychology*. Milton Keynes: Open University Press.

Wilde, A., & Diekman, A. B. (2005). Cross-cultural similarities and differences in dynamic stereotypes: A comparison between Germany and the United States. *Psychology of Women Quarterly*, *29*(2), 188-196.

Williams, J. E., & Best, D. L. (1990). *Cross-cultural research and methodology series, Vol. 6. Measuring sex stereotypes: A multination study (Rev. ed.).* Thousand Oaks, CA: Sage Publications, Inc.

Willis, P. (1977). *Learning to labour.* Westmead, England: Saxon House.

Willness, C. R., Steel, P., & Lee, K. (2007). A meta-analysis of the antecedents and consequences of workplace sexual harassment. *Personnel Psychology, 60*(1), 127–162.

Witt, S. D. (1997). Boys will be boys, and girls will be… hard to find: Gender representation in third grade basal readers. *Education and society*, *15*(1), 47-57.

Wittig, M. (1985). The mark of gender. *Feminist Issues,* 5(2), 3–12.

Wolsiefer, K., Westfall, J., & Judd, C. M. (2017). Modeling stimulus variation in three common implicit attitude tasks. *Behavior Research Methods*, *49*(4), 1193-1209.

Wood, W., Christensen, P. N., Hebl, M. R., & Rothgerber, H. (1997). Conformity to sex-typed norms, affect, and the self-concept. *Journal of Personality and Social Psychology*, *73*(3), 523-538

World Economic Forum. The Global Gender Gap Report. (2018). Accessed on 24th February 2019 at http://www3.weforum.org/docs/WEF_GGGR_2018.pdf

Wright, P. H. (1988). Interpreting research on gender differences in friendship: A case for moderation and a plea for caution. *Journal of Social and Personal Relationships, 5*(3), 367-373.

Wright, P. H., & Scanlon, M. B. (1991). Gender role orientations and friendship: Some attenuation, but gender differences abound. *Sex Roles, 24*(9-10), 551-566.

Xu, K., Nosek, B., & Greenwald, A. (2014). Psychology data from the race implicit association test on the project implicit demo website. *Journal of Open Psychology Data, 2*(1), e3.

Yoder, J. D., Fischer, A. R., Kahn, A. S., & Groden, J. (2007). Changes in students' explanations for gender differences after taking a psychology of women class: More constructionist and less essentialist. *Psychology of Women Quarterly, 31*(4), 415-425.

Yoder, J.D. (1995). Socialization practices: Learning to be ourselves in a gender polarized world. *Psychology of Women Quarterly, 21*(3), 51-79.

Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences, 22*(2), 100-110.

Zalk, S. R. (1990). *Men in the academy: A psychological profile of harassment.* In M. A Paludi (Ed.). *Ivory power: Sexual harassment on campus* (pp 141-175). Albany, NY: State University of New York Press.

Zarate, A. M. & Smith, E. R.*(*1990). Person categorization and stereotyping. *Social Cognition, 8*(2), 161-185.

Zeigler-Hill, V., Besser, A., Morag, J., & Campbell, W. K. (2016). The dark triad and sexual harassment proclivity. *Personality and Individual Differences, 89*, 47-54.

Zimman, L., & Hall, K. (2016). *Language, gender, and sexuality.* New York: Oxford University Press.

Zimmerman, D. H., and West, C. (1975). Sex Role, Interruptions and Silences in Conversation. In: B. Thorne and N. Henley (Eds). *Language and Sex: Differences and Dominance* (pp 105-129). Rowley, MA: Newbury House.

Zitelny, H., Shalom, M., & Bar-Anan, Y. (2017). What is the implicit gender-science stereotype? Exploring correlations between the gender-science IAT and self-report measures. *Social Psychological and Personality Science*, *8*(7), 719-735.

Appendix 1

Consent Form

In agreeing to participate in this research I understand the following:

This research is being conducted by _____, an undergraduate student at the Department of Psychology, Maynooth University, under the supervision of Dr. Carol Murphy. It is the responsibility of the student to adhere to professional ethical guidelines in their dealings with participants and the collection and handling of data. If I have any concerns about participation, I understand that I may refuse to participate or withdraw at any stage of my participation.

I have been informed as to the general nature of the study. I understand that as a requirement of participating in the study I will be exposed to a computer-based task which will involve the presentation of words related to gender. I am also happy to complete a series of questionnaires that will ask me questions about my attitude to sex and gender matters.

All data from the study will be treated confidentially. My name will not be recorded and so my data will not be identified by name at any stage of the data analysis or in the final report. The data will be compiled, analysed and submitted in a report to the Psychology Department, Maynooth University.

I understand that no clinical judgement can be made of me on the basis of my participation or performance during this research and that because this is a group-based study my own individual responses on the questionnaire and the computer-based tasks are of no interest to the researchers.

At the conclusion of my participation, any questions or concerns I have will be fully addressed.

I may withdraw from this study at any time and may withdraw my data at the conclusion of my participation if I still have concerns. However, I understand that once I leave the experiment, I can no longer withdraw my data as it will not be identifiable by name.

I am over 18 years of age.

Signed in duplicate

_____Participant

_____ Researcher

_____ Date

Appendix 2

Gender and Desirability Ratings Study 1 (n = 40)

Trait List	Masculinity-Femininity Rating	Desirability in Men Rating	Desirability in Women Rating
Abrasive	2.65	2.08	2.03
Adaptable	2.95	3.80	3.83
Affectionate	3.97	3.95	4.25
Aggressive	2.00	1.78	1.80
Arrogant	2.18	1.98	1.65
Artistic	3.48	3.60	3.83
Athletic	2.08	3.90	3.30
Bitchy	4.08	1.50	1.65
Blunt	3.00	2.58	2.63
Bossy	3.75	1.90	2.00
Callous	2.60	1.65	1.50
Capable	2.93	4.08	4.08
Charismatic	2.60	4.23	4.08
Communicative	3.55	4.13	4.25
Compassionate	3.75	4.00	4.23
Competitive	2.00	3.40	2.90
Conscientious	3.38	3.50	3.60
Decisive	2.70	3.95	3.80
Dependent	3.37	2.53	2.88
Dishonest	2.80	1.40	1.40
Dismissive	2.88	1.70	1.65
Dominant	2.05	2.70	2.45
Domineering	2.60	2.35	2.13
Driven	3.15	4.28	3.83
Emotional	4.17	3.15	3.53
Empathetic	3.90	3.88	4.00
Forceful	2.38	2.08	1.83
Gentle	4.13	3.60	4.10
Helpful	3.47	4.10	4.03
Helpless	3.35	1.40	1.93
Independent	3.00	4.25	4.10
Inefficient	2.72	1.65	1.80
Insecure	3.65	1.85	1.78
Insensitive	2.20	1.83	1.73
Jealous	3.38	1.90	1.70
Loyal	3.00	4.48	4.35
Nurturing	4.25	3.53	4.35

Optimistic	3.20	4.10	4.13
Passive	2.93	2.08	2.53
Pessimistic	2.93	1.78	1.80
Polite	3.65	4.13	4.13
Practical	2.75	4.00	3.73
Punctual	3.40	3.75	3.78
Rational	2.85	3.75	3.80
Secretive	3.28	1.80	1.85
Self-sufficient	2.80	1.53	1.53
Selfish	2.82	3.90	3.95
Sensitive	3.98	3.45	4.08
Sincere	3.35	4.13	4.40
Sociable	3.20	4.18	4.13
Stubborn	2.90	2.23	2.23
Submissive	3.33	1.98	2.45
Tactful	3.28	3.48	3.35
Tactless	2.53	1.75	1.68
Unemotional	1.88	2.05	1.95
Unpredictable	3.20	2.65	2.68
Unreliable	2.60	1.35	1.65
Vain	3.17	1.78	1.83
Violent	2.20	1.28	1.33
Weak	3.40	1.73	2.08
Witty	2.60	4.30	4.13

Appendix 3

Gender and Desirability Ratings Study 1 (n = 228)

Trait List	Masculinity-Femininity Rating	Desirability in Men Rating	Desirability in Women Rating
Abrasive	2.71	1.88	1.89
Adaptable	3.08	4.16	4.17
Affectionate	3.75	4.14	4.49
Aggressive	2.16	1.67	1.56
Arrogant	2.21	1.63	1.59
Artistic	3.39	3.61	3.89
Athletic	2.46	3.92	3.64
Bitchy	4.02	1.54	1.64
Blunt	2.59	2.51	2.41
Bossy	3.51	1.98	1.93
Callous	2.89	1.73	1.71
Capable	3.02	4.32	4.28
Charismatic	2.67	4.27	4.15
Communicative	3.61	4.32	4.41
Compassionate	3.71	4.32	4.45
Competitive	2.50	3.40	3.32
Conscientious	3.35	3.70	3.73
Decisive	2.76	4.03	4.02
Dependent	3.31	2.53	2.63
Dishonest	2.86	1.32	1.26
Dismissive	2.74	1.66	1.64
Dominant	2.32	2.56	2.43
Domineering	2.64	2.35	2.22
Driven	3.11	4.33	4.23
Emotional	3.89	3.24	3.52
Empathetic	3.65	3.94	4.11
Forceful	2.29	1.83	1.84
Gentle	3.86	3.82	4.21
Helpful	3.24	4.29	4.35
Helpless	3.25	1.63	1.75
Independent	3.08	4.22	4.29
Inefficient	2.78	1.67	1.64
Insecure	3.57	1.83	1.91
Insensitive	2.41	1.72	1.61
Jealous	3.26	1.80	1.81
Loyal	3.06	4.65	4.68
Nurturing	4.02	3.74	4.32

Optimistic	3.23	4.26	4.31
Passive	3.14	2.25	2.19
Pessimistic	2.80	1.73	1.64
Polite	3.39	4.28	4.46
Practical	2.83	4.19	3.99
Punctual	3.17	3.92	3.98
Rational	2.88	3.93	3.94
Secretive	3.12	1.71	1.72
Self-sufficient	3.05	4.21	4.21
Selfish	3.00	1.48	1.42
Sensitive	3.80	3.57	3.89
Sincere	3.15	4.30	4.44
Sociable	3.13	4.37	4.40
Stubborn	2.86	2.11	2.08
Submissive	3.37	2.15	2.20
Tactful	3.26	3.40	3.46
Tactless	2.68	2.00	1.97
Unemotional	2.21	1.86	1.70
Unpredictable	3.11	2.75	2.76
Unreliable	2.60	1.30	1.39
Vain	3.21	1.76	1.85
Violent	2.21	1.29	1.22
Weak	3.37	1.86	2.02

Note. Traits were rated on a five-point Likert-Style scale from very masculine (1) to very feminine (5).

Appendix 4

Generic IRAP Experimental Protocol

Protocol

1. Assign participant code based on gender of participant (see counterbalancing sheet below).
2. Predictor task (LSH, human-as-male task, or hiring task) and demographic information (age and gender) via Google form
3. IRAP(s) based on counterbalancing sheet below.
4. Modern sexism scale via Google form

Counterbalancing

Gender of participant	**Participant code** (NB probably not the order in which they are run!)	**First IRAP**	**Second IRAP**
Female	1	a	b
Female	2	a	b
Female	3	b	a
Female	4	b	a
Female	5	a	b
Female	6	a	b
Female	7	b	a
Female	8	b	a
Female	9	a	b
Female	10	a	b
Female	11	b	a
Female	12	b	a
Female	13	a	b
Female	14	a	b
Female	15	b	a
Female	16	b	a
Female	17	a	b
Female	18	a	b
Female	19	b	a
Female	20	b	a
Female	21	a	b
Female	22	a	b
Female	23	b	a
Female	24	b	a
Female	25	a	b
Female	26	a	b
Female	27	b	a
Female	28	b	a
Male	29	a	b

Male	30	a	b
Male	31	b	a
Male	32	b	a
Male	33	a	b
Male	34	a	b
Male	35	b	a
Male	36	b	a
Male	37	a	b
Male	38	a	b
Male	39	b	a
Male	40	b	a
Male	41	a	b
Male	42	a	b
Male	43	b	a
Male	44	b	a
Male	45	a	b
Male	46	a	b
Male	47	b	a
Male	48	b	a
Male	49	a	b
Male	50	a	b
Male	51	b	a
Male	52	b	a
Male	53	a	b
Male	54	a	b
Male	55	b	a
Male	56	b	a

Appendix 5

Modern Sexism Scale (Swim et al., 1995)

Subject: _____
Gender: _____
Age: ____

Below are a number of statements measuring your attitudes and beliefs toward gender. **Please read each statement carefully using the scale below to make your choice.**

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree

1	Women are generally not as smart as men	1	2	3	4	5
2	I would be equally comfortable having a woman as a boss as a man.*	1	2	3	4	5
3	It is more important to encourage boys than to encourage girls to participate in athletics.	1	2	3	4	5
4	When both parents are employed and their child gets sick at school, the school should call the mother rather than the father.	1	2	3	4	5
5	Women are just as capable of thinking logically as men.*	1	2	3	4	5
6	Discrimination against women is no longer a problem in Ireland.	1	2	3	4	5
7	Women often miss out on good jobs due to sexual discrimination.*	1	2	3	4	5
8	It is rare to see women treated in a sexist manner on television.	1	2	3	4	5
9	Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences.	1	2	3	4	5
10	It is easy to understand the anger of women's groups in Ireland.*	1	2	3	4	5

*** Item is reverse-scored. Sum items to give an overall score.**

Appendix 6

Heteronormativity Attitudes and Beliefs Scale (Habarth, 2015)

Subject: _____
Gender: _____
Age: ____

Below are a number of statements measuring your attitudes and beliefs toward gender and sexuality. **Please read each statement carefully using the scale below to make your choice.**

Strongly disagree	Disagree	Slightly disagree	Exactly neutral	Slightly agree	Agree	Strongly agree
1	2	3	4	5	6	7

Femininity and masculinity are determined by biological factors, such as genes and hormones, before birth.	1	2	3	4	5	6	7
There are only two sexes: male and female.	1	2	3	4	5	6	7
All people are either male or female	1	2	3	4	5	6	7
In intimate relationships, women and men take on roles according to gender for a reason; it's really the best way to have a successful relationship	1	2	3	4	5	6	7
In intimate relationships, people should act only according to what is traditionally expected of their gender.	1	2	3	4	5	6	7
Gender is the same thing as sex.	1	2	3	4	5	6	7
It's perfectly okay for people to have intimate relationships with people of the same sex.	1	2	3	4	5	6	7
The best way to raise a child is to have a mother and a father raise the child together.	1	2	3	4	5	6	7
In healthy intimate relationships, women may sometimes take on stereotypical 'male' roles, and men may sometimes take on stereotypical 'female' roles.*	1	2	3	4	5	6	7
Sex is complex; in fact, there might even be more than 2 sexes.*	1	2	3	4	5	6	7

Gender is a complicated issue, and it doesn't always match up with biological sex.*	1	2	3	4	5	6	7
Women and men need not fall into stereotypical gender roles when in an intimate relationship.*	1	2	3	4	5	6	7
People should partner with whomever they choose, regardless of sex or gender.*	1	2	3	4	5	6	7
There are particular ways that men should act and particular ways that women should act in relationships.	1	2	3	4	5	6	7
People who say that there are only two legitimate genders are mistaken.	1	2	3	4	5	6	7
Gender is something we learn from society.*	1	2	3	4	5	6	7

*** Item is reverse-scored. Sum all items (including reverse-scored items) to give an overall score.**

Appendix 7

This experiment explores how we associate certain personality traits with men and women.

Specifically, whether our gut reaction is to assume that men and women have different personality traits relative to one another.

We therefore make a distinction between “masculine” personality traits and “feminine” personality traits. You will need to learn which list each personality trait fits into in order to complete this experiment.

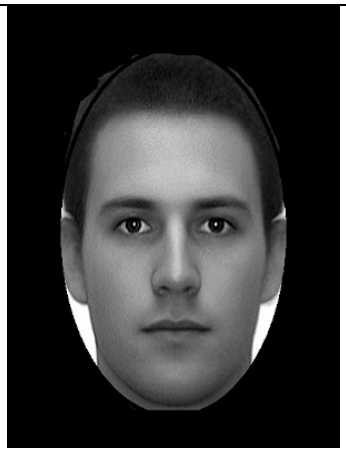
You might not agree that each of the below traits are either masculine or feminine. That’s ok, just try to learn which are places on each list.

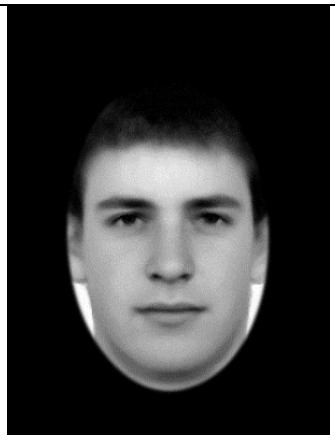
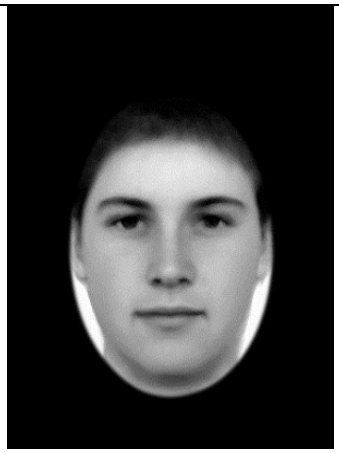
Masculine personality traits	**Feminine personality traits**
Witty	Nurturing
Competitive	Gentle
Decisive	Sensitive
Charismatic	Affectionate

Appendix 8

Androcentrism Task

Below you will be presented with a series of faces. For each image, please select whether you think the person is male or female. There is no need to spend too long on each item, but do try to be as accurate as possible.

1		a. Male b. Female
2		a. Female b. Male
3		a. Female b. Male

4		a. Male b. Female
5		a. Female b. Male
6		a. Male b. Female
7		c. Male d. Female

Appendix 9

This experiment explores how we perceive other people's emotions, and whether is related to gender.

Specifically, whether our gut reaction is to assume that men and women have more or less complex emotional lives than one another.

We therefore make a distinction between "simple" emotions and "complex" emotions. You will need to learn which list each emotion fits into in order to complete this experiment.

You might not agree that each of the below emotions are either simple or complex. That's ok, just try to learn which are places on each list.

Simple emotions	Complex emotions
Fear	Apprehension
Anger	Contempt
Happiness	Pride
Sadness	Disappointment

Appendix 10

Likelihood to Sexually Harass Scale (Pryor, 1987)

Instructions
On the sheets that follow you will find 10 brief scenarios that describe 10 different interactions between males and females. In each case you will be asked to imagine that you are the main male character in the scenario. Then you will be asked to rate how likely it is that you would perform each of several different behaviours in the described social context. Assume in each scenario that no matter what you choose to do, nothing bad would be likely to happen to you as a result of your action. Try to answer each question as honestly as you can. Your answers will be completely anonymous. No one will ever try to discover your identity, no matter what you say on the questionnaire.

Scenario #1
Imagine that you are an executive in a large corporation. You are 42 years old. Your income is above average for people at your job level. You have had numerous job offers from other companies. You feel very secure in your job. One day your personal secretary decides to quit her job and you have the task of replacing her. The personnel department sends several applicants over for you to interview. All seem to be equally qualified for the job. One of the applicants, Michelle S., explains during her interview that she desperately needs the job. She is 23 years old, single and has been job hunting for about a month. You find yourself very attracted to her. She looks at you in a way that possibly conveys she is also attracted to you. How likely are you to do the following things in this situation?
a. Would you give her the job over the other applicants? (Circle a number to indicate your response.)
b. Assuming that you are secure enough in your job that no possible reprisals could happen to you, would you offer her the job in exchange for sexual favours? (Circle a number to indicate your response.)
c. Assuming that you fear no reprisals on your job, would you ask her to meet you later for dinner to discuss her possible employment?

Scenario #2
Imagine that you are the owner and manager of an expensive restaurant. One day, while going over the receipts, you discover that one of the waitresses has made some errors in her checks. She has undercharged several customers. The mistake costs you $100. In talking to some of the other employees, you find that the particular customers involved were friends of the waitress. You call her into your office and ask her to explain her behaviour. The waitress confesses to having intentionally undercharged her friends. She promises that she will never repeat this dishonest act and tells you that she will do anything to keep her job. The waitress is someone you have always found particularly attractive. She is a divorcee and about 25 years old. How likely are you to do the following things in this situation?
a. Would you let her keep her job?
b. Would you let her keep her job in exchange for sexual favours?
c. Would you ask her to meet you for dinner after work to discuss the problem?

Scenario #3

Imagine that you are the manager of a shipping company. One day your supervisor asks you to study the possibility of buying several computers for the office. You call up several competing companies that sell computers. Each company sends a sales representative over to your office who describes the company's products. A salesperson from company "A" calls you and asks to come to your office. You agree and the next day a very attractive woman shows up. She can offer no real reason for buying her company's products over those of the other companies. However, she seems very sexy. How likely are you to do the following things in this situation?
a. Would you recommend her line of computers?
b. Assuming that you are secure enough in your job that no possible reprisals could happen to you, would you agree to recommend her line of computers in exchange for sexual favours?
c. Given the same assumptions as the last question above, would you ask her to meet you later for dinner to discuss the choice of computers?

Scenario #4
Imagine that you are a Hollywood film director. You are casting for a minor role in a film you are planning. The role calls for a particularly stunning actress, one with a lot of sex appeal. How likely are you to do the following things in this situation?
a. Would you give the role to the actress whom you personally found sexiest?
b. Would give the role to an actress who agreed to have sex with you?
c. Would ask the actress to whom you were most personally attracted to talk with you about the role over dinner?

Scenario #5
Imagine that you are the owner of a modelling agency. Your agency specializes in sexy female models used in television commercials. One of your models, Amy T., is a particularly ravishing brunette. You stop her after work one day and ask her to have dinner with you. She coldly declines your offer and tells you that she would like to keep your relationship with her "strictly business." A few months later you find that business is slack and you have to lay off some of your employees. You can choose to lay off Amy or one of four other women. All are good models, but someone has to go. How likely are you to do the following things in this situation?
a. Would you fire Amy?
b. Assuming that you are unafraid of possible reprisals, would you offer to let Amy keep her job in return for sexual favours?
c. Would you ask Amy to dinner so that you could talk over her future employment?

Scenario #6
Imagine that you are a college professor. You are 38 years old. You teach in a large midwestern university. You are a full professor with tenure. You are renowned in your field (Abnormal Psychology) and have numerous offers for other jobs. One day following the return of an examination to a class, a female student stops in your office. She tells you that her score is one point away from an "A" and asks you if she can do some extra credit project to raise her score. She tells you that she may not have a sufficient grade to get into graduate school without the "A." Several other students have asked you to do extra credit assignments and you have declined to let them. This particular woman is a stunning blonde. She sits in the front row of the class every day and always wears short skirts. You find her extremely sexy. How likely are you to do the following things in this situation?
a. Would you let her carry out a project for extra credit (e.g. write a paper)?

b. Assuming that you are very secure in your job and the university has always tolerated professors who make passes at students, would you offer the student a chance to earn extra credit in return for sexual favours?
c. Given the same assumptions as in the question above, would you ask her to join you for dinner to discuss the possible extra credit assignments?

Scenario #7
Imagine that you are a college student at a large Midwestern university. You are a junior who just transferred from another school on the East coast. One night at a bar you meet an attractive female student named Rhonda. Rhonda laments to you that she is failing a course in English Poetry. She tells you that she has a paper due next week on the poet, Shelley, and fears that she will fail since she has not begun to write it. You remark that you wrote a paper last year on Shelley at your former school. Your paper was given an A+. She asks you if you will let her use your paper in her course. She wants to just retype it and put her name on it. How likely are you to do the following things in this situation?
a. Would you let Rhonda use your paper?
b. Would you let Rhonda use your paper in exchange for sexual favours?
c. Would you ask Rhonda to come to your apartment to discuss the matter?

Scenario #8
Imagine that you are the editor for a major publishing company. It is your job to read new manuscripts of novels and decide whether they are worthy of publication. You receive literally hundreds of manuscripts per week from aspiring novelists. Most of them are screened by your subordinates and thrown in the trash. You end up accepting about one in a thousand for publication. One night you go to a party. There you meet a very attractive woman named Betsy. Betsy tells you that she has written a novel and would like to check into getting it published. This is her first novel. She is a dental assistant. She asks you to read her novel. How likely are you to do the following things in this situation.
a. Would you agree to read Betsy's novel?
b. Would you agree to reading Betsy's novel in exchange for sexual favours?
c. Would you ask Betsy to have dinner with you the next night to discuss your reading her novel?

Scenario #9
Imagine that you are a physician. You go over to the hospital one day to make your rounds visiting your patients. In looking over the records of one of your patients, you discover that one of the attending nurses on the previous night shift made an error in administering drugs to your patient. She gave the wrong dosage of a drug. You examine the patient and discover that no harm was actually done. He seems fine. However, you realize that the ramifications of the error could have been catastrophic under other circumstances. You pull the files and find out who made the error. It turns out that a new young nurse named Wendy H. was responsible. You have noticed Wendy in some of your visits to the hospital and have thought of asking her out to dinner. You realize that she could lose her job if you report this incident. How likely are you to do each of the following things?
a. Would you report Wendy to the hospital administration?
b. Assuming that you fear no reprisals, would you tell Wendy in private that you will not report her if she will have sex with you?

c. Assuming that you fear no reprisals, would you ask Wendy to join you for dinner to discuss the incident?

Scenario #10

Imagine that you are the news director for a local television station. Due to some personnel changes you have to replace the anchor woman for the evening news. Your policy has always been to promote reporters from within your organization when an anchor woman vacancy occurs. There are several female reporters from which to choose. All are young, attractive, and apparently qualified for the job. One reporter, Loretta W., is someone whom you personally find very sexy. You initially hired her, giving her a first break in the TV news business. How likely are you to do the following things in this situation?

a. Would give Loretta the job?

b. Assuming that you fear no reprisals in your job, would you offer Loretta the job in exchange for sexual favours?

c. Assuming that you fear no reprisals in your job, would you ask her to meet you after work for dinner to discuss the job?